

ON STRATEGIES OF SAMPLING FINITE POPULATIONS IN ONE AND TWO PHASES

(A REVISED VERSION)

ARUN KUMAR ADHIKARY

INDIAN STATISTICAL INSTITUTE
CALCUTTA

1982

Revised 1984

PREFACE

This thesis is being submitted to the Indian Statistical Institute (ISI) in fulfilment of the primary requirements for the award of the degree of Doctor of Philosophy in Statistics.

No part of this thesis was submitted to any other Institute for any degree, diploma, certificate etc.

In this thesis we have mainly presented results of our investigation on optimal strategies for surveying finite populations in two phases using selection procedures with varying probabilities. Incidentally we have added a few more results to the available literature on uni-phase sampling as are considered important in formulating appropriate double sampling plans. Optimal and pragmatically useful single phase sampling strategies employing 'randomized' rather than 'direct' social enquiries to take account of stigmatizing quantitative characters are also derived and discussed. A case study based on an enquiry among a sample of workers of the Indian Statistical Institute, Calcutta, illustrating an application of our proposed procedure for 'randomized' trial is reported in Appendix 4. This thesis includes materials which in some form constitute contents of (1) a paper by me

jointly with Arijit Chaudhuri being published in Journal of Indian Society of Agricultural Statistics, XXXIV, 1982 under the title

"On the relative efficiencies of several strategies for estimating a finite population mean"

and of (2) the under-noted technical reports :

- (i)* A study of the efficiency of Midzuno-Sen strategy relative to ratio and ratio - type estimators under a linear regression model. Technical Report No. ASC/81/1, Indian Statistical Institute, Calcutta.
- (ii) On optimality of double sampling strategies with varying probabilities. Technical Report No. ASC/81/3, Indian Statistical Institute, Calcutta.
- (iii) On sampling strategies with randomized response trials and their properties and relative efficiencies. Technical Report No. ASC/81/5, Indian Statistical Institute, Calcutta.
- (iv) On double sampling strategies with uniform admissibility and other optimality properties. Technical Report No. ASC/81/10, Indian Statistical Institute, Calcutta.

*To appear in Biometrika as "On the efficiency of Midzuno-Sen strategy relative to several ratio-type estimators under a model."

This dissertation was prepared under the active and enthusiastic guidance of Dr. Arijit Chaudhuri to whom I owe a deep debt of gratitude for his constant encouragement and careful supervision throughout the work and his permission to include some of our joint works in this thesis.

I also wish to record my gratitude to Prof. J. Roy who kindly granted me enough off-time to take up a research programme for the Ph.D. degree of ISI. I must also acknowledge my indebtedness to Prof. B.P. Adhikari, Director, ISI, whose helpful intervention enabled me to have an access to official records needed for the implementation of a case-study reported in Appendix 4. Our gratitude is also due to Prof. J.N.K. Rao who initiated our interest in double sampling problems and to Dr. B.K. Sinha whose occasional remarks led to partial revision of our preliminary drafts and to the examiner whose suggestions led to this revised and improved version.

I also wish to thank Messrs. M. Pal, S. Dasgupta, P. Sen, N. Behara and K. Banerjee of the Computer Science Unit of ISI who helped me in various ways in getting over various hurdles in submitting the thesis in the present form. I also hereby acknowledge the kind co-operation of my colleagues of ISI in my implementation of the case study mentioned earlier.

Finally I must express my sincerest thanks to Shri
and Dilip Bardhan
Jyotirmay Sarma, who typed the thesis with utmost care,
competence and patience.

Arun Kumar Adhikary

September, 1982

August, 1984

Applied Statistics, Surveys
and Computing Division,
Indian Statistical Institute,
Calcutta.

C O N T E N T S

	SUMMARY OF TOPICS	1
CHAPTER 1	: FORMULATION OF THE PROBLEM AND SUMMARY OF RESULTS	2-15
1.1	: On sampling strategies in two phases	2
1.2	: On the performances of some single- phase sampling strategies	8
1.3	: On the performances of sampling strategies with randomized response trials	13
CHAPTER 2	: ON SAMPLING STRATEGIES IN TWO PHASES	16-104
2.1	: Summary	16
2.2	: Introduction and formulation of the problem	17
2.3	: The results	18
2.4	: Concluding remarks concerning choice of two-phase strategies	100
2.5	: Unbiased estimation of variances for estimators in two-phase sampling	104

CHAPTER 3	: ON THE PERFORMANCES OF SOME SINGLE-PHASE SAMPLING STRATEGIES	105-193
3.1	: Summary	105
3.2.1	: Relative efficiencies of several strategies for estimating a finite population mean	108
3.2.2	: Exact (small sample) formulae concerning relative efficiencies of sampling strategies under Model I with gamma-distributed auxiliary variates	112
3.2.3	: Asymptotic formulae concerning relative efficiencies of strategies under Model II	114
3.2.4	: Relative efficiencies of the strategies	116
3.3.1	: On the efficiency of Midzuno/Sen strategy relative to several ratio- type estimators under a model	126
3.3.2	: A model and the strategies	127
3.3.3	: The biases and mean square errors	129

3.3.4 :	Approximate design - expectation of square errors	145
3.3.5 :	A few analytical but approximate results on MSE's	148
3.4.1 :	On the efficiencies of unbiased regression and regression - type estimators in finite populations	150
3.4.2 :	Formulation of the problem and the results	150
3.4.3 :	Variance estimation	159
3.5.1 :	A study of the effect of variate transformations on strategies of sampling finite populations	163
3.5.2 :	Modified strategies and certain models	165
3.5.3 :	Comparison of modified strategies with the corresponding original ones	168
3.5.4 :	Comparison of relative efficiencies of strategies under Model M_1	170
3.5.5 :	Comparison of relative efficiencies of strategies under Model M_2	173

3.5.6 :	Numerical values of relative efficiencies of strategies	175
3.6.1 :	On a use of variate - transformation in improving efficiency in two specific sampling strategies	187
3.6.2 :	A super-population model and asymptotic efficiencies of strategies	189

CHAPTER 4	:	SAMPLING STRATEGIES WITH RANDOMIZED RESPONSE TRIALS	194-227
-----------	---	---	---------

4.1 :	Summary	194
4.2 :	Introduction	195
4.3 :	Certain 'non-existence', 'admissibility' and 'optimality' results	196
4.4 :	Relative efficiencies of a few usual sampling strategies modified under randomized trials	210
4.5 :	Variance estimation	220

APPENDICES	228-249
------------	---------

REFERENCES	(i)-(ix)
------------	----------

A SUPPLEMENTARY CHAPTER	1-26
-------------------------	------

TITLE OF THE THESIS : ON STRATEGIES OF SAMPLING FINITE
POPULATIONS IN ONE AND TWO PHASES

BY

ARUN KUMAR ADHIKARY
INDIAN STATISTICAL INSTITUTE
CALCUTTA, INDIA.

SUMMARY OF TOPICS :

In estimating a finite population total (or mean) on taking samples in two phases with varying probabilities admissibility, uniform admissibility and other optimal properties under various formulations with Bayesian and non-Bayesian approaches have been established for certain strategies and estimators after showing non-existence of uniformly best strategies in general. Non-trivialities of these extensions over their 'single-phase' counterparts have been taken into considerations. For one phase sampling too, a few additions have been made to supplement the earlier literature to facilitate comparison among several strategies under appropriate modelling along with numerical illustrations. Investigations are also reported covering situations with surveys 'randomized' rather than 'direct' to preserve confidentiality relating to quantitative but stigmatizing characters, deriving, inter alia, a few optimality results.

CHAPTER - 1

FORMULATION OF THE PROBLEM AND SUMMARY OF RESULTS

1.1 ON SAMPLING STRATEGIES IN TWO PHASES :

We suppose that on the units of a finite population (of size N) $U = (1, \dots i, \dots N)$ certain size-measures W_j 's (say) are available and two variables x and y are closely positively correlated with W -values. Our purpose is to estimate the population total $Y = \sum_{j=1}^N y_j$ (or mean $\bar{Y} = \frac{Y}{N}$) of y using a suitable estimation procedure on choosing an appropriately designed sample. If it is relatively cheap to ascertain x -values rather than the y 's then the adoption of a suitably designed double sampling strategy seems to be called for - the first phase sample being drawn to provide data on x alone yielding an estimator for its total $X = \sum_{j=1}^N x_j$ and the second phase sample - a subsample from the first phase sample being intended to produce y -values in addition. Our next job will be to use the accumulated data on x and y for the over-all sample to get an estimate of Y . In what follows we set forth our plans to carry out such a mission in an organized way. Incidentally, we wish to see what may be ideally aimed at, what is not realizable and what are the features of the strategy incidentally available for adoption in realistic circumstances. Rao/Bellhouse (1978) seem to be

the pioneers in searching for optimal strategies with such an approach. Pursuing with their work we happen to get a number of results to be discussed below in brief (the details narrated in Chapter 2).

Let p_1 be a design for first-phase sample-selection (s' being a typical sample) and p_2 be the design for sub-sampling in the second phase (s being a typical sub-sample from s') so that p is the over-all double-phase sampling design. Here in general, s is a sub-sequence of s' which is a sequence of units of U . In particular, they may be sets of distinct units. In either case, we will write, $s \subset s'$. For a sample $\bar{s} = (s', s)$ so drawn we get values on x_j for $j \in s'$ and y_j for $j \in s$, the observed sample data being denoted as $d = (\bar{s}, x_j, y_j \mid j \in s, j' \in s')$. An estimator for Y based on d is taken as the sum of two components respectively (unbiasedly with respect to p) estimating $(Y-X)$ (based on y_j, x_j for $j \in s$) and X (based on x_j for $j \in s'$). Among such unbiased estimators, whatever p may be unless amounting to a census, one with the minimum variance uniformly in x_j 's and y_j 's (permitted to be any real number) is not available - an extension of Basu's (1971) result in case of single-phase sampling. However, if the estimator for Y based on d be taken as the sum of three

estimators for Y , $-X$ (based on $y_j, x_j, j \in s$) and X (based on $x_j, j \in s'$) homogeneously linear and unbiased (HLU) in y 's, x 's and x 's respectively, then (1) a unique uniformly minimum variance (UMV) estimator exists if p be a uni-cluster (uc) design (an extension of Hege (1965)/Hanurav(1966) result) and (2) no UMV estimator exists if p be a non-uni-cluster (nuc) design (an extension of Godambe's (1955) result). For any design (admitting an unbiased estimator), however, an estimator is identified as admissible in the above class of HLU estimators (extension of Godambe (1960), Roy/Chakravarti (1960) result in uni-phase sampling). However, if the estimators for Y , $-X$ and X as above are permitted to be any unbiased estimators (not necessarily of the HLU form) then an admissible estimator is available under severe restrictions on the class of such estimators and also on the class of designs p . The restrictions are motivated by the consideration that the estimators for Y and $-X$ based on second-phase sample alone must separately be admissible as estimators respectively for well-known admissible estimators for Y and $-X$ that might be based on y_j and x_j values for $j \in s'$ if $y_j, x_j, j \in s'$ were available, supposing that the estimator for X based on $x_j, j \in s'$ is also a usual admissible (based on uni-phase sampling) unbiased estimator for X . The

class of strategies involving the above estimator based on designs with fixed expected sample sizes is also shown to be uniformly admissible - extending Ramakrishnan's (1975) work. The details follow in Chapter 2. Further, if we consider the class of estimators for Y such that X is estimated using x_j 's for $j \in s'$ and $Y - X$ is estimated using $u_j = (y_j - x_j)$'s for $j \in s$ both the estimators being HLE's then 'uniform admissibility' is established for a sub-class of strategies within the above class of estimators and within the class of designs having fixed (1) sizes of s' 's and s 's at n' and n (say, respectively) and fixed (2) inclusion probabilities with respect to designs p_1 and p (extension of Sekkappan's (1973) work in uniphase sampling). Removing the restriction of linearity in estimators and conditions on inclusion probabilities but fixing average sample sizes at both phases and imposing unbiasedness condition on estimators uniform admissibility has been established for another sub-class of strategies extending the works of Ramakrishnan (1975).

Next we consider some customary models and under them find optimal strategies in the sense of attaining minimal values for model - expected variances of estimators. Whenever models are considered it is understood that p_2 is independent

of x -values. We first extend Cassel/Särndal/Wretman's (1976) (CSW, say) work to find a lower bound on the model - expected variance of an estimator (for Y) taken as the sum of three BLUE's for Y , $-X$ and X (as earlier) postulating y 's and x 's to be realizations on random variables (also denoted as y 's and x 's for simplicity) with finite variances. In doing so we assume common inter-correlations separately among y -pairs, x -pairs and y, x pairs. Besides, we also have to impose a mild restriction (relaxable if y, x pairs be uncorrelated) on the class of estimators. In case the model - expectations and standard deviations (of y_1 's and x_1 's) are proportional to known size-measures (W_1 's, say) then the lower bound is shown to be attained by a particular estimator based on designs with inclusion - probabilities for designs p as well as p_1 proportional to W 's.

In case the above - noted correlations are all absent, then in the same class of estimators one with the minimal expected variance is identified if based on fixed sample-size p and p_1 designs with inclusion - probabilities proportional to known size-measures (the so - called IPPS designs) provided the competitors are also based on IPPS designs, there being no conditions on the model variances, only the means being proportional to the size - measures. A corresponding result

on relaxing linearity restriction on estimators but imposing 'independence' assumption on the variables is established extending Godambe/Joshi's (1965) work by imposing additional conditions on the class of designs and estimators (details follow in Chapter 2).

Next we postulate a random permutation model (the results also carry over to postulation of exchangeable priors) for the vector with N co-ordinates of the form $(\frac{x_j}{a_j}, \frac{x_j}{b_j}, \frac{y_j}{b_j})$ with a_j 's and b_j 's as specified positive constants with $\sum_{j=1}^N a_j = n'$, $\sum_{j=1}^N b_j = n$ so as to show that the model - expected variance of a particular estimator is constant if inclusion - probabilities for p_1 and p respectively coincide with a_j 's and b_j 's which never exceeds that for any unbiased estimator (as the sum of unbiased estimators for Y , $-X$ and X) for Y based on any fixed ^{average} sample size design with average first and second phase sample - sizes as n' and n respectively. This is an extension of Godambe/Thompson's (1973) works in uniphase case.

Finally, we make a truly Bayesian approach, following Godambe (1969) basing the post-survey analysis on data already collected, no matter how by considering the posterior risk, the prior being formed after surveying the first-phase

sample on the eve of taking the second-phase sample. By showing that given any vector of variate -values there exists a prior assigning it a positive probability, 'admissibility' of an estimator (a Bayes estimator, of course) is proved for any two-phase sample design and the 'uniform admissibility' for a class of strategies with the above estimator and any design with fixed average sample sizes in either phase.

1.2 ON THE PERFORMANCES OF SOME SINGLE-PHASE SAMPLING STRATEGIES

In respect of two-phase sampling we have considered certain 'optimal' strategies available under certain circumstances. But in practice, it is difficult to check when these circumstances are plausible. So an alternative strategy may fare better than a so-called 'optimal' one in a situation not favouring the latter. So, various alternatives 'may' and 'should' be tried out in actual surveys. We refrain from suggesting them here because from the available literature on single-phase sampling one may be clever enough to pick up and mix up alternatives in either phase and permuting them in different ways. In one-phase sampling various 'optimal' strategies are well-known and it is also well-known that often those without such 'optimality properties' may steal the show with an edge over the optimal ones under specific circumstances. In what follows we throw a further light on

the relative performances of a few single phase sampling strategies so as to suggest intelligent choices from among them in suitable manners while adopting double-sampling procedures. Before we do so it may be of interest to note the following (applicable to double sampling):

Suppose in the first phase an SRSWOR, say, s' is chosen and one contemplates to use the first phase survey data on x_j for $j \in s'$ in choosing the second phase sub-sample s from s' and estimating Y directly from second phase sample using y_j values for j in s . For example, three alternatives may be, on using the normed size-measures $q_j = x_j / \sum_{j \in s'} x_j$ ($\forall j \in s'$) to (1) take an IPPS (inclusion-probability proportional to q_j 's for second phase sampling design p_2) sample s and use the Horvitz/Thompson (1952) (HT, say) estimator $e_1 = \frac{N}{n} \cdot \frac{1}{n} \sum_{j \in s} y_j / q_j$ or (2) take s using q_j 's and employing Rao/Hartley/Cochran (1962), (RHC, say) scheme and employing their estimator viz. $e_2 = \frac{N}{n} \sum_{j \in s} \frac{y_j}{q_j} \cdot S_j$, (where S_j 's are sums of q_j values over units falling in particular groups formed in applying RHC scheme) or (3) apply Midzuno/Sen (1952, 1953) (MS, say) scheme using q_j 's to choose s and employ Lahiri's (1951) procedure (LMS, in brief) to use the estimator $e_3 = \frac{N}{n} \sum_{j \in s} y_j / \sum_{j \in s} q_j$. If a linear (through the origin) regression model for y on x be postulated

with uncorrelated errors with zero means and conditional variances proportional to x^g (for $0 \leq g \leq 2$, as usual), then realising that the model-expected variances of HT estimator based on IPPS design, RHC estimator and LMS estimator each based on common size-measures and a common sample size in uni-phase sampling, denoted as M_1, M_2, M_3 (say, respectively) satisfy the relations [Vide Chaudhuri/ Arnab (1979).]

$$\begin{array}{lll} M_1 < M_2 < M_3 & \text{if } g > 1 \\ = & = & \text{if } g = 1 \\ > & > & \text{if } g < 1 \end{array}$$

we immediately get an exactly analogous relative standings for e_1, e_2 and e_3 (c.f. details in Chapter 2) to reach appropriate conclusions.

However, at present at least we have to concede that in case the first-phase sample is chosen with varying probabilities we are unable to suggest concrete but simple double-sampling strategies with second-phase sampling design utilizing the realized x_1 -values in the initial sample.

Turning to uni-phase sampling we may observe the following concerning the choice of strategies. Some of the

popular strategies are: (i) Horvitz/Thompson (1952) estimator based on a fixed sample-size IPPS-design; (ii) Rao/Hartley/Cochran (1962) strategy, (iii, iv) the ratio estimator based on SRSWOR or on Midzuno/Sen (1952, 1953) scheme of sampling (v) sample mean of ratios of y_1 's to x_1 's for SRSWOR, (vi) Hansen/Hurwitz (1943) strategy, (vii, viii) Raj's (1956) and Murthy's (1957) estimators based on PPSWOR sampling. Analytical comparisons of the first six with the last two seem difficult because the variances of latter are complicated. So, we will consider the first six only and later we will consider some variants of them derived on certain modifications. First we consider linear regression (not necessarily through the origin) models with identically independently gamma-distributed regressors and uncorrelated errors with zero means but conditional variances as functions of the regressors. The model-expected biases and mean square errors (MSE) of the above six strategies are then compared - analytically and numerically. Allowing more generality on the distributions for the regressor variates but assuming existence of second order moments certain asymptotic results also have been derived under assumption of large population and/or sample sizes. In these derivations law of large numbers has been applied on the x_1 's but not on the errors e_1 's (say) because

the latter are not assumed independently or identically distributed and about the conditional error variances no assumption being made to ensure validity of law of large numbers (LLN). However, in case LLN is also assumed to hold for the errors certain modifications on the relative standings among the strategies are also noted. Comparatively, the Lahiri/Midzuno/Sen (1951, 1952, 1953) (LMS) strategy is observed to stand out as a very promising one under several circumstances not favourable to the HT strategy studied with many optimal properties. So, next we consider this strategy vis-a-vis the strategies involving selection by SRSWOR method and using ratio estimator and a few derivatives of it based on half-sampling. Under other similar models too, LMS strategy figures impressively in this comparative study. In this investigation the status of regression estimator based on SRSWOR is difficult to ascertain because it does not admit a simple model-expected MSE expression except when errors are homoscedastic.

Singh and Srivastava (1980) have given a sampling scheme yielding an unbiased regression estimator and an unbiased regression type strategy. By model sampling from bivariate normal populations they have demonstrated superiority of their unbiased regression strategy to biased regression

estimation through SRSWOR sampling. But, considering more general models involving linear regression with homoscedasticity we identify situations giving varying levels of efficiencies for them under different circumstances.

In case the postulated regression line passes through the origin assuming x_i 's as fixed constants, results concerning relative efficiencies of various strategies are available in the literature. If, however, the regression line has a non-zero intercept α (say) with an unknown magnitude then if one may have some idea about its magnitude, it seems plausible to modify the strategies using origin-shifted y 's rather than the y 's themselves along with other necessary adjustments. That this is so is illustrated with reference to some standard strategies and their modifications. In addition relative efficiencies of several popular strategies modified with variate transformations (rather translations) are worked out with postulated identically independently gamma distributed regressors (details given in Chapter 3).

1.3 ON THE PERFORMANCES OF SAMPLING STRATEGIES WITH RANDOMIZED RESPONSE TRIALS

Sometimes the variable y of interest may relate to some sensitive character whose revelation may carry touches of

social and moral stigma. Then in a survey respondents (human beings, naturally) should be unwilling to truthfully divulge such values. An answer is then provided by what are called randomized rather than direct surveys - introduced by Warner (1965) with respect to qualitative characters and developed later by many workers covering quantitative ones as well. We investigate how sampling strategies with varying probabilities relating to such sensitive and hence confidential quantitative characters should be modified if the enquiry has to be 'randomized'. Under certain assumptions we obtain optimal sampling strategies, note non-existence of a uniformly best strategy and derive uniformly admissible strategies in case of randomized surveys, on modifying strategies available only with direct surveys. One major assumption here is that the population is so homogeneous (rendered so through stratification, if necessary) that in any specific situation for the unknown variate values y_i 's ($i = 1, \dots, N$), it is possible to specify an anticipated set of l_i values (say M in number, M may be equal to, more than or even less than N) supposed to cover the y -values. Then, the randomized device requires a respondent to divulge either the true y -value with a certain chance (c , say) or with the complementary chance $(1-c)$ to report one of the l_i -values taken at random out of all the l_i 's listed. Appropriate sampling strategies

are then required to be derived. Our solutions are detailed in Chapter 4. Some of the results available with direct surveys carry over with little modifications. However, the price paid due to randomization in terms of loss of efficiency and difficulties of discriminating among strategies is also earmarked in that Chapter.

CHAPTER-2

ON SAMPLING STRATEGIES IN TWO PHASES

2.1 SUMMARY:

Certain optimal properties are established for strategies for estimating a finite population total for a variable y of interest from a sample to be taken in two phases using second phase sample data on y and first phase sample data on a correlated variable x (appropriately combined) using varying probability sampling schemes in both phases. Recognising that known optimality results under unified set-ups for single phase sampling do not extend straightaway to double sampling situation, results have been derived, in the latter case, to show inter alia (i) non-existence of UMVUE's in certain classes of double sampling strategies, (ii) existence of UMVHLUE's for unicluster designs, (iii) admissibility of an estimator within certain homogeneous linear unbiased classes as well as within a general class of unbiased estimators, (iv) uniform admissibility (with Bayesian and non-Bayesian arguments) of classes of strategies, (v) minimality of model-expected design-variance of an estimator under super-population modelling.

INTRODUCTION AND FORMULATION OF THE PROBLEM

Our purpose is to investigate if certain well-known optimality results concerning strategies for sampling a finite population in a single phase extend in an obvious manner if the sampling is implemented in two phases. With specific formulations we show (we believe, in a non-trivial way) in what follows that some of them actually do. From a given finite population U of size N we suppose a first phase sample s' is chosen with a probability $p_1(s')$ and in the second phase a sub-sample s is chosen therefrom with a conditional probability $p_2(s) = p_2(s/s')$, the over-all sample $\bar{s} = (s', s)$ having the selection probability $p(\bar{s}) = p_1(s') \cdot p_2(s/s') = p_1(s') p_2(s)$, say, for brevity. Our problem is to unbiasedly estimate the finite population total Y for the variate (y) - values observed in the second phase on utilising variate (x) - values observed in the first phase sample and other auxiliary information available both at sampling and estimation stage, on choosing appropriate designs p_1, p_2 and estimators in suitable classes and if necessary on postulating reasonable models, warranted by the situations at hand.

2.3 THE RESULTS

Writing $y = (y_1 | i \in s)$ and $x' = (x_1 | i \in s')$ and $x = (x_1 | i \in s)$ when a sample $\bar{s} = (s', s)$ is observed with probability $p(\bar{s})$ we may use an unbiased (with respect to p) estimator $e = e(\bar{s}, y, x')$ for Y . Let this e be of the form $e(\bar{s}, y, x') = e_1(x, y, x) + e_2(s', x') = e_1 + e_2$ (say) such that $E_p(e_1) = Y - X$ and $E_p(e_2) = X$ so that $E_p(e) = Y$. Denoting these classes of designs and estimators as D and C respectively we have several results presented below. This may be generalized requiring the conditions $E_p(e_1) = Y + \psi_1 X$ and $E_p(e_2) = \psi_2 X$ such that ψ_1 and ψ_2 are constants satisfying $\psi_1 + \psi_2 = 0$. Choosing $\psi_2 = -\psi_1 = \psi$ (say), where ψ is some parameter dependent on an assumed model, Rao/Bellhouse (1978) (vide Theorem 4.5, page 136) obtained some results in this context. Noting that with such generalizations the final estimators must obviously depend on usually unknown parameters we consider the simpler and less general case as above in what follows. First, following Basu (1971), we have the

Theorem 2.3.1 : In the class C there does not exist a UNV (uniformly minimum variance) estimator for any design in the class D unless one has a census.

Proof: Let us write $\underline{A} = (A_1, \dots, A_1, \dots, A_N)$ for $\underline{Y} = (\mathcal{Y}_1, \dots, \mathcal{Y}_1, \dots, \mathcal{Y}_N)$ where the y-variate value $\mathcal{Y}_i$ equals a fixed value A_i ($i=1, \dots, N$) and $\underline{B} = (B_1, \dots, B_1, \dots, B_N)$ for $\underline{X} = (\mathcal{X}_1, \dots, \mathcal{X}_1, \dots, \mathcal{X}_N)$ where the x-variate value $\mathcal{X}_i$ equals a fixed value B_i ($i=1, \dots, N$) and also write 'a' for ($\mathcal{Y}_i = A_i$ lies), 'b' for ($\mathcal{X}_i = B_i$ lies) and 'b' for ($\mathcal{X}_i = B_i$ lies'). If an estimator $\bar{e}(\bar{s}, y, x') = \bar{e}_1(s, y, x) + \bar{e}_2(s', x')$ be the UMV among those in the class C (with obvious notations), then we may define an estimator

$$e^* = e^*(\bar{s}, y, x') = \bar{e}(\bar{s}, y, x') - \bar{e}(\bar{s}, a, b') + A$$

for fixed and known vectors $\underline{A}$ and $\underline{B}$

$$(\text{here } A = \sum_{i=1}^N A_i \text{ and } B = \sum_{i=1}^N B_i).$$

Writing

$$e_1^*(s, y, x) = \bar{e}_1(s, y, x) - \bar{e}_1(s, a, b) + (A - B)$$

$$\text{and } e_2^*(s', x') = \bar{e}_2(s', x') - \bar{e}_2(s', b') + B$$

it may be checked that e^* belongs to the class C. Also it follows that $\text{Var}(e^*) = 0$ when $\underline{Y} = \underline{A}$ and $\underline{X} = \underline{B}$. Since $\underline{A}$ and $\underline{B}$ are arbitrary, it follows that $\bar{e}$ cannot be UMV unless $\text{Var}(\bar{e})$ is uniformly zero which is possible only in case p corresponds to taking a census.

Next, we consider the class $C_1 (= C_{HL})$ of homogeneous linear unbiased estimators (HLUE) for Y of the form (following Rao/Bellhouse's (1978) notation)

$$\begin{aligned} e_{bs} &= \sum_{j \in s} b_{sj} y_j + \sum_{j \in s} b'_{sj} x_j + \sum_{j \in s'} b''_{sj} x_j \\ &= e_{bs}(y) + e'_{b's}(x) + e''_{b''s'}(x'), \text{ (say) where} \end{aligned}$$

b_{sj} , b'_{sj} , b''_{sj} 's are quantities independent of y_j 's and x_j 's so chosen that (of course implying $b_{sj} = b_{sj}$, $b'_{sj} = b'_{sj}$, $b''_{sj} = b''_{sj}$ throughout)

$$Y = E_p e_{bs}(y) = \sum_s \left[\sum_{s \subset s'} \left(\sum_{j \in s} b_{sj} y_j \right) p_2(s) \right] p_1(s')$$

$$- X = E_p e'_{b's}(x) = \sum_{s'} \left[\sum_{s \subset s'} \left(\sum_{j \in s} b'_{sj} x_j \right) p_2(s) \right] p_1(s')$$

($\sum_{s \subset s'}$ denoting summation over all s contained in s')

and

$$X = E_p e''_{b''s'}(x') = \sum_{s'} \left(\sum_{j \in s'} b''_{sj} x_j \right) p_1(s').$$

Let us write the inclusion-probabilities as

$$P_j = \sum_1 p_1(s') \text{ and } Q_j = \sum_1 \sum_2 p_2(s) p_1(s'),$$

$\sum_1 = \text{sum over } s' \ni j$, $\sum_2 = \text{sum over } s \ni j$. Later, we will write $\sum_3 = \text{sum over } s' \ni j, k$, $\sum_4 = \text{sum over } s \ni j, k$ and $\sum_5 = \text{sum over } s' - s \ni j$ (it being implied throughout that $s \subset s'$).

For every design in D we will assume throughout that $P_j > 0$, $Q_j > 0 \forall j$. Let us denote by $D_1 = D_u$ a sub-class of D (the so-called class of uni-cluster designs) such that every pair of samples s_1', s_2' with $p_1(s_1'), p_1(s_2') > 0$ must be either disjoint or (effectively) equivalent (both sharing a common set of distinct units) and any pair of sub-samples s_1, s_2 from any sample s' (with $p_1(s') > 0$) with $p_2(s_1|s') > 0$, $p_2(s_2|s') > 0$ is also either disjoint or (effectively) equivalent. Then, we have the

Theorem 2.3.2. For every design in $D_1 = D_u$, there exists a UMV estimator (for Y) in the class $C_1 = C_{HL}$ and is of the form

$$e_{b\bar{s}}^* = \sum_{j \in s} \frac{y_j - x_j}{Q_j} + \sum_{j \in s'} \frac{x_j}{P_j}.$$

For any design outside the class D_1 there does not exist a UMV estimator in the class C_1 .

Proof. For $e_{b\bar{s}}$ within C_1 we have

$$\begin{aligned} E(e_{b\bar{s}}) &= \sum_{j=1}^N y_j E_1 \left\{ \sum_2 b_{\bar{s}j} p_2(s) \right\} p_1(s') \\ &+ \sum_{j=1}^N x_j E_1 \left\{ \sum_2 b'_{\bar{s}j} p_2(s) \right\} p_1(s') \\ &+ \sum_{j=1}^N x_j E_1 b''_{\bar{s}j} p_1(s') \end{aligned}$$

$$\iff \sum_1 \left\{ \sum_2 b_{sj} p_2(s) \right\} p_1(s') = 1 \quad \forall j$$

$$\sum_1 \left\{ \sum_2 b'_{sj} p_2(s) \right\} p_1(s') = -1 \quad \forall j$$

$$\text{and } \sum_1 b''_{sj} p_1(s') = 1 \quad \forall j.$$

If an estimator in C for any design in D is to be a UMVE it must be independent of order and multiplicities with which units occur in samples because otherwise it will lie outside a complete class within C satisfying these two conditions and hence can be improved upon by Rao-Blackwellisation. A brief argument for this fact, following Murthy (1957), may obviously be put forward as follows:

For any ordered $\bar{s} = (s', s)$ with $s' = (i_1, \dots, i_n)$, say, $s = (i_{r+1}, \dots, i_{r+n})$, say, which is a subsequence of s' both with ordered units with any multiplicities, let $t = (\bar{s}, y, x')$ be an unbiased estimator for Y with respect to any ordered design p i.e. $E_p(t) = Y$. We may construct unordered samples of the form $\bar{S} = (S', S)$ where

$$S' = (j_1, \dots, j_{j'}) \equiv \text{the set of distinct units in } s',$$

$$S = (k_1, \dots, k_{j'}) \equiv \text{the set of distinct units in } s$$

and an unordered design P such that $P(\bar{S}) = \sum' p(\bar{s})$, $\sum'$ being sum over all $\bar{s}$'s with distinct units of s' 's in S' and

distinct units of s 's in S . Then, we may construct a new estimator $T = T(\bar{S}, y, x')$, say, such that $T = T(\bar{S}, y, x') = \frac{\sum' t(\bar{s}, y, x') p(\bar{s})}{P(\bar{S})}$. Then,

$$E_p(T) = \sum'' T(\bar{S}, y, x') P(\bar{S}), \text{ (where } \sum'' \text{ is sum over all } \bar{S}'\text{'s)}$$

$$= \sum t(\bar{s}, y, x') p(\bar{s}) = Y, \sum \text{ is sum over all } \bar{s}'\text{'s};$$

i.e. T is unbiased for Y .

Now

$$E_p(tT) = \sum'' T(\bar{S}, y, x') \sum' t(\bar{s}, y, x') p(\bar{s})$$

$$= \sum'' T^2(\bar{S}, y, x') P(\bar{S}) = E_p(T^2)$$

and hence $\text{Cov}(t, T) = E_p(T^2) - Y^2 = V_p(T)$ and hence $V_p(t) - V_p(T) = E_p(t - T)^2 \geq 0$ implying that t cannot have a smaller variance than that of T unless t coincides with T with probability 1. The latter is possible only if t is independent of order and/or multiplicities of units in $\bar{s}$.

If we have a design in $D_1 = D_u$, then writing $\bar{s}(j)$ for a set of all (effectively equivalent) samples (with s in s' with $p_1(s') > 0$) containing a unit j , we must have for an estimator $\hat{y}$ in C_1 to be a UMVE, that

$$b_{\bar{s}j} = b_{\bar{s}(j)j}, \quad b'_{\bar{s}j} = b'_{\bar{s}(j)j} \text{ and } b''_{\bar{s}j} = b''_{\bar{s}(j)j} \quad \forall \bar{s} \in \bar{s}(j).$$

From this it follows that

$$b_{\bar{s}(j)j} = \frac{1}{Q_j}, \quad b'_{\bar{s}(j)j} = -\frac{1}{Q_j} \quad \text{and} \quad b''_{\bar{s}(j)j} = \frac{1}{P_j} \quad \forall j.$$

So $e_{\bar{b}\bar{s}}^*$ is a unique member of C_1 and hence is the UMV in C_1 .

In order to justify the second assertion of the theorem let us now appeal to C.R. Rao's (1952) theorem applied in a manner that is followed essentially by Ericson (1974). We will now consider designs in D which are outside D_1 and for definiteness denote them as D_2 . For a design in D_2 , an estimator $\bar{e}_{\bar{b}\bar{s}}$ in C_1 is UMV if

$$\text{Cov}(\bar{e}_{\bar{b}\bar{s}}, \bar{e}_{\bar{b}\bar{s}} - e_{\bar{b}\bar{s}}) = 0 \quad \text{for every member } e_{\bar{b}\bar{s}} \text{ in } C_1$$

so that $\bar{e}_{\bar{b}\bar{s}} - e_{\bar{b}\bar{s}} = e_{\bar{a}\bar{s}}$ (say) is an unbiased estimator of 0, so that $e_{\bar{a}\bar{s}}$ may be said to belong to a class C_{01} of unbiased estimators of 0. Writing, in an obvious manner

$$\begin{aligned} \bar{e}_{\bar{b}\bar{s}} &= \sum_{j \in S} \bar{b}_{\bar{s}j} y_j + \sum_{j \in S} \bar{b}'_{\bar{s}j} x_j + \sum_{j \in S'} \bar{b}''_{\bar{s}j} x_j \\ e_{\bar{a}\bar{s}} &= \sum_{j \in S} (\bar{b}_{\bar{s}j} - b_{\bar{s}j}) y_j + \sum_{j \in S} (\bar{b}'_{\bar{s}j} - b'_{\bar{s}j}) x_j \\ &\quad + \sum_{j \in S'} (\bar{b}''_{\bar{s}j} - b''_{\bar{s}j}) x_j \\ &= \sum_{j \in S} a_{\bar{s}j} y_j + \sum_{j \in S} a'_{\bar{s}j} x_j + \sum_{j \in S'} a''_{\bar{s}j} x_j \quad (\text{say}), \end{aligned}$$

we have $\text{Cov}(\bar{e}_{bs}, \bar{e}_{bs} - e_{bs})$

$$= \text{Cov}(\bar{e}_{bs}, e_{as})$$

$$= \sum_{j=1}^N y_j^2 \Sigma_1 \left\{ \Sigma_2 \bar{b}_{sj} a_{sj} p_2(s) \right\} p_1(s')$$

$$+ \sum_{j+k=1}^N y_j y_k \Sigma_3 \left\{ \Sigma_4 \bar{b}_{sj} a_{sk} p_2(s) \right\} p_1(s')$$

$$+ \sum_{j=1}^N x_j^2 \Sigma_1 \bar{b}''_{sj} a''_{sj} p_1(s')$$

$$+ \Sigma_1 \left\{ \Sigma_2 \bar{b}'_{sj} (a'_{sj} + a''_{sj}) p_2(s) \right\} p_1(s')$$

$$+ \Sigma_1 \left\{ \Sigma_2 \bar{b}''_{sj} a'_{sj} p_2(s) \right\} p_1(s') \int$$

$$+ \sum_{j+k=1}^N x_j x_k \Sigma_3 \bar{b}''_{sj} a''_{sk} p_1(s')$$

$$+ \Sigma_3 \left\{ \Sigma_4 \bar{b}'_{sj} a'_{sk} p_2(s) \right\} p_1(s')$$

$$+ \Sigma_3 \left\{ \Sigma_2 (\bar{b}'_{sj} a''_{sk} + a'_{sj} \bar{b}''_{sk}) p_2(s) \right\} p_1(s') \int$$

$$+ \sum_{j=1}^N x_j y_j \Sigma_1 \left\{ \Sigma_2 (a_{sj} (\bar{b}'_{sj} + \bar{b}''_{sj}) + \bar{b}_{sj} (a'_{sj} + a''_{sj})) p_2(s) \right\} p_1(s')$$

$$+ \sum_{j+k=1}^N x_k y_j \Sigma_3 \left\{ \Sigma_4 (\bar{b}'_{sk} a_{sj} + \bar{b}_{sj} a'_{sk}) p_2(s) \right\} p_1(s')$$

$$+ \Sigma_3 \left\{ \Sigma_2 (\bar{b}''_{sk} a_{sj} + \bar{b}_{sj} a''_{sk}) p_2(s) \right\} p_1(s') \int.$$

So, $\text{Cov}(\bar{e}_{b\bar{s}}, e_{a\bar{s}}) = 0$ iff the following hold together, viz.,

$$\Sigma_1 \{ \Sigma_2 \bar{b}_{sj} a_{sk} p_2(s) \} p_1(s') = 0 \quad \forall j \quad (2.3.1)$$

$$\Sigma_3 \{ \Sigma_4 \bar{b}_{sj} a_{sk} p_2(s) \} p_1(s') = 0 \quad \forall j \neq k \quad (2.3.2)$$

$$\begin{aligned} & \Sigma_1 \bar{b}_{sj}'' a_{sj}'' p_1(s') + \Sigma_1 \{ \Sigma_2 \bar{b}_{sj}' (a_{sj}' + a_{sj}'') p_2(s) \} p_1(s') \\ & + \Sigma_1 \{ \Sigma_2 \bar{b}_{sj}'' a_{sj}' p_2(s) \} p_1(s') = 0 \quad \forall j \end{aligned} \quad (2.3.3)$$

$$\begin{aligned} & \Sigma_3 \bar{b}_{sj}'' a_{sk}'' p_1(s') + \Sigma_3 \{ \Sigma_4 \bar{b}_{sj}' a_{sk}' p_2(s) \} p_1(s') \\ & + \Sigma_3 \{ \Sigma_2 (\bar{b}_{sj}' a_{sk}'' + a_{sj}' \bar{b}_{sk}'') p_2(s) \} p_1(s') = 0 \quad \forall j \neq k \end{aligned} \quad (2.3.4)$$

$$\begin{aligned} & \Sigma_1 \{ \Sigma_2 (a_{sj}'' (\bar{b}_{sj}' + \bar{b}_{sj}'') + \bar{b}_{sj} (a_{sj}' + a_{sj}'')) p_2(s) \} p_1(s') \\ & = 0 \quad \forall j \end{aligned} \quad (2.3.5)$$

and

$$\begin{aligned} & \Sigma_3 \{ \Sigma_4 (\bar{b}_{sk}' a_{sj}'' + \bar{b}_{sj} a_{sk}') p_2(s) \} p_1(s') \\ & + \Sigma_3 \{ \Sigma_2 (\bar{b}_{sk}'' a_{sj}'' + \bar{b}_{sj} a_{sk}'') p_2(s) \} p_1(s') = 0 \quad \forall j \neq k \end{aligned} \quad (2.3.6)$$

In addition, we have (because of unbiasedness of $e_{b\bar{s}}$ and $\bar{e}_{b\bar{s}}$),

$$\Sigma_1 \{ \Sigma_2 a_{sj} p_2(s) \} p_1(s') = 0 \quad \forall j \quad (2.3.7)$$

$$\sum_1 \left\{ \sum_2 a'_{sj} p_2(s) \right\} p_1(s') = 0 \quad \forall j \quad (2.3.8)$$

$$\text{and} \quad \sum_1 a''_{sj} p_1(s') = 0 \quad \forall j \quad (2.3.9)$$

Let $\bar{s}_1 = (s'_1, s_1)$, $\bar{s}_2 = (s'_2, s_2)$ be two samples with $p(\bar{s}_1) > 0$, $p(\bar{s}_2) > 0$; $j \in \bar{s}_1 = (s'_1, s_1)$ and $j \in \bar{s}_2 = (s'_2, s_2)$ and $\bar{b}_{s_1j} \neq \bar{b}_{s_2j}$ for a supposedly UMVE $\bar{e}_b$ (of course, if one exists). Let an alternative e_b be chosen so that $\bar{b}_{sk} = b_{sk} \quad \forall k \neq j$ and $\forall \bar{s}$ except for $\bar{s}_1, \bar{s}_2 \rightarrow e_b$ is different from $\bar{e}_b$ but $E_p(e_b) = Y = E_p(\bar{e}_b)$.

$$\text{Then } 0 = E_p(\bar{e}_b - e_b) = a_{s_1j} p(\bar{s}_1) + a_{s_2j} p(\bar{s}_2)$$

$$\Rightarrow a_{s_1j} p(\bar{s}_1) = -a_{s_2j} p(\bar{s}_2) \quad (2.3.10)$$

But this contradicts (2.3.1) because, using (2.3.10), we have,

$$\begin{aligned} & \sum_1 \sum_2 \bar{b}_{sj} a_{sj} p_2(s) p_1(s') \\ &= \bar{b}_{s_1j} a_{s_1j} p(\bar{s}_1) + \bar{b}_{s_2j} a_{s_2j} p(\bar{s}_2) \\ &= (\bar{b}_{s_1j} - \bar{b}_{s_2j}) a_{s_1j} p(\bar{s}_1) \neq 0. \end{aligned}$$

To complete the proof we may note further the following.

Suppose, again, $j \in \bar{s}_1 = (s'_1, s_1)$, $j \in \bar{s}_2 = (s'_2, s_2)$ such that $p(\bar{s}_1) > 0$, $p(\bar{s}_2) > 0$ and for a supposedly UMVE $\bar{e}_b$ (if exists), in C_1 for a design in D_2 , namely,

$$\bar{e}_{b\bar{s}} = \sum_{j \in \bar{s}} \bar{b}_{\bar{s}j} y_j + \sum_{j \in \bar{s}} \bar{b}'_{\bar{s}j} x_j + \sum_{j \in \bar{s}'} \bar{b}''_{\bar{s}j} x_j,$$

$\bar{b}_{\bar{s}_1j} \neq \bar{b}_{\bar{s}_2j}$, $\bar{b}'_{\bar{s}_1j} \neq \bar{b}'_{\bar{s}_2j}$, $\bar{b}''_{\bar{s}_1j} \neq \bar{b}''_{\bar{s}_2j}$ and for a

different unbiased estimator $e_{b\bar{s}}$, $\bar{b}_{\bar{s}k} = b_{\bar{s}k}$,

$\bar{b}'_{\bar{s}k} = b'_{\bar{s}k}$ and $\bar{b}''_{\bar{s}k} = b''_{\bar{s}k}$ for all $k \neq j$ and for all $\bar{s}$

except $\bar{s}_1$ and $\bar{s}_2$. Then,

$$a_{\bar{s}_1j} p(\bar{s}_1) + a_{\bar{s}_2j} p(\bar{s}_2) = 0 \quad (2.3.11)$$

$$a'_{\bar{s}_1j} p(\bar{s}_1) + a'_{\bar{s}_2j} p(\bar{s}_2) = 0 \quad (2.3.12)$$

$$\text{and } a''_{\bar{s}_1j} p(\bar{s}_1) + a''_{\bar{s}_2j} p(\bar{s}_2) = 0. \quad (2.3.13)$$

But these contradict (2.3.5), because, using (2.3.11),

(2.3.12) and (2.3.13), we have

$$\sum_1 \sum_2 \int a_{\bar{s}j} (\bar{b}'_{\bar{s}j} + \bar{b}''_{\bar{s}j}) + \bar{b}_{\bar{s}j} (a'_{\bar{s}j} + a''_{\bar{s}j}) p_2(s) p_1(s')$$

$$\begin{aligned}
 &= a_{\bar{s}_1 j} (\bar{b}'_{\bar{s}_1 j} + \bar{b}''_{\bar{s}_1 j}) p(\bar{s}_1) + a_{\bar{s}_2 j} (\bar{b}'_{\bar{s}_2 j} + \bar{b}''_{\bar{s}_2 j}) p(\bar{s}_2) \\
 &+ \bar{b}_{\bar{s}_1 j} (a'_{\bar{s}_1 j} + a''_{\bar{s}_1 j}) p(\bar{s}_1) + \bar{b}_{\bar{s}_2 j} (a'_{\bar{s}_2 j} + a''_{\bar{s}_2 j}) p(\bar{s}_2) \\
 &= (\bar{b}'_{\bar{s}_1 j} - \bar{b}'_{\bar{s}_2 j}) a_{\bar{s}_1 j} p(\bar{s}_1) + (\bar{b}''_{\bar{s}_1 j} - \bar{b}''_{\bar{s}_2 j}) a_{\bar{s}_1 j} p(\bar{s}_1) \\
 &+ (\bar{b}_{\bar{s}_1 j} - \bar{b}_{\bar{s}_2 j}) a'_{\bar{s}_1 j} p(\bar{s}_1) + (\bar{b}_{\bar{s}_1 j} - \bar{b}_{\bar{s}_2 j}) a''_{\bar{s}_1 j} p(\bar{s}_1) \\
 &\neq 0.
 \end{aligned}$$

Let $j \in \bar{s}_1 = (s'_1, s_1)$ and $j \in \bar{s}_2 = (s'_2, s_2)$ such that $p(\bar{s}_1) > 0$, $p(\bar{s}_2) > 0$ and for a supposedly UMVE $\bar{e}_b$ in C_1 for a design in D_2 , namely $\bar{e}_{b\bar{s}} = \sum_{j \in \bar{s}} \bar{b}_{\bar{s}j} y_j + \sum_{j \in \bar{s}} \bar{b}'_{\bar{s}j} x_j + \sum_{j \in \bar{s}} \bar{b}''_{\bar{s}j} x_j$, $\bar{b}'_{\bar{s}_1 j} \neq \bar{b}'_{\bar{s}_2 j}$, $\bar{b}''_{\bar{s}_1 j} \neq \bar{b}''_{\bar{s}_2 j}$ and for a different unbiased estimator $e_{b\bar{s}}$, $\bar{b}'_{sk} = b'_{sk}$ and $\bar{b}''_{sk} = b''_{sk}$ for all $k \neq j$ and for all $\bar{s}$ except $\bar{s}_1$ and $\bar{s}_2$.

Then,

$$a'_{\bar{s}_1 j} p(\bar{s}_1) + a'_{\bar{s}_2 j} p(\bar{s}_2) = 0 \quad (2.3.14)$$

$$\text{and } a''_{\bar{s}_1 j} p_1(s'_1) + a''_{\bar{s}_2 j} p_1(s'_2) = 0 \quad (2.3.15)$$

But these contradict (2.3.3), because, using (2.3.14) and (2.3.15), we have

$$\begin{aligned}
 & \sum_1 \bar{b}_{sj}'' a_{sj}'' p_1(s') + \sum_1 \left\{ \sum_2 \bar{b}_{sj}' (a_{sj}' + a_{sj}'') p_2(s) \right\} p_1(s') \\
 & + \sum_1 \left\{ \sum_2 \bar{b}_{sj}'' a_{sj}' p_2(s) \right\} p_1(s') \\
 & = \bar{b}_{s_1j}'' a_{s_1j}'' p_1(s'_1) + \bar{b}_{s_2j}'' a_{s_2j}'' p_1(s'_2) + a_{s_1j}' (\bar{b}_{s_1j}' + \bar{b}_{s_1j}'') p(\bar{s}_1) \\
 & + a_{s_2j}' (\bar{b}_{s_2j}' + \bar{b}_{s_2j}'') p(\bar{s}_2) + \bar{b}_{s_1j}' a_{s_1j}'' p(\bar{s}_1) \\
 & + \bar{b}_{s_2j}' a_{s_2j}'' p(\bar{s}_2) \\
 & = (\bar{b}_{s_1j}'' - \bar{b}_{s_2j}'') a_{s_1j}'' p_1(s'_1) + [(\bar{b}_{s_1j}' - \bar{b}_{s_2j}') + (\bar{b}_{s_1j}'' - \bar{b}_{s_2j}'')] \\
 & \quad \cdot a_{s_1j}' p(\bar{s}_1) \\
 & + (\bar{b}_{s_1j}' - \bar{b}_{s_2j}') a_{s_1j}'' p(\bar{s}_1)
 \end{aligned}$$

≠ 0.

Finally let $j \in \bar{s}_1 = (s'_1, s_1)$ and $j \in \bar{s}_2 = (s'_2, s_2)$ such that $p(\bar{s}_1) > 0$, $p(\bar{s}_2) > 0$ and $\bar{b}_{s_1j} = \bar{b}_{s_2j}$, $\bar{b}_{s_1j}' = \bar{b}_{s_2j}'$

and $\bar{b}_{s_1 j}'' = \bar{b}_{s_2 j}''$, for a supposedly UMVE $\bar{e}_b$. Note that in such cases, $\bar{b}_{s j} = \bar{b}_j = \frac{1}{Q_j}$, $\bar{b}_{s j}' = \bar{b}_j' = -\frac{1}{Q_j}$ and $\bar{b}_{s j}'' = \bar{b}_j'' = \frac{1}{P_j} \forall j$ and $\forall \bar{s}$ with $p(\bar{s}) > 0$.

Let $k \in \bar{s}_1$ but $k \notin \bar{s}_2$.

Let an alternative $e_{b\bar{s}}$ be so chosen that $\bar{b}_{sk} = b_{sk}$, $\bar{b}_{sk}' = b_{sk}'$ and $\bar{b}_{sk}'' = b_{sk}''$ for all $k \neq j$ and for all $\bar{s}$ except $\bar{s}_1$ and $\bar{s}_2$ so that (2.3.11), (2.3.12) and (2.3.13) hold.

Then (2.3.2), (2.3.4) and (2.3.6) are also contradicted, because, we have

$$\begin{aligned} & \sum_3 \sum_4 \bar{b}_{s j} a_{s k} p_2(s) p_1(s') \\ &= \bar{b}_{s_1 j} a_{s_1 k} p(\bar{s}_1) = \bar{b}_j a_{s_1 k} p(\bar{s}_1) = \frac{1}{Q_j} a_{s_1 k} p(\bar{s}_1) \end{aligned}$$

$\neq 0$

$$\begin{aligned} & \sum_3 \bar{b}_{s j}'' a_{s k}'' p_1(s') + \sum_3 \left\{ \sum_4 \bar{b}_{s j}' a_{s k}' p_2(s) \right\} p_1(s') \\ &+ \sum_3 \left\{ \sum_2 (\bar{b}_{s j}' a_{s k}'' + a_{s j}' \bar{b}_{s k}'') p_2(s) \right\} p_1(s') \end{aligned}$$

$$\begin{aligned}
 &= \bar{b}_{s_1 j}'' a_{s_1 k}'' p_1(s_1') + \bar{b}_{s_1 j}' a_{s_1 k}' p(\bar{s}_1) \\
 &\quad + (\bar{b}_{s_1 j}' a_{s_1 k}'' + a_{s_1 j}' \bar{b}_{s_1 k}'') p(\bar{s}_1) \\
 &= \bar{b}_j'' a_{s_1 k}'' p_1(s_1') + \bar{b}_j' a_{s_1 k}' p(\bar{s}_1) \\
 &\quad + (\bar{b}_j' a_{s_1 k}'' + \bar{b}_k'' a_{s_1 j}') p(\bar{s}_1)
 \end{aligned}$$

* 0

$$\begin{aligned}
 &\Sigma_3 \Sigma_4 (\bar{b}_{sk}' a_{sj} + \bar{b}_{sj} a_{sk}') p_2(s) p_1(s') \\
 &+ \Sigma_3 \Sigma_2 (\bar{b}_{sk}'' a_{sj} + \bar{b}_{sj} a_{sk}'') p_2(s) p_1(s') \\
 &= (\bar{b}_{s_1 k}' a_{s_1 j} + \bar{b}_{s_1 j} a_{s_1 k}') p(\bar{s}_1) + (\bar{b}_{s_1 k}'' a_{s_1 j} + \bar{b}_{s_1 j} a_{s_1 k}'') p(\bar{s}_1) \\
 &= (\bar{b}_k' a_{s_1 j} + \bar{b}_j a_{s_1 k}' + \bar{b}_k'' a_{s_1 j} + \bar{b}_j a_{s_1 k}'') p(\bar{s}_1)
 \end{aligned}$$

* 0.

From these we conclude that there does not exist a UMV estimator for Y in C_1 for any design in D_2 i.e. in D outside D_1 .

Next, we investigate the properties of the estimator e_{bs}^* when it is based on any arbitrary design of the class D with $P_j > 0$, $Q_j > 0 \quad \forall j$.

Writing $Q_{jk} = \sum_3 \sum_4 p_2(s) p_1(s')$, $P_{jk} = \sum_3 p_1(s')$ for the inclusion - probabilities of pairs of units for two-phase and first-phase sampling and $R_{jk} = \sum_3 \sum_2 p_2(s) p_1(s')$ (with which we associate no simple physical meaning but which we consider because of their uses in the analysis to follow) and also $u_j = y_j - x_j$, the variance of e_{bs}^* , on algebraic simplification, may be written as

$$\begin{aligned} V_p(e_{bs}^*) &= \sum_{j=1}^N u_j^2 \left(\frac{1}{Q_j} - 1 \right) + \sum_{j \neq k=1}^N u_j u_k \left(\frac{Q_{jk}}{Q_j Q_k} - 1 \right) \\ &\quad + \sum_{j=1}^N x_j^2 \left(\frac{1}{P_j} - 1 \right) + \sum_{j \neq k=1}^N x_j x_k \left(\frac{P_{jk}}{P_j P_k} - 1 \right) \\ &\quad + 2 \sum_{j=1}^N x_j u_j \left(\frac{1}{P_j} - 1 \right) + 2 \sum_{j \neq k=1}^N x_k u_j \left(\frac{R_{jk}}{Q_j P_k} - 1 \right) \\ &= \sum_{j=1}^N y_j^2 \left(\frac{1}{Q_j} - 1 \right) + \sum_{j \neq k=1}^N u_j u_k \left(\frac{Q_{jk}}{Q_j Q_k} - 1 \right) \\ &\quad + \sum_{j=1}^N x_j^2 \left(\frac{1}{Q_j} - \frac{1}{P_j} \right) + \sum_{j \neq k=1}^N x_j x_k \left(\frac{P_{jk}}{P_j P_k} - 1 \right) \end{aligned}$$

$$+ 2 \sum_{j=1}^N x_j y_j \left(\frac{1}{P_j} - \frac{1}{Q_j} \right) + 2 \sum_{j+k=1}^N x_k u_j \left(\frac{R_{jk}}{Q_j P_k} - 1 \right).$$

Assuming $P_{jk} > 0$, $Q_{jk} > 0$ and $R_{jk} > 0 \forall j \neq k$, we have two unbiased estimators for it; viz.

$$\begin{aligned} v_1 = & \sum_{j \in s} u_j^2 \left(\frac{1 - Q_j}{Q_j^2} \right) + \sum_{j \neq k \in s} u_j u_k \left(\frac{Q_{jk} - Q_j Q_k}{Q_{jk} Q_j Q_k} \right) \\ & + \sum_{j \in s} x_j^2 \left(\frac{1 - P_j}{P_j^2} \right) + \sum_{j \neq k \in s} x_j x_k \left(\frac{P_{jk} - P_j P_k}{P_{jk} P_j P_k} \right) \\ & + 2 \sum_{j \in s} x_j u_j \frac{(1 - P_j)}{P_j Q_j} + 2 \sum_{\substack{j \neq k \\ j \in s, k \in s}} x_k u_j \frac{(R_{jk} - Q_j P_k)}{R_{jk} Q_j P_k} \end{aligned}$$

- (unbiased for any sampling design p)

$$\begin{aligned} v_2 = & \sum_{j < k \in s} \frac{(Q_j Q_k - Q_{jk})}{Q_{jk}} \left(\frac{u_j}{Q_j} - \frac{u_k}{Q_k} \right)^2 \\ & + \sum_{j < k \in s} \frac{(P_j P_k - P_{jk})}{P_{jk}} \left(\frac{x_j}{P_j} - \frac{x_k}{P_k} \right)^2 \\ & + 2 \sum_{j \in s} x_j u_j \frac{(1 - P_j)}{P_j Q_j} + 2 \sum_{\substack{j \neq k \\ j \in s, k \in s}} x_k u_j \left(\frac{R_{jk} - Q_j P_k}{R_{jk} Q_j P_k} \right) \end{aligned}$$

- (unbiased for any fixed-sample-size design).

For any $e_{b\bar{s}}$ in C_1 again the variance is

$$\begin{aligned}
 v_p(e_{b\bar{s}}) &= \sum_{j=1}^N y_j^2 \left[\sum_1 \sum_2 b_{\bar{s}j}^2 p_2(s) p_1(s') - 1 \right] \\
 &+ \sum_{j \neq k=1}^N y_j y_k \left[\sum_3 \sum_4 b_{\bar{s}j} b_{\bar{s}k} p_2(s) p_1(s') - 1 \right] \\
 &+ \sum_{j=1}^N x_j^2 \left[\sum_1 \sum_2 b_{\bar{s}j}^{\prime 2} p_2(s) p_1(s') + \sum_1 b_{\bar{s}j}^{\prime \prime 2} p_1(s') \right. \\
 &+ 2 \sum_1 \sum_2 b_{\bar{s}j}^{\prime} b_{\bar{s}j}^{\prime \prime} p_2(s) p_1(s') \left. \right] \\
 &+ \sum_{j \neq k=1}^N x_j x_k \left[\sum_3 \sum_4 b_{\bar{s}j}^{\prime} b_{\bar{s}k}^{\prime} p_2(s) p_1(s') \right. \\
 &+ \sum_3 b_{\bar{s}j}^{\prime \prime} b_{\bar{s}k}^{\prime \prime} p_1(s') + 2 \sum_3 \sum_2 b_{\bar{s}j}^{\prime} b_{\bar{s}k}^{\prime \prime} p_2(s) p_1(s') \left. \right] \\
 &+ 2 \sum_{j=1}^N x_j y_j \left[\sum_1 \sum_2 b_{\bar{s}j} b_{\bar{s}j}^{\prime} p_2(s) p_1(s') \right. \\
 &+ \sum_1 \sum_2 b_{\bar{s}j} b_{\bar{s}j}^{\prime \prime} p_2(s) p_1(s') \left. \right] \\
 &+ 2 \sum_{j \neq k=1}^N x_k y_j \left[\sum_3 \sum_4 b_{\bar{s}j} b_{\bar{s}k}^{\prime} p_2(s) p_1(s') \right. \\
 &+ \sum_3 \sum_2 b_{\bar{s}j} b_{\bar{s}k}^{\prime \prime} p_2(s) p_1(s') \left. \right].
 \end{aligned}$$

For unbiasedness of $c_{b\bar{s}}$, necessary and sufficient conditions are

$$\sum_1 \sum_2 b_{\bar{s}j} p_2(s) p_1(s') = 1 \quad \forall j$$

$$\sum_1 \sum_2 b_{sj}^1 p_2(s) p_1(s') = -1 \quad \forall j$$

$$\text{and } \sum_1 b_{sj}'' p_1(s') = 1 \quad \forall j.$$

The last condition may be written as

$$\sum_1 \sum_2 b_{sj}'' p_2(s) p_1(s') + \sum_1 \sum_5 b_{sj}'' p_2(s) p_1(s') = 1 \quad \forall j.$$

Now by Cauchy - Schwarz's inequality, we get

$$\begin{aligned} \sum_1 \sum_2 b_{sj}^2 p_2(s) p_1(s') &\geq \frac{[\sum_1 \sum_2 b_{sj} p_2(s) p_1(s')]^2}{\sum_1 \sum_2 p_2(s) p_1(s')} \\ &= \frac{1}{Q_j} \quad \forall j \end{aligned} \quad (2.3.16)$$

Also

$$\begin{aligned} &\sum_1 \sum_2 b_{sj}^1 p_2(s) p_1(s') + \sum_1 b_{sj}''^2 p_1(s') \\ &\quad + 2 \sum_1 \sum_2 b_{sj}^1 b_{sj}'' p_2(s) p_1(s') \\ &= \sum_1 \sum_2 (b_{sj}^1 + b_{sj}'')^2 p_2(s) p_1(s') + \sum_1 \sum_5 b_{sj}''^2 p_2(s) p_1(s') \end{aligned}$$

$$\geq \frac{\int \Sigma_1 \Sigma_2 (b_{sj}^I + b_{sj}^n) p_2(s) p_1(s') J^2}{\Sigma_1 \Sigma_2 p_2(s) p_1(s')} + \frac{\int \Sigma_1 \Sigma_5 b_{sj}^n p_2(s) p_1(s') J^2}{\Sigma_1 \Sigma_5 p_2(s) p_1(s')}$$

(by Cauchy-Schwarz's inequality)

$$= \frac{\int^{-1+1-\Sigma_1 \Sigma_5} b_{sj}^n p_2(s) p_1(s') J^2}{Q_j} + \frac{\int \Sigma_1 \Sigma_5 b_{sj}^n p_2(s) p_1(s') J^2}{P_j - Q_j}$$

$$= \int \Sigma_1 \Sigma_5 b_{sj}^n p_2(s) p_1(s') J^2 \int \frac{1}{Q_j} + \frac{1}{P_j - Q_j} J.$$

Again, by Cauchy-Schwarz's inequality, we have

$$\begin{aligned} & \Sigma_1 \Sigma_5 b_{sj}^n p_2(s) p_1(s') \\ & \geq \frac{\int \Sigma_1 \Sigma_5 \sqrt{b_{sj}^n} p_2(s) p_1(s') J^2}{\Sigma_1 \Sigma_5 p_2(s) p_1(s')} \end{aligned}$$

Now this becomes an equality when

$$b_{sj}^n = \frac{1}{P_j}, \text{ in which case the right-hand side expression}$$

$$\text{becomes, on simplification, } \frac{P_j - Q_j}{P_j}$$

$$\text{so that } \Sigma_1 \Sigma_5 b_{sj}^n p_2(s) p_1(s') \geq \frac{P_j - Q_j}{P_j}.$$

Hence we have

$$\begin{aligned}
 & \Sigma_1 \Sigma_2 b_{sj}^{'2} p_2(s) p_1(s') + \Sigma_1 b_{sj}^{'2} p_1(s') \\
 & \quad + 2 \Sigma_1 \Sigma_2 b_{sj}' b_{sj}'' p_2(s) p_1(s') \\
 & \geq \left[\frac{p_j - q_j}{p_j} \right]^2 \left[\frac{1}{q_j} + \frac{1}{p_j - q_j} \right] = \frac{1}{q_j} - \frac{1}{p_j} \quad (2.3.17)
 \end{aligned}$$

$$\begin{aligned}
 & \text{Finally, } \Sigma_1 \Sigma_2 b_{sj}' b_{sj}' p_2(s) p_1(s') \\
 & \quad + \Sigma_1 \Sigma_2 b_{sj}'' b_{sj}'' p_2(s) p_1(s') \\
 & = \Sigma_1 \Sigma_2 b_{sj} (b_{sj}' + b_{sj}'') p_2(s) p_1(s') \\
 & \geq \frac{\left[\Sigma_1 \Sigma_2 \sqrt{b_{sj}'} \sqrt{b_{sj}'} + b_{sj}''} p_2(s) p_1(s') \right]^2}{\Sigma_1 \Sigma_2 p_2(s) p_1(s')}
 \end{aligned}$$

by Cauchy-Schwarz's inequality so long as applicable.

However, this becomes an equality when

$$b_{sj} = \frac{1}{q_j}, \quad b_{sj}' = -\frac{1}{q_j} \quad \text{and} \quad b_{sj}'' = \frac{1}{p_j} \quad \forall j,$$

in which case the right-hand side becomes, on simplification,

$$\frac{1}{p_j} - \frac{1}{q_j} \quad (\text{may be negative, of course}).$$

Hence it follows that

$$\begin{aligned} \sum_1 \sum_2 b_{sj}^- b_{sj}^+ p_2(s) p_1(s') + \sum_1 \sum_2 b_{sj}^- b_{sj}^+ p_2(s) p_1(s') \\ \geq \frac{1}{P_j} - \frac{1}{Q_j} . \end{aligned} \quad (2.3.18)$$

If $\underline{Y}$ and $\underline{X}$ be such that $y_j \neq 0$, $x_j \neq 0$, x_j and y_j are of the same sign but $y_k = 0 \ \forall \ k \neq j$ and $x_k = 0 \ \forall \ k \neq j$, so that we write such a point $(\underline{Y}, \underline{X})$ as $(\underline{Y}^{(j)}, \underline{X}^{(j)}) = \underline{Z}^{(j)}$, then at such a point

$$V_p(e_{bs}) - V_p(e_{bs}^*) = y_j^2 a_j + x_j^2 b_j + 2 x_j y_j c_j \geq 0 \quad (2.3.19)$$

because $a_j, b_j, c_j \geq 0 \ \forall \ j$ from above.

For e_{bs} different from e_{bs}^* for at least one sample with $p(\bar{s}) > 0$, at least one of a_j, b_j, c_j is strictly positive for at least one j so that (2.3.19) is a strict inequality at least for one j and for one $\bar{s}$ with $p(\bar{s}) > 0$. Hence we have the

Theorem 2.3.3. In the class C_1 of estimators for Y for any design in the class D , the estimator e_{bs}^* is admissible (in the usual sense in survey sampling theory).

However if the estimators for Y , $-X$ and X as above are permitted to be any unbiased estimator (not necessarily of the HLU form) then an admissible estimator is available under severe restrictions on the class of such estimators and also on the class of designs p . The restrictions are motivated by the consideration that the estimators for Y and $-X$ based on second-phase sample alone must separately be admissible as estimators respectively for well-known admissible estimators for Y and $-X$ that might be based on y_j and x_j values for $j \in s'$ if y_j , x_j , $j \in s'$ were available, supposing that the estimator for X based on x_j , $j \in s'$ is also a usual admissible (based on uniphase sampling) unbiased estimator for X .

For Y as before we shall consider unbiased estimators of the form $e = e(\bar{s}, y, x') = e(s, s', y_1, x_1; i \in s, j \in s')$ with $s \subset s'$, the sample $\bar{s} = (s', s)$ being chosen according to a design p of the class D with

$$P_j = \sum_1 p_1(s') > 0 \quad \forall j \quad \text{and}$$

$$Q_j = \sum_1 \sum_2 p_2(s) p_1(s') > 0 \quad \forall j.$$

But we shall impose further restrictions on the estimators e and the designs p to suit our purpose as described below.

Here in the first phase of sampling we are estimating **X** and the estimator concerned should be chosen in an optimal way. So the estimator may be suggested as

$$e_2^* (s', x') = e_2^* (s', x_j \mid j \in s') = \sum_{j \in s'} \frac{x_j}{P_j},$$

the usual HT estimator known to have many optimal properties.

Obviously, if the intention was to estimate $Y-X$ and in the first-phase sample, data were available as $(s', y_j, x_j \mid j \in s')$ one would be inclined to use $e_2^* (s', y_j, x_j \mid j \in s')$

$$= \sum_{j \in s'} \frac{y_j - x_j}{P_j} \text{ to estimate } Y-X \text{ and hence the estimator}$$

$$e_2^* (s', y_j \mid j \in s') = \sum_{j \in s'} \frac{y_j}{P_j} \text{ to estimate } Y \text{ itself. But}$$

the requisite data being unavailable and the second-phase sample data being realized as $(s, y_i, x_i \mid i \in s)$ one may intend to estimate $Y-X$ after having obtained an estimator

$$e_2^* (s', x') = \sum_{j \in s'} \frac{x_j}{P_j} \text{ for } X, \text{ so as to get a pooled}$$

estimate for Y . So, one should be inclined to choose an

estimator $e_1 = e_1(s, y_i, x_i \mid i \in s)$ for $(Y-X)$ from second phase sample data such that it should estimate $\sum_{j \in s'} \frac{y_j - x_j}{P_j}$

first in an appropriate way because the latter would estimate

(Y-X) optimally if the detailed first phase sample data were available as $d' = d(s') = (s', y_j, x_j | j \in s')$. Such an estimator should have a conditional expectation, given d' , as $\sum_{j \in s'} \frac{y_j - x_j}{p_j}$ and it should be of the form

$$\sum_{j \in s} \frac{y_j - x_j}{p_j} \cdot \frac{1}{\sum_2 p_2(s)} \quad \text{so as to be an HT estimator}$$

based on second phase sample data for $\sum_{j \in s'} \frac{y_j - x_j}{p_j}$.

This would take the form

$$e_1^* = e_1^*(s, y, x) = \sum_{j \in s} \frac{y_j - x_j}{Q_j}, \quad \text{provided we}$$

impose, as we do from now on, the condition on the design p that

$$\sum_2 p_2(s) = \frac{Q_j}{p_j}, \quad j = 1, \dots, N \quad (2.3.20)$$

(vide an example in Appendix 1).

The motivation for this is eventually to establish the admissibility of e^* . The optimum estimator for Y thus suggested is clearly

$$e^* = e^*(\bar{s}, y, x') = e_1^*(s, y, x) + e_2^*(s', x').$$

$$\text{Here } e_2^*(s', x') = \sum_{j \in s'} \frac{x_j}{p_j}; \quad \text{also}$$

$$e_2^*(s', y') = \sum_{j \in s'} \frac{y_j}{p_j}, \quad e_2^*(s', y', x') = \sum_{j \in s'} \frac{y_j - x_j}{p_j},$$

$$e_1^*(s, y, x) = \sum_{j \in s} \frac{y_j - x_j}{q_j} \quad \text{such that}$$

$$E [e_1^*(s, y, x) | d' = d(s')] = \sum_{s \subset s'} \left(\sum_{j \in s} \frac{y_j - x_j}{q_j} \right) p_2(s)$$

$$= \sum_{j \in s'} \frac{y_j - x_j}{q_j} \sum_2 p_2(s) = \sum_{j \in s'} \frac{y_j - x_j}{p_j}$$

$$= e_2^*(s', y', x') = e_2^*(s', y') - e_2^*(s', x')$$

$$\text{so that } E [e^*(\bar{s}, y, x') | d' = d(s')] = e_2^*(s', y')$$

$$\text{and that } E [e^*(\bar{s}, y, x')] = E [e_2^*(s', y')] = Y.$$

From these considerations, we shall consider the following form of estimators for Y based on a design p subject to (2.3.20), namely,

$$e = e(\bar{s}, y, x') = e(s, s', y_1, x_1 | \text{ies}, j \in s')$$

$$= e_1(s, y, x) + e_2(s', x')$$

$$\text{with } e_1 = e_1(s, y, x) = e_1^*(s, y, x) + h_1(s, y, x)$$

$$e_2 = e_2(s', x') = e_2^*(s', x') + h_2(s', x')$$

such that $E [e_2(s', x')] = X$

and $E [e(\bar{s}, y, x') | d' = d(s')] = e_2^*(s', y')$,

the latter being suggested by the conditional argument put forward as above. Then, h_1, h_2 are required to satisfy the condition

$$E [h_1(s, y, x) + h_2(s', x') | d' = d(s')] = 0$$

$$\forall s' \text{ with } p_1(s') > 0.$$

Because of this we shall write

$$h_1(s, y, x) + h_2(s', x') = h(\bar{s}, y, x')$$

such that

$$E[h(\bar{s}, y, x') | d' = d(s')] = \sum_{s \in s'} h(\bar{s}, y, x') p_2(s) = 0 \quad (2.3.21)$$

This condition is reasonable because we note that

$$E(e_1 | d') = e_2^*(s', y') - e_2^*(s', x') + E(h_1 | d')$$

$$\text{so that } E(e | d') = e_2^*(s', y') - e_2^*(s', x') + E(h_1 | d')$$

$$+ e_2^*(s', x') + h_2(s', x')$$

$$= e_2^*(s', y') + E(h_1 | d') + h_2(s', x')$$

which should equal $e_2^*(s', y')$ because given the data d' , an optimal estimator for Y is $e_2^*(s', y') = \sum_{j \in s'} \frac{y_j}{P_j}$, as is known from Godambe/Joshi's (1965) results.

So, we impose the condition (2.3.21) on the class of estimators for Y considered above. When (2.3.20) and (2.3.21) are imposed we say that our strategy (e, p) belongs to a class C^* . Now we have

Theorem 2.3.4. In the class of estimators e for any design p such that (e, p) belongs to the class C^* of strategies, the estimator e^* is admissible.

Proof. We observe that for

$$e = e(\bar{s}, y, x') = e_1^*(s, y, x) + e_2^*(s', x') \\ + h_1(s, y, x) + h_2(s', x')$$

$$V(e) = V_{P_1} \left(\sum_{j \in s'} \frac{y_j}{P_j} \right) + E_{P_1} V_{P_2} \left[\sum_{j \in s} \frac{y_j - x_j}{Q_j} + h \right] | d'$$

$$\text{and } V(e^*) = V_{P_1} \left(\sum_{j \in s'} \frac{y_j}{P_j} \right) + E_{P_1} V_{P_2} \left[\sum_{j \in s} \frac{y_j - x_j}{Q_j} \right] | d'.$$

Given the sample s' and the data $d' = d(s')$ both $\sum_{j \in s} \frac{y_j - x_j}{Q_j}$

and $\sum_{j \in s} \frac{y_j - x_j}{Q_j} + h$ are unbiased estimators for $\sum_{j \in s'} \frac{y_j - x_j}{P_j}$

with respect to the design p_2 . But from Godambe/Joshi

(1965) it is known that among the class of all unbiased estimators of $\sum_{j \in s'} \frac{y_j - x_j}{P_j}$ the estimator $\sum_{j \in s} \frac{y_j - x_j}{P_j} \cdot \frac{1}{\sum_2 p_2(s)}$

$= \sum_{j \in s} \frac{y_j - x_j}{Q_j}$ is admissible for it (for every given s'). So

for no h such that $E_{p_2}(h | s') = 0 \quad \forall s' \text{ with } p_1(s') > 0$,

$V_{p_2} \left[\sum_{j \in s} \frac{y_j - x_j}{Q_j} + h \right] | d' J$ can be uniformly less than

$V_{p_2} \left[\sum_{j \in s} \frac{y_j - x_j}{Q_j} | d' J \right]$ and strictly so for at least one

parametric point $(y_1 - x_1, \dots, y_N - x_N)$. Consequently the same conclusion follows concerning $V(e)$ versus $V(e^*)$ implying admissibility of e^* as an estimator for Y . Q.E.D.

Let D^* be a class of designs p such that $\sum_{s'} n(s') p_1(s') =$

n' , $\sum_{s'} \sum_s n(s) p_2(s) p_1(s') = n$ with $n, n' (n < n' < N)$

as pre-assigned positive integers. Let D^{**} denote the class

of sampling designs p in D^* subject to the additional

restriction (2.3.20). Then, we have,

Theorem 2.3.5. Among strategies in C^* with p in D^{**} , the strategy (e^*, p^*) with p^* in D^{**} is uniformly admissible.

Proof. Suppose not. Then there must exist (e, p) with $e \in C^*$ and $p \in D^{**}$ dominating it (in the sense that

$$V_p(e) \leq V_{p^*}(e^*) \quad (2.3.22)$$

$\forall (X, Y)$ and strictly so for at least one).

For $e \in C^*$ and $p \in D^{**}$,

$$V_p(\cdot) = V_{p_1}(\sum_{j \in S} \frac{y_j}{P_j}) + E_{p_1} V_{p_2}(\sum_{j \in S} \frac{y_j - x_j}{Q_j} + h \mid d' = d(s'))$$

$$V_{p^*}(e^*) = V_{p_1^*}(\sum_{j \in S} \frac{y_j}{P_j^*}) + E_{p_1^*} V_{p_2^*}(\sum_{j \in S} \frac{y_j - x_j}{Q_j^*} \mid d' = d(s'))$$

writing P_j^* and Q_j^* for the inclusion - probabilities for the sampling design p^* (similarly keeping a provision to use the symbols P_{jk}^* , Q_{jk}^* in what follows with obvious meanings).

Let, it possible, $V_{p_1}(\sum_{j \in S} \frac{y_j}{P_j}) \leq V_{p_1^*}(\sum_{j \in S} \frac{y_j}{P_j^*}) \quad \forall \underline{Y}$.

Noting that $\sum_{j \in S} \frac{y_j}{P_j}$ and $\sum_{j \in S} \frac{y_j}{P_j^*}$ are HT estimators based

on designs p_1 and p_1^* respectively we observe it to follow that

$$V_{p_1} \left(\sum_{j \in s} \frac{y_j}{p_j} \right) = V_{p_1^*} \left(\sum_{j \in s} \frac{y_j}{p_j^*} \right),$$

vide Ramakrishnan (1975) and also that $p_j = p_j^* \quad \forall j$,

$p_{jk} = p_{jk}^* \quad \forall j \neq k$. Now given s' and the data $d' = d(s')$,

$\sum_{j \in s} \frac{y_j - x_j}{q_j}$ is a Horvitz / Thompson (1952) estimator for

$\sum_{j \in s} \frac{y_j - x_j}{p_j}$ and $\sum_{j \in s} \frac{y_j - x_j}{q_j} + h$ is a different unbiased estimator

for $\sum_{j \in s} \frac{y_j - x_j}{p_j}$. So from Godambe / Joshi (1965) it follows

that for no h such that $E_{p_2}(h | s') = 0 \quad \forall s'$ with

$p_1(s') > 0$, $V_{p_2} \left(\sum_{j \in s} \frac{y_j - x_j}{q_j} + h \mid d' \right)$ can be uniformly

less than $V_{p_2} \left(\sum_{j \in s} \frac{y_j - x_j}{q_j} \mid d' \right)$ and strictly so for at

least one parametric point $(y_1 - x_1, \dots, y_N - x_N)$.

Now $E_{p_1} V_{p_2} \left(\sum_{j \in s} \frac{y_j - x_j}{q_j} \mid d' = d(s') \right)$

$$= \sum_{j=1}^N u_j^2 \left(\frac{1}{q_j} - \frac{1}{p_j} \right) + \sum_{j+k=1}^N u_j u_k \left(\frac{q_{jk}}{q_j q_k} - \frac{p_{jk}}{p_j p_k} \right)$$

$$\text{and } E_{p_1^*} V_{p_2^*} \left(\sum_{j \in S} \frac{y_j - x_j}{Q_j^*} \mid d' = d(s') \right)$$

$$= \sum_{j=1}^N u_j^2 \left(\frac{1}{Q_j^*} - \frac{1}{P_j^*} \right) + \sum_{j \neq k=1}^N u_j u_k \left(\frac{Q_{jk}^*}{Q_j^* Q_k^*} - \frac{P_{jk}^*}{P_j^* P_k^*} \right)$$

$$\text{where } u_j = y_j - x_j.$$

$$\text{Let if possible, } E_{p_1^*} V_{p_2^*} \left(\sum_{j \in S} \frac{y_j - x_j}{Q_j^*} \mid d' = d(s') \right)$$

$$\leq E_{p_1^*} V_{p_2^*} \left(\sum_{j \in S} \frac{y_j - x_j}{Q_j^*} \mid d' = d(s') \right) \quad (2.3.23)$$

$$\Rightarrow \sum_{j=1}^N u_j^2 \left(\frac{1}{Q_j^*} - \frac{1}{P_j^*} \right) + \sum_{j \neq k=1}^N u_j u_k \left(\frac{Q_{jk}^*}{Q_j^* Q_k^*} - \frac{P_{jk}^*}{P_j^* P_k^*} \right) \leq 0.$$

Choosing $u_j \neq 0$, $u_k = 0 \quad \forall k \neq j$ we get

$$u_j^2 \left(\frac{1}{Q_j^*} - \frac{1}{P_j^*} \right) \leq 0$$

$$\Rightarrow Q_j^* = P_j^* \quad \forall j.$$

Again choosing $u_j \neq 0$, $u_k \neq 0$, $u_l = 0 \quad \forall l \neq j, k$,

we get

$$u_j u_k \left(\frac{Q_{jk}^*}{Q_j^* Q_k^*} - \frac{P_{jk}^*}{P_j^* P_k^*} \right) \leq 0$$

$$\implies Q_{jk} = Q_{jk}^* \quad \forall \quad j \neq k.$$

So strict inequality cannot hold in (2.3.23).

So $V_p(e)$ cannot be uniformly less than $V_{p^*}(e^*)$ and hence the theorem is proved.

Next we consider the class C' of estimators

$$e = \sum_{j \in S} b_{sj} (y_j - x_j) + \sum_{j \in S'} b'_{sj} x_j$$

and designs in the class D' such that for every p in D' ,

$$p_1(s') > 0 \implies n(s') = n', \quad p_2(s) > 0 \implies n(s) = n;$$

and P_j, Q_j 's are pre-assigned.

For a specific subclass of C' namely

$$e^* = \sum_{j \in S} b_j (y_j - x_j) + \sum_{j \in S'} b'_j x_j \quad \text{with} \quad b_j, b'_j > 1 \quad \forall j,$$

$$\sum_{j=1}^N \frac{1}{b_j} = n, \quad \sum_{j=1}^N \frac{1}{b'_j} = n' \quad \text{and any design } p^* \text{ in } D',$$

(with $p^*(\bar{s}) = p_1^*(s') p_2^*(s/s') = p_1^*(s') p_2^*(s)$, say) we

have following Sekkappan (1973) and Sekkappan/Thompson (1975)

Theorem 2.3.6. Among all strategies (e, p) with e in C' , p in D' , any strategy (e^*, p^*) is 'uniformly admissible'.

[Remark I. As usual (vide, e.g., Godambe (1969)) a strategy is 'uniformly admissible' with loss as the square error i.e. $(e - Y)^2$, if the value of the design-average of the loss for no other strategy with the designs in the same class is uniformly (i.e. for all vectors $(\underline{X}, \underline{Y})$ with $\underline{X} = (x_1, \dots, x_N)$, $\underline{Y} = (y_1, \dots, y_N)$) less or equal and strictly less for at least one vector-point. In case the design is fixed and one considers only contending estimators, the concept is one of 'admissibility']

Proof. Suppose not. Then a strategy (e, p) with p in D' , say,

$$e = \sum_{j \in S} p_{sj} (y_j - x_j) + \sum_{j \in S'} p'_{sj} x_j \text{ in } C',$$

must exist such that for every $(\underline{X}, \underline{Y})$ in $R^N \times R^N$,

$$\sum_{s'} \sum_s \left[\sum_{j \in S} p_{sj} u_j + \sum_{j \in S'} p'_{sj} x_j - \sum_{j=1}^N u_j - \sum_{j=1}^N x_j \right]^2 p_2(s) p_1(s')$$

$$\leq \sum_{s'} \sum_s \left[\sum_{j \in S} b_{sj} u_j + \sum_{j \in S'} b'_{sj} x_j - \sum_{j=1}^N u_j - \sum_{j=1}^N x_j \right]^2 p_2^*(s) p_1^*(s')$$

(2.3.24)

(and strictly so for at least one $(\underline{X}, \underline{Y})$),

$$\text{or, } \sum_{s'} \sum_s \left[\sum_{j \in s} \frac{\rho_{sj}}{b_j} r_j + \sum_{j \in s'} \frac{\rho'_{sj}}{b'_j} t_j - \sum_{j=1}^N r_j/b_j - \sum_{j=1}^N \frac{t_j}{b'_j} \right]^2$$

$$\leq \sum_{s'} \sum_s \left[\sum_{j \in s} r_j + \sum_{j \in s'} t_j - \sum_{j=1}^N r_j/b_j - \sum_{j=1}^N t_j/b'_j \right]^2 p_2^*(s) p_1^*(s')$$

(writing $u_j = y_j - x_j$, $\underline{u} = (u_1, \dots, u_N)$, $r_j = b_j u_j$

and $t_j = b'_j x_j$)

or (on simplifications)

$$\sum_1 \sum_2 \frac{\rho_{sj}^2}{b_j^2} p_2(s) p_1(s') - 2 \sum_1 \sum_2 \frac{\rho_{sj}}{b_j^2} p_2(s) p_1(s')$$

$$- Q_j + 2Q_j/b_j \int \leq 0 \quad \forall j$$

$$\sum_1 \frac{\rho'_{sj}}{b_j^2} p_1(s') - 2 \sum_1 \frac{\rho'_{sj}}{b_j^2} p_1(s') - P_j + 2P_j/b'_j \int \leq 0 \quad \forall j.$$

Writing

$$\alpha_{sj} = \rho_{sj}/b_j - 1, \text{ for } j \in s$$

$$\delta_j = \sum_1 \sum_2 \alpha_{sj} p_2(s) p_1(s') \text{ and noting}$$

$$\delta_j + Q_j = \sum_1 \sum_2 \frac{\rho_{sj}}{b_j} p_2(s) p_1(s')$$

one gets

$$\sum_1 \sum_2 \alpha_{sj}^2 p_2(s) p_1(s') + 2 \delta_j \left(1 - \frac{1}{b_j}\right) \leq 0$$

$$\text{implying } \delta_j \leq 0 \quad \forall j.$$

For the choice $x_j = 0$ and $y_j = \frac{1}{b_j} \quad \forall j$, one has

$$\sum_{s'} \sum_s \left[\sum_{j \in s} \alpha_{sj} \right]^2 p_2(s) p_1(s') \leq 0$$

$$\Rightarrow \sum_{j \in s} \alpha_{sj} = 0 \quad \forall s \Rightarrow \sum_{s'} \sum_s \left[\sum_{j \in s} \alpha_{sj} \right] p_2(s) p_1(s') = 0$$

$$\Rightarrow \sum_{j=1}^N \delta_j = 0 \Rightarrow \delta_j = 0 \quad \forall j$$

$$\Rightarrow \alpha_{sj} = 0, \quad \beta_{sj} = b_j \quad \forall s \quad \text{with } p(\bar{s}) > 0.$$

$$\text{Similarly writing } \alpha'_{sj} = \beta'_{sj} / b'_j - 1, \quad \delta'_j = \sum_1 \alpha'_{sj} p_1(s')$$

$$\text{one gets } \alpha'_{sj} = 0 \quad \forall s' \quad \text{with } p_1(s') > 0 \Rightarrow \beta'_{sj} = b'_j \quad \forall s'$$

$$\text{with } p_1(s') > 0 \Rightarrow e = e^* \quad \forall \bar{s} \quad \text{with } p(\bar{s}) > 0$$

$$\Rightarrow \sum_{s'} \sum_s \left[\sum_{j \in s} b_j u_j + \sum_{j \in s'} b'_j x_j - \sum_{j=1}^N u_j - \sum_{j=1}^N x_j \right]^2 p_2(s) p_1(s')$$

$$\leq \sum_{s'} \sum_s \left[\sum_{j \in s} b_j u_j + \sum_{j \in s'} b'_j x_j - \sum_{j=1}^N u_j - \sum_{j=1}^N x_j \right]^2 p_2^*(s) p_1^*(s')$$

whence on noting $\sum_1 p_1(s') = P_j = \sum_1 p_1^*(s')$,

$$\sum_1 \sum_2 p_2(s) p_1(s') = Q_j = \sum_1 \sum_2 p_2^*(s) p_1^*(s') \text{ and}$$

$$\text{writing } Q_{jk}^* = \sum_3 \sum_4 p_2^*(s) p_1^*(s'), \quad P_{jk}^* = \sum_3 p_1^*(s')$$

$$\text{and } R_{jk}^* = \sum_3 \sum_2 p_2^*(s) p_1^*(s')$$

one gets

$$\begin{aligned} & \sum_{j \neq k=1}^N \sum b_j b_k u_j u_k Q_{jk} + \sum_{j \neq k=1}^N \sum b_j' b_k' x_j x_k P_{jk} \\ & + 2 \sum_{j \neq k=1}^N \sum b_j b_k' u_j x_k R_{jk} \leq \sum_{j \neq k=1}^N \sum b_j b_k u_j u_k Q_{jk}^* \\ & + \sum_{j \neq k=1}^N \sum b_j' b_k' x_j x_k P_{jk}^* + 2 \sum_{j \neq k=1}^N \sum b_j b_k' u_j x_k R_{jk}^* . \end{aligned}$$

Choosing $x_j = 0 \quad \forall j$, $y_j \neq 0$, $y_k \neq 0$, $y_l = 0 \quad \forall l \neq j, k$,

this yields

$$\begin{aligned} b_j b_k y_j y_k Q_{jk} & \leq b_j b_k y_j y_k Q_{jk}^* \\ \implies Q_{jk} & = Q_{jk}^* \quad \forall j, k \quad (j \neq k). \end{aligned}$$

Choosing $x_j = y_j \quad \forall j$, $x_j \neq 0$, $x_k \neq 0$, $x_l = 0 \quad \forall l \neq j, k$,

one gets

$$b_j' b_k' x_j x_k P_{jk} \leq b_j' b_k' x_j x_k P_{jk}^*$$

$$\implies P_{jk} = P_{jk}^* \quad \forall j, k \quad (j \neq k).$$

choosing $y_j \neq 0$, $y_k = 0 \quad \forall k \neq j$ and $x_k \neq 0$, $x_l = 0 \quad \forall l \neq k$, one gets

$$b_j' b_k' y_j x_k R_{jk} \leq b_j' b_k' y_j x_k R_{jk}^*$$

$$\implies R_{jk} = R_{jk}^* \quad \forall j, k \quad (j \neq k)$$

$\implies$ (2.3.24) cannot be strict $\implies$ contradiction and hence proof.

Remark II. We are unable to obtain a corresponding result with a wider class of estimators of the form

$$e_{b\bar{s}} = \sum_{j \in s} b_{sj} y_j + \sum_{j \in s} b'_{sj} x_j + \sum_{j \in s'} b''_{sj} x_j.$$

A useful example of (e^*, p^*) is one with

$$e^* = \sum_{j \in s} \frac{u_j}{Q_j} + \sum_{j \in s'} \frac{x_j}{P_j}.$$

Let C'' be a class of unbiased estimators (for Y) :

$$e = e(\bar{s}, \underline{Y}, \underline{X}) = e_1(s, \underline{U}) + e_2(s', \underline{X}) = e_1 + e_2 \text{ (say)}$$

such that $E_p(e_1) = Y - X$ and $E_p(e_2) = X$ where

$$\underline{X} = (x_1, \dots, x_N), \quad \underline{Y} = (y_1, \dots, y_N) \text{ and } \underline{U} = (u_1, \dots, u_N).$$

In particular, $e^* = e^*(\bar{s}, \underline{Y}, \underline{X}) = e_1^*(s, \underline{U}) + e_2^*(s', \underline{X})$

$$= \sum_{j \in s} \frac{u_j}{Q_j} + \sum_{j \in s'} \frac{x_j}{P_j} \text{ (say) is inside } C''.$$

Theorem 2.3.7. In C'' , e^* is 'admissible' for any chosen design p .

Proof. If $x_j = y_j \quad \forall j$, then for any e in C'' ,

$$V(e) - V(e^*) = V \left[e_2(s', \underline{X}) \right] - V \left[\sum_{j \in s'} \frac{x_j}{P_j} \right].$$

Now from Godambe / Joshi (1965) it follows that

$$V \left[e_2(s', \underline{X}) \right] > V \left[\sum_{j \in s'} \frac{x_j}{P_j} \right]$$

for at least one $\underline{X}$ in R^N (since $\sum_{j \in s'} \frac{x_j}{P_j}$ is 'admissible' in the class of all unbiased estimators for X for any design).

$\implies e^*$ is admissible in C'' Q.E.D.

A possibly better proof is given in Appendix 2.

Following Ramakrishnan (1975) one more optimal property for e^* in C^n is available. Let D^n be a class of designs p such that $\sum_{s'} n(s') p_1(s') = n'$, $\sum_{s'} \sum_s n(s) p_2(s) p_1(s') = n$ with n, n' ($n < n' < N$) as pre-assigned positive integers. Then, we have

Theorem 2.3.8. Among strategies (e, p) with e in C^n and p in D^n , every strategy (e^*, p^*) with p^* in D^n is 'uniformly admissible'.

Proof. Suppose not. Then a strategy (e, p) with p in D^n , say, is uniformly better than (e^*, p^*) in the sense that

$$\sum_{s'} \sum_s e^2(\bar{s}, \underline{y}, \underline{x}) p_2(s) p_1(s') \leq \sum_{s'} \sum_s e^{*2}(\bar{s}, \underline{y}, \underline{x}) p_2^*(s) p_1^*(s') \quad (2.3.25)$$

uniformly in $(\underline{x}, \underline{y})$ and strictly so for at least one such point.

For any $e = e(\bar{s}, \underline{y}, \underline{x})$ in C^n , we must have

$$e(\bar{s}, \underline{y}, \underline{x}) = 0 \quad \forall \quad \bar{s} \quad \text{if} \quad y_j = 0 = x_j \quad \forall \quad j$$

$$e_2(s', \underline{x}) = 0 \quad \forall \quad s' \quad \text{if} \quad x_j = 0 \quad \forall \quad j$$

$$\text{and} \quad e_1(s, \underline{u}) = 0 \quad \forall \quad s \quad \text{if} \quad x_j = y_j \quad \forall \quad j,$$

by definition of an estimator.

Let $(\underline{X}', \underline{Y}')$ be a point such that $\underline{U}' = \underline{Y}' - \underline{X}'$, $y'_j \neq 0$, $y'_k = 0 \quad \forall k \neq j$ and $x'_j = 0 \quad \forall j$. Then,

$$\Sigma_1 \Sigma_2 e_1(s, \underline{U}') p_2(s) p_1(s') = y'_j \text{ and } \Sigma_1 \Sigma_2 e_1^2(s, \underline{U}') p_2(s) p_1(s') \geq y_j'^2 / Q_j \text{ giving us } Q_j \geq Q_j^* \implies Q_j = Q_j^* \implies e_1(s, \underline{U}) = \frac{y_j}{Q_j}$$

for any s with $p(\bar{s}) > 0$ and any point $(\underline{X}, \underline{Y})$ such that the set of co-ordinates $(x_j, y_j) \cdot j \in s$ includes only one $y_j \neq 0$ and all $x_j = 0$.

Let for a point $(\underline{X}'', \underline{Y}'')$, $x''_j = y''_j \quad \forall j$ and $x''_j \neq 0$, $x''_k = 0 \quad \forall k \neq j$. Then, $\Sigma_1 e_2(s', \underline{X}'') p_1(s') = x''_j$

$$\implies \Sigma_1 e_2^2(s', \underline{X}'') p_1(s') \geq x_j''^2 / P_j \implies P_j \geq P_j^*$$

$$\implies P_j = P_j^* \implies e_2(s', \underline{X}) = x_j / P_j \quad \forall s' \text{ with } p_1(s') > 0$$

and $\forall \underline{X}$ such that the set of co-ordinates $x_j \cdot j \in s'$ includes only one non-vanishing co-ordinate x_j .

Let for a point $(\underline{X}''', \underline{Y}''')$, $y''_j \neq 0$, $y''_k \neq 0$, $y''_1 = 0 \quad \forall 1 \neq j, k$ and $x''_j = 0 \quad \forall j$. Then

$$\Sigma_3 \Sigma_4 e_1(s, \underline{U}''') p_2(s) p_1(s') = Q_{jk} (y''_j / Q_j + y''_k / Q_k)$$

$$\Rightarrow \Sigma_3 \Sigma_4 e_1^2(s, \underline{U}''') p_2(s) p_1(s') \geq Q_{jk} \left(\frac{y_j'''}{Q_j} + \frac{y_k'''}{Q_k} \right)^2$$

by Cauchy - Schwarz's inequality.

Again we have

$$\begin{aligned} \Sigma_3 \Sigma_4 e_1^2(s, \underline{U}''') p_2(s) p_1(s') &\leq Q_{jk} \left(y_j''^2 / Q_j^2 + y_k''^2 / Q_k^2 \right) \\ &\quad + 2 Q_{jk}^* \frac{y_j'' y_k''}{Q_j Q_k} \end{aligned}$$

$$\Rightarrow Q_{jk} y_j'' y_k'' \leq Q_{jk}^* y_j'' y_k'' \Rightarrow Q_{jk} = Q_{jk}^* .$$

Let for a point $(\underline{x}^{(iv)}, \underline{y}^{(iv)})$, $x_j^{(iv)} = y_j^{(iv)} \quad \forall j$ and $x_j^{(iv)} \neq 0$, $x_k^{(iv)} \neq 0$, $x_l^{(iv)} = 0 \quad \forall l \neq j, k$. Then

$$\Sigma_3 e_2(s', \underline{x}^{(iv)}) p_1(s') = P_{jk} (x_j^{(iv)} / P_j + x_k^{(iv)} / P_k)$$

$$\Rightarrow \Sigma_3 e_2^2(s', \underline{x}^{(iv)}) p_1(s') \geq P_{jk} (x_j^{(iv)} / P_j + x_k^{(iv)} / P_k)^2$$

by Cauchy - Schwarz's inequality.

Again,

$$\begin{aligned} \Sigma_3 e_2^2(s', \underline{x}^{(iv)}) p_1(s') &\leq P_{jk} (x_j^{2(iv)} / P_j^2 + x_k^{2(iv)} / P_k^2) \\ &\quad + 2 P_{jk}^* \frac{x_j^{(iv)} x_k^{(iv)}}{P_j P_k} \end{aligned}$$

$$\Rightarrow P_{jk} x_j^{(iv)} x_k^{(iv)} \leq P_{jk}^* x_j^{(iv)} x_k^{(iv)}$$

$$\Rightarrow P_{jk} = P_{jk}^* .$$

Let for a point $(\underline{X}^{(v)}, \underline{Y}^{(v)})$, $y_j^{(v)} = 0 \ \forall j$, $x_j^{(v)} \neq 0$, $x_k^{(v)} = 0 \ \forall k \neq j$. Then, $e_2(s', \underline{X}^{(v)}) = x_j^{(v)}/P_j$ and

$$\Sigma_1 \Sigma_2 e_1(s, \underline{U}^{(v)}) p_2(s) p_1(s') = - x_j^{(v)} \Rightarrow \Sigma_1 \Sigma_2 e_1^2(s, \underline{U}^{(v)}) p_2(s) p_1(s') \geq x_j^{2(v)} / Q_j . \text{ Again we have,}$$

$$\Sigma_1 \Sigma_2 e_1^2(s, \underline{U}^{(v)}) p_2(s) p_1(s') \leq x_j^{2(v)} / Q_j$$

$\Rightarrow e_1(s, \underline{U}) = - \frac{x_j}{Q_j} \ \forall s$ with $p(\bar{s}) > 0$ and any point $(\underline{X}, \underline{Y})$ such that the set of co-ordinates $(x_j, y_j) \cdot j \in s$ includes only one $x_j \neq 0$ and all $y_j = 0$.

Let for a point $(\underline{X}^{(vi)}, \underline{Y}^{(vi)})$, $y_j^{(vi)} \neq 0$, $y_k^{(vi)} = 0 \ \forall k \neq j$ and $x_k^{(vi)} \neq 0$, $x_l^{(vi)} = 0 \ \forall l \neq k$. Then,

$$\Sigma_3 \Sigma_4 e_1(s, \underline{U}^{(vi)}) p_2(s) p_1(s') = Q_{jk} (y_j^{(vi)} / Q_j - x_k^{(vi)} / Q_k)$$

$$\Rightarrow \Sigma_3 \Sigma_4 e_1^2(s, \underline{U}^{(vi)}) p_2(s) p_1(s') \geq Q_{jk} \left(\frac{y_j^{(vi)}}{Q_j} - \frac{x_k^{(vi)}}{Q_k} \right)^2 .$$

Again, $\sum_j \sum_k e_1^2(s, \underline{U}^{(vi)}) p_2(s) p_1(s')$

$$\leq Q_{jk} (y_j^{(vi)}/Q_j - x_k^{(vi)}/Q_k)^2 + \frac{2y_j^{(vi)} x_k^{(vi)}}{Q_j P_k} (R_{jk}^* - R_{jk})$$

$$\Rightarrow y_j^{(vi)} x_k^{(vi)} R_{jk} \leq y_j^{(vi)} x_k^{(vi)} R_{jk}^*$$

$$\Rightarrow R_{jk} = R_{jk}^*$$

$$\Rightarrow \sum_{s'} \sum_s e^{*2}(\bar{s}, \underline{Y}, \underline{X}) p_2^*(s) p_1^*(s') = \sum_{s'} \sum_s e^{*2}(\bar{s}, \underline{Y}, \underline{X}) p_2(s) p_1(s')$$

So, in (2.3.25) inequality cannot be strict.

$\Rightarrow$ contradiction and hence proof.

In order to investigate further properties of e_{bs}^* for designs in D , we shall in what follows confine within the sub-class $D(n', n)$ of fixed sample size designs within D such that for any design p in $D(n', n)$, we require the size of any sample s' with $p_1(s') > 0$ to be an integer n' and that of any sample s (a subset of s' with $p_1(s') > 0$), with $p_2(s|s') > 0$ to be another integer $n (< n')$, the units in samples and sub-samples being all distinct. For such designs in $D(n', n)$, following Horvitz/Thompson (1952),

Yates/Grundy (1953), Godambe (1955), Hanurav (1966) and others, we may easily check the following properties (defining suitable indicator functions), namely,

$$\sum_{j=1}^N P_j = n', \quad \sum_{\substack{k=1 \\ (+j)}}^N P_{jk} = (n'-1)P_j, \quad \sum_{j \neq k=1}^N P_{jk} = n'(n'-1),$$

$$\sum_{j=1}^N Q_j = n, \quad \sum_{\substack{k=1 \\ (+j)}}^N Q_{jk} = (n-1)Q_j, \quad \sum_{j \neq k=1}^N Q_{jk} = n(n-1)$$

with obvious interpretations concerning inclusion-probabilities and further $\sum_{\substack{k=1 \\ (+j)}}^N R_{jk} = (n'-1)Q_j$, $\sum_{j \neq k=1}^N R_{jk} = n(n'-1)$,

for which we do not offer any physical interpretation but which we require to use in our subsequent analysis. We give below the proofs of some results (proofs in cases that follow obviously from available literature are mostly omitted).

$$\text{Let } R(s', s, j) = \begin{cases} 1 & \text{if } j \in s \text{ or } s \rightarrow j \text{ } (s \subset s') \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} \text{Then } \sum_{j=1}^N Q_j &= \sum_{j=1}^N \sum_1 \sum_2 p_2(s) p_1(s') \\ &= \sum_{j=1}^N \sum_{s'} \sum_{s \subset s'} p_2(s) p_1(s') R(s', s, j) \end{aligned}$$

$$\begin{aligned}
 &= \sum_{s'} \sum_{s \subseteq s'} p_2(s) p_1(s') \sum_{j=1}^N R(s', s, j) \\
 &= n.
 \end{aligned}$$

$$\begin{aligned}
 \text{Let } t(s', s, j, k) &= 1 \text{ if } j, k \in s \text{ or } s \rightarrow j, k (s \subseteq s') \\
 &= 0 \text{ otherwise}
 \end{aligned}$$

$$\begin{aligned}
 \text{Then } \sum_{\substack{k=1 \\ \neq j}}^N Q_{jk} &= \sum_{\substack{k=1 \\ \neq j}}^N \sum_3 \sum_4 p_2(s) p_1(s') \\
 &= \sum_{k=1}^N \sum_{s'} \sum_{s \subseteq s'} p_2(s) p_1(s') t(s', s, j, k) \\
 &= \sum_{s'} \sum_{s \subseteq s'} p_2(s) p_1(s') \sum_{\substack{k=1 \\ \neq j}}^N t(s', s, j, k) \\
 &= \sum_{s'} \sum_{s \subseteq s'} p_2(s) p_1(s') (n-1) R(s', s, j) \\
 &= (n-1) \sum_1 \sum_2 p_2(s) p_1(s') \\
 &= (n-1) Q_j.
 \end{aligned}$$

$$\text{and } \sum_{j \neq k=1}^N \sum_{j=1}^N Q_{jk} = \sum_{j=1}^N (n-1) Q_j = n(n-1).$$

$$\begin{aligned}
 \text{Let } t^*(s', s, j, k) &= 1 \text{ if } j, k \in s' \text{ and } j \in s \text{ or } s' \rightarrow j, k \text{ and } s \rightarrow j \\
 &= 0 \text{ otherwise.}
 \end{aligned}$$

$$\begin{aligned}
 \text{Then, } \sum_{\substack{k=1 \\ \neq j}}^N R_{jk} &= \sum_{\substack{k=1 \\ \neq j}}^N \sum_3 \sum_2 p_2(s) p_1(s') \\
 &= \sum_{\substack{k=1 \\ \neq j}}^N \sum_{s'} \sum_{s \subset s'} p_2(s) p_1(s') t^*(s', s, j, k) \\
 &= \sum_{s'} \sum_{s \subset s'} p_2(s) p_1(s') \sum_{\substack{k=1 \\ \neq j}}^N t^*(s', s, j, k) \\
 &= \sum_{s'} \sum_{s \subset s'} p_2(s) p_1(s') (n'-1) R(s', s, j) \\
 &= (n'-1) \sum_1 \sum_2 p_2(s) p_1(s') \\
 &= (n'-1) Q_j
 \end{aligned}$$

$$\text{and } \sum_{j \neq k=1}^N \sum R_{jk} = \sum_{j=1}^N (n'-1) Q_j = n(n'-1).$$

Further we shall assume for estimators in C_1 and for designs in D that $\text{Cov}_p \left(\sum_{j \in s} b_{sj} y_j, \sum_{j \in s} b_{sj}^I x_j + \sum_{j \in s'} b_{sj}^{II} x_j \right) \geq 0$ whatever y, x may be and when this condition is imposed we say that the estimator e_{bs} belongs to a subclass C_2 of C_1 . Here $\sum_{j \in s} b_{sj} y_j$ is a linear unbiased estimator for Y and $\sum_{j \in s} b_{sj}^I x_j + \sum_{j \in s'} b_{sj}^{II} x_j$ is a linear unbiased estimator for zero. Since y and x are non-negatively correlated this condition seems reasonable and

will be found useful in subsequent analysis. We shall now postulate a super-population model and assume the availability of known positive size measures W_j 's and in terms of the model we shall assume the following forms of expectations with respect to the distribution of the vectors $\underline{Y}$ and $\underline{X}$ which in terms of the super-population are now supposed to be stochastic. However, W_j 's are supposed to be fixed (non-stochastic) quantities. Denoting by E_M the expectation-operator with respect to the model M (say), we assume

$$E_M(y_j) = \theta_j, E_M(y_j - \theta_j)^2 = \sigma_j^2, E_M(y_j - \theta_j)(y_k - \theta_k) = \rho \sigma_j \sigma_k \quad \forall j \neq k$$

$$E_M(x_j) = \theta'_j, E_M(x_j - \theta'_j)^2 = \sigma_j'^2, E_M(x_j - \theta'_j)(x_k - \theta'_k) = \rho' \sigma_j' \sigma_k' \quad \forall j \neq k$$

$$C_M(y_j, x_j) = E_M(y_j - \theta_j)(x_j - \theta'_j) = \sigma_j''^2 \quad (\text{so written as a squared quantity to emphasize non-negativity of correlation between } y \text{ and } x),$$

$$E_M(y_j - \theta_j)(x_k - \theta'_k) = \psi \sigma_j'' \sigma_k'' \quad (\text{written conveniently so with}$$

ψ as a constant which should be non-negative in view of non-negative correlation between y and x) where θ_j, θ'_j 's are any real numbers $\forall j$, $\sigma_j, \sigma_j', \sigma_j''$ are any positive real numbers $\forall j$ and the correlation co-efficients are such that $0 \leq \rho, \rho' \leq 1$. In addition, we further assume $\psi \leq 1$.

Assuming the above model we have the

Theorem 2.3.9. For any design p in D and for any estimator $e_{b\bar{s}}$ in the class C_2 we have

$$\begin{aligned} E_M V_p(e_{b\bar{s}}) \geq & (1-\rho) \sum_{j=1}^N \sigma_j^2 \left(\frac{1}{Q_j} - 1 \right) + (1-\rho') \sum_{j=1}^N \sigma_j'^2 \left(\frac{1}{Q_j} - \frac{1}{P_j} \right) \\ & + 2(1-\psi) \sum_{j=1}^N \sigma_j''^2 \left(\frac{1}{P_j} - \frac{1}{Q_j} \right) \end{aligned} \quad (2.3.26)$$

Proof $E_M V_p(e_{b\bar{s}})$

$$\begin{aligned} &= \sum_{j=1}^N (\sigma_j^2 + \theta_j^2) \sum_1 \sum_2 b_{sj}^2 p_2(s) p_1(s') \\ &+ \sum_{j \neq k=1}^N (\rho \sigma_j \sigma_k + \theta_j \theta_k) \sum_3 \sum_4 b_{sj} b_{sk} p_2(s) p_1(s') \\ &+ \sum_{j=1}^N (\sigma_j'^2 + \theta_j'^2) \sum_1 \sum_2 b_{sj}^2 p_2(s) p_1(s') + \sum_1 b_{sj}''^2 p_1(s') \\ &\quad + 2 \sum_1 \sum_2 b_{sj}' b_{sj}'' p_2(s) p_1(s') \int \\ &+ \sum_{j \neq k=1}^N (\rho' \sigma_j' \sigma_k' + \theta_j' \theta_k') \sum_3 \sum_4 b_{sj}' b_{sk}' p_2(s) p_1(s') \\ &\quad + \sum_3 b_{sj}'' b_{sk}'' p_1(s') + 2 \sum_3 \sum_2 b_{sj}' b_{sk}'' p_2(s) p_1(s') \int \\ &+ 2 \sum_{j=1}^N (\sigma_j''^2 + \theta_j \theta_j') \sum_1 \sum_2 b_{sj} b_{sj}' p_2(s) p_1(s') \\ &\quad + \sum_1 \sum_2 b_{sj} b_{sj}'' p_2(s) p_1(s') \int \end{aligned}$$

$$\begin{aligned}
 & + 2 \sum_{j \neq k=1}^N (\psi \sigma_j^* \sigma_k^* + \theta_j \theta_k') \left[\sum_{j \neq k=1}^N b_{sj}^- b_{sk}^+ p_2(s) p_1(s') \right. \\
 & \quad \left. + \sum_{j \neq k=1}^N b_{sj}^- b_{sk}^* p_2(s) p_1(s') \right] \\
 & - \sum_{j=1}^N (\sigma_j^2 + \theta_j^2) - \sum_{j \neq k=1}^N (\rho \sigma_j \sigma_k + \theta_j \theta_k) \\
 & = T + E_p \left[\sum_{j \in S} b_{sj}^- \theta_j + \sum_{j \in S} b_{sj}^+ \theta_j' + \sum_{j \in S} b_{sj}^* \theta_j' \right]^2 - \theta^2 \\
 & \text{(where } \theta = \sum_{j=1}^N \theta_j \text{ and } T \text{ denotes the terms not involving } \theta_j, \theta_j' \text{'s in the previous expression)} \\
 & \geq (1-\rho) \sum_{j=1}^N \sigma_j^2 \sum_{j \in S} b_{sj}^2 p_2(s) p_1(s') + \rho E_p \left(\sum_{j \in S} b_{sj}^- \sigma_j \right)^2 \\
 & - \rho \left(\sum_{j=1}^N \sigma_j \right)^2 - (1-\rho) \sum_{j=1}^N \sigma_j^2 + (1-\rho') \sum_{j=1}^N \sigma_j'^2 \left[\sum_{j \in S} b_{sj}^+ p_2(s) p_1(s') \right. \\
 & \quad \left. + \sum_{j \in S} b_{sj}^* p_1(s') + 2 \sum_{j \in S} b_{sj}^- b_{sj}^* p_2(s) p_1(s') \right] \\
 & + \rho' E_p \left(\sum_{j \in S} b_{sj}^+ \sigma_j' + \sum_{j \in S} b_{sj}^* \sigma_j' \right)^2 \\
 & + 2 (1-\psi) \sum_{j=1}^N \sigma_j^* \sum_{j \in S} b_{sj}^- b_{sj}^+ p_2(s) p_1(s') \\
 & + \sum_{j \in S} b_{sj}^- b_{sj}^* p_2(s) p_1(s') \left]
 \end{aligned}$$

$$+ 2 \Psi E_p \left[\left(\sum_{j \in S} b_{sj}^- \sigma_j'' \right) \left(\sum_{j \in S} b_{sj}^+ \sigma_j'' + \sum_{j \in S'} b_{sj}'' \sigma_j'' \right) \right]$$

$$\left[\text{because } E_p \left(\sum_{j \in S} b_{sj}^- \theta_j + \sum_{j \in S} b_{sj}^+ \theta_j' + \sum_{j \in S'} b_{sj}'' \theta_j' \right)^2 \right.$$

$$\geq E_p^2 \left(\sum_{j \in S} b_{sj}^- \theta_j + \sum_{j \in S} b_{sj}^+ \theta_j' + \sum_{j \in S'} b_{sj}'' \theta_j' \right) = \theta^2 \left. \right]$$

$$\geq (1-\rho) \sum_{j=1}^N \sigma_j^2 \left[\sum_1 \sum_2 b_{sj}^2 p_2(s) p_1(s') - 1 \right]$$

$$+ (1-\rho') \sum_{j=1}^N \sigma_j'^2 \left[\sum_1 \sum_2 b_{sj}^{\prime 2} p_2(s) p_1(s') + \sum_1 b_{sj}^{\prime 2} p_1(s') \right.$$

$$\left. + 2 \sum_1 \sum_2 b_{sj}^+ b_{sj}'' p_2(s) p_1(s') \right]$$

$$+ 2 (1-\Psi) \sum_{j=1}^N \sigma_j^{\prime \prime 2} \left[\sum_1 \sum_2 b_{sj}^- b_{sj}^+ p_2(s) p_1(s') \right.$$

$$\left. + \sum_1 \sum_2 b_{sj}^- b_{sj}'' p_2(s) p_1(s') \right]$$

$$+ 2 \Psi E_p \left[\left(\sum_{j \in S} b_{sj}^- \sigma_j'' \right) \left(\sum_{j \in S} b_{sj}^+ \sigma_j'' + \sum_{j \in S} b_{sj}'' \sigma_j'' \right) \right]$$

$$\left[\text{because, } E_p \left(\sum_{j \in S} b_{sj}^- \sigma_j \right)^2 \geq E_p^2 \left(\sum_{j \in S} b_{sj}^- \sigma_j \right) = \left(\sum_{j=1}^N \sigma_j \right)^2, \right.$$

$$E_p \left(\sum_{j \in S} b_{sj}^+ \sigma_j' + \sum_{j \in S'} b_{sj}'' \sigma_j' \right)^2$$

$$\begin{aligned} &\geq E_p^2 \left(\sum_{j \in S} b_{sj}^i \sigma_j^i + \sum_{j \in S'} b_{sj}^n \sigma_j^i \right) = 0 \int \\ &\geq (1-f) \sum_{j=1}^N \sigma_j^2 \left(\frac{1}{Q_j} - 1 \right) + (1-f') \sum_{j=1}^N \sigma_j'^2 \left(\frac{1}{Q_j} - \frac{1}{P_j} \right) \\ &\quad + 2(1-\psi) \sum_{j=1}^N \sigma_j^{n2} \left(\frac{1}{P_j} - \frac{1}{Q_j} \right) \end{aligned}$$

Using the inequalities (2.3.16) - (2.3.18) and the assumption that

$$\text{Cov}_p \left\{ \sum_{j \in S} b_{sj} y_j, \sum_{j \in S} b_{sj}^i x_j + \sum_{j \in S'} b_{sj}^n x_j \right\} \geq 0$$

which implies that

$$\begin{aligned} &E_p \left[\left(\sum_{j \in S} b_{sj} \sigma_j^n \right) \left(\sum_{j \in S} b_{sj}^i \sigma_j^n + \sum_{j \in S'} b_{sj}^n \sigma_j^n \right) \right] \\ &\geq E_p \left[\sum_{j \in S} b_{sj} \sigma_j^n \right] \cdot E_p \left[\sum_{j \in S} b_{sj}^i \sigma_j^n + \sum_{j \in S'} b_{sj}^n \sigma_j^n \right] \\ &= \left(\sum_{j=1}^N \sigma_j^n \right) \cdot 0 = 0 \int \end{aligned}$$

The lower bound (2.3.26) is attained if the following conditions hold simultaneously, viz.,

$$\sum_{j \in S} b_{sj} \theta_j + \sum_{j \in S} b_{sj}^i \theta_j^i + \sum_{j \in S'} b_{sj}^n \theta_j^i = \sum_{j=1}^N \theta_j,$$

$$\sum_{j \in S} b_{sj}^- \sigma_j = \sum_{j=1}^N \sigma_j, \quad \sum_{j \in S} b_{sj}^+ \sigma_j' + \sum_{j \in S} b_{sj}'' \sigma_j'' = 0,$$

$$\left(\sum_{j \in S} b_{sj}^- \sigma_j'' \right) \left(\sum_{j \in S} b_{sj}^+ \sigma_j'' + \sum_{j \in S} b_{sj}'' \sigma_j'' \right) = 0,$$

$$b_{sj}^- = \frac{1}{Q_j}, \quad b_{sj}^+ = -\frac{1}{Q_j} \quad \text{and} \quad b_{sj}'' = \frac{1}{P_j} \quad \forall j.$$

These conditions (which we may describe as (2.3.27)) will hold if the following conditions (which we may describe as (2.3.28)) hold together, namely

p belongs to $D_{\pi}(n', n)$ which is a sub-class of designs $D(n', n)$ for which $P_j = \frac{n' W_j}{W}$, $Q_j = \frac{n W_j}{W} \quad \forall j$ (i.e. those with inclusion-probabilities proportional to size i.e. these are IPPS or the so-called π ps designs), and the model M is reduced to a form M_1 (say) for which $\theta_j = \rho W_j$, $\theta_j' = \rho' W_j$, $\sigma_j = \sigma W_j$, $\sigma_j' = \sigma' W_j$ i.e. the model-expectations and model-standard deviations are proportional to size-measures as in Godambe/Joshi (1965) and Cassel/Särndal/Wretman (1976), for example (here $W = \sum_{j=1}^N W_j$, ρ, ρ' are any real quantities and σ, σ' are any real positive numbers) and the b_{sj}^- , b_{sj}^+ , b_{sj}'' 's are as above in (2.3.27) and $\sigma_j'' = \sigma'' W_j$ where σ'' is any real positive number.

When (2.3.28) holds, then we may check that $e_{b\bar{s}}$ reduces to

$$e_{b\bar{s}}^* = \sum_{j \in S} \frac{y_j - x_j}{Q_j} + \sum_{j \in S'} \frac{x_j}{P_j}$$

and the right hand side of (2.3.26) equals $E_M V_p (e_{b\bar{s}}^*)$

which works out (under (2.3.28)) as

$$\begin{aligned} E_M V_p (e_{b\bar{s}}^*) &= \frac{W^2}{n} [\sigma^2 + \sigma'^2 - 2 \sigma''^2] + (\beta - \beta')^2 J \\ &+ \frac{W^2}{n'} [\sigma'^2 + \beta'^2] + W^2 (1 - \frac{1}{n'}) (\rho' \sigma'^2 + \beta'^2) \\ &+ \frac{2W^2}{n} [(\sigma''^2 - \sigma'^2) + \beta' (\beta - \beta')] J \\ &+ 2W^2 (1 - \frac{1}{n'}) [(\psi \sigma''^2 - \rho' \sigma'^2) + \beta' (\beta - \beta')] J \\ &- \sigma^2 \sum_{j=1}^N W_j^2 - \rho \sigma^2 \sum_{j \neq k=1}^N W_j W_k - \beta^2 W^2 \\ &= (1 - \rho) \sigma^2 \sum_{j=1}^N W_j^2 (\frac{1}{Q_j} - 1) + (1 - \rho') \sigma'^2 \sum_{j=1}^N W_j^2 (\frac{1}{Q_j} - \frac{1}{P_j}) \\ &+ 2(1 - \psi) \sigma''^2 \sum_{j=1}^N W_j^2 (\frac{1}{P_j} - \frac{1}{Q_j}). \end{aligned}$$

So, we have

Theorem 2.3.10. For any design p in $D_{\pi}(n', n)$ if the model specified in (2.3.28) holds then, for any estimator e_{bs} in the class C_2 we have

$$\begin{aligned} E_M V_p(e_{bs}) &\geq (1 - \rho) \sigma^2 \sum_{j=1}^N w_j^2 \left(\frac{1}{Q_j} - 1 \right) \\ &+ (1 - \rho') \sigma'^2 \sum_{j=1}^N w_j^2 \left(\frac{1}{Q_j} - \frac{1}{P_j} \right) \\ &+ 2(1 - \psi) \sigma''^2 \sum_{j=1}^N w_j^2 \left(\frac{1}{P_j} - \frac{1}{Q_j} \right) \\ &= E_M V_p(e_{bs}^*) \end{aligned}$$

From theorems 2.3.9 and 2.3.10 we have the following

Corollary 2.3.9. For any design p in D and for any estimator e_{bs} in the class C_1 (and not just in C_2), in case $\psi = 0$ for the model M , we have

$$\begin{aligned} E_M V_p(e_{bs}) &\geq (1 - \rho) \sum_{j=1}^N \sigma_j^2 \left(\frac{1}{Q_j} - 1 \right) \\ &+ (1 - \rho') \sum_{j=1}^N \sigma_j'^2 \left(\frac{1}{Q_j} - \frac{1}{P_j} \right) \\ &+ 2 \sum_{j=1}^N \sigma_j''^2 \left(\frac{1}{P_j} - \frac{1}{Q_j} \right) \end{aligned}$$

Corollary 2.3.10. For any design p in $D_{\pi}(n', n)$ if the model specified in (2.3.28) holds with $\Psi = 0$, then for any estimator $e_{b\bar{s}}$ in the class C_1 we have

$$\begin{aligned} E_M V_p(e_{b\bar{s}}) &\geq (1-\rho) \sigma^2 \sum_{j=1}^N w_j^2 \left(\frac{1}{Q_j} - 1 \right) \\ &+ (1-\rho') \sigma'^2 \sum_{j=1}^N w_j^2 \left(\frac{1}{Q_j} - \frac{1}{P_j} \right) \\ &+ 2 \sigma''^2 \sum_{j=1}^N w_j^2 \left(\frac{1}{P_j} - \frac{1}{Q_j} \right) \\ &= E_M V_p(e_{b\bar{s}}^*). \end{aligned}$$

Remark These results are extended versions of cassel/Särndal/Wretman's (1976) works for uni-phase sampling. Starting with an 'uncorrelated' super-population model (to be called M_2) (i.e. y_i 's are pairwise uncorrelated, so are x_i 's and y_i 's are uncorrelated with x_j 's for $i \neq j$) corresponding optimality results for $e_{b\bar{s}}^*$ as extended (double sampling) versions of Godambe's (1955) works are also derivable likewise. The obvious demonstration is as follows:

Let under a super-population set-up y_j 's be pairwise uncorrelated, x_j 's be pairwise uncorrelated and y_j 's be

uncorrelated with x_k 's for $j \neq k$. Let $E_M(y_j) = \theta_j$,
 $V_M(y_j) = E_M(y_j - \theta_j)^2 = \sigma_j^2$, $E_M(x_j) = 0_j$, $V_M(x_j) =$
 $E_M(x_j - \theta_j')^2 = \sigma_j'^2$, $E_M(y_j - \theta_j)(x_j - \theta_j') = \sigma_j^{*2}$
 (using a squared quantity to emphasize y_j 's are non-negatively
 correlated with x_j 's), $j = 1, \dots, N$.

Assuming the above model we have the

Theorem 2.3.11. For any design p in D and for any estimator
 $e_{b\bar{s}}$ in the class C_1 we have

$$E_M V_p(e_{b\bar{s}}) \geq \sum_{j=1}^N \sigma_j^2 \left(\frac{1}{Q_j} - 1 \right) + \sum_{j=1}^N \sigma_j'^2 \left(\frac{1}{Q_j} - \frac{1}{F_j} \right) \\
+ 2 \sum_{j=1}^N \sigma_j^{*2} \left(\frac{1}{F_j} - \frac{1}{Q_j} \right) \quad (2.3.29)$$

Proof. Following Godambe/Thompson (1977). we can write

$$E_M V_p(e_{b\bar{s}}) = E_p V_M(e_{b\bar{s}}) + E_p \Delta_M^2(e_{b\bar{s}}) - V_M(Y)$$

$$\text{where } \Delta_M(e_{b\bar{s}}) = E_M(e_{b\bar{s}}) - E_M(Y).$$

$$\text{Now } E_p V_M(e_{b\bar{s}}) = \sum_{j=1}^N \sigma_j^2 \sum_1 \sum_2 b_{sj}^2 p_2(s) p_1(s') \\
+ \sum_{j=1}^N \sigma_j'^2 \left[\sum_1 \sum_2 b_{sj}^2 p_2(s) p_1(s') + \sum_1 b_{sj}^2 p_1(s') \right]$$

$$\begin{aligned}
 & + 2 \sum_1 \sum_2 b_{sj}^I b_{sj}^{II} p_2(s) p_1(s') J \\
 & + 2 \sum_{j=1}^N \sigma_j^{*2} \left[\sum_1 \sum_2 b_{sj}^I (b_{sj}^I + b_{sj}^{II}) p_2(s) p_1(s') J \right] \\
 & \geq \sum_{j=1}^N \frac{\sigma_j^2}{Q_j} + \sum_{j=1}^N \sigma_j'^2 \left(\frac{1}{Q_j} - \frac{1}{P_j} \right) + 2 \sum_{j=1}^N \sigma_j^{*2} \left(\frac{1}{P_j} - \frac{1}{Q_j} \right)
 \end{aligned}$$

using the inequalities (2.3.16) - (2.3.18).

Hence we have,

$$\begin{aligned}
 E_M V_p(e_{b\bar{s}}) & \geq E_p V_M(e_{b\bar{s}}) - V_M(Y) \\
 & \geq \sum_{j=1}^N \sigma_j^2 \left(\frac{1}{Q_j} - 1 \right) + \sum_{j=1}^N \sigma_j'^2 \left(\frac{1}{Q_j} - \frac{1}{P_j} \right) + 2 \sum_{j=1}^N \sigma_j^{*2} \left(\frac{1}{P_j} - \frac{1}{Q_j} \right)
 \end{aligned}$$

Q.E.D.

Under a simplified form of the above model M_2 we have the following theorem of practical importance

Theorem 2.3.12. For any design p in $D_\pi(n', n)$ for which

$$P_j = \frac{n' W_j}{W} \quad \text{and} \quad Q_j = \frac{n W_j}{W} \quad \forall j \quad \text{and when the model } M_2$$

holds with the restrictions $\theta_j = \beta W_j$, $\theta_j' = \beta' W_j$ for

β, β' some unknown positive constants (M_2 may then be called M_3), we have for any estimator $e_{b\bar{s}}$ in C_1 ,

$$E_M V_p(e_{b\bar{s}}) \geq E_M V_p(e_{b\bar{s}}^*)$$

Proof. Writing

$$E_M V_p(e_{bs}^*) = E_p V_M(e_{bs}^*) + E_p \Delta_M^2(e_{bs}^*) - V_M(Y)$$

and noting that $E_p V_M(e_{bs}^*) \geq E_p V_M(e_{bs}^*)$

and $\Delta_M(e_{bs}^*) = 0$ in this case, we have

$$E_M V_p(e_{bs}^*) - E_M V_p(e_{bs}^*) \geq E_p \Delta_M^2(e_{bs}^*),$$

proving the theorem.

Extending Godambe/Joshi's (1965) results one more optimality of e^* follows under another model (M_4 , say) stipulating that y_j 's are distributed independently of each other, x_j 's distributed independently of each other and (y_j, x_j) 's distributed independently of each other i.e. for example (y_j, x_j) is distributed independently of (y_k, x_k) , (y_1, x_1) etc. for $j \neq k \neq 1$. Also, let y_j 's be uncorrelated with x_k 's for $j \neq k$ and let $E_M(y_j) = \theta_j$, $V_M(y_j) = E_M(y_j - \theta_j)^2 = \sigma_j^2$, $E_M(x_j) = \theta_j'$, $V_M(x_j) = E_M(x_j - \theta_j')^2 = \sigma_j'^2$, $E_M(y_j - \theta_j)(x_j - \theta_j') = \sigma_j''^2$ (using a squared quantity to emphasize y_j 's are non-negatively correlated with x_j 's), $j = 1, \dots, N$.

Now we have the

Theorem 2.3.13. When the model M_4 holds for every strategy in C^* , we have

$$E_M V_P(e) \geq \sum_{j=1}^N \sigma_j^2 \left(\frac{1}{Q_j} - 1 \right) + \sum_{j=1}^N \sigma_j'^2 \left(\frac{1}{Q_j} - \frac{1}{P_j} \right) \\ + 2 \sum_{j=1}^N \sigma_j''^2 \left(\frac{1}{P_j} - \frac{1}{Q_j} \right)$$

Proof. We write

$$e(\bar{s}, y, x') = e^*(\bar{s}, y, x') + h(\bar{s}, y, x')$$

$$\text{where } 0 = E \left[h(\bar{s}, y, x') | d' \right] = \sum_{s \in S'} h(\bar{s}, y, x') p_2(s)$$

$$\text{implying } \sum_{s \in S'} h(\bar{s}, y, x) p_2(s) = - \sum_{s \in S'} h(\bar{s}, y, x') p_2(s) \\ - h_2(s', x') \quad (2.3.30)$$

Then,

$$E_P V_M(e) = E_P V_M(e^*) + E_P V_M(h) + 2 E_P C_M(e^*, h),$$

(C_M denoting covariance with respect to model M_4).

Writing $\epsilon_j = y_j - \theta_j$, $\epsilon_j' = x_j - \theta_j'$, so that

$$E_M(\epsilon_j) = 0 = E_M(\epsilon_j'),$$

$$\begin{aligned}
 p_M(e^*, h) &= E_M E_P \left[\sum_{j \in S} \frac{\varepsilon_j - \varepsilon_j'}{Q_j} + \sum_{j \in S'} \frac{\varepsilon_j'}{P_j} \right] h(\bar{s}, y, x') \\
 &= E_M E_{P_1} \left[\sum_{s \in S'} \left(\sum_{j \in S} \frac{\varepsilon_j - \varepsilon_j'}{Q_j} \right) h(\bar{s}, y, x') p_2(s) \right. \\
 &\quad \left. + \sum_{j \in S'} \frac{\varepsilon_j'}{P_j} \sum_{s \in S'} h(\bar{s}, y, x') p_2(s) \right]
 \end{aligned}$$

(For details, see Appendix 3).

= 0, by independence of (y_j, x_j) and (y_k, x_k) 's $\forall j \neq k$

$$\text{So, } E_P V_M(e) \geq E_P V_M(e^*) \quad (2.3.31)$$

Following Godambe/Thompson (1977), we can write

$$E_M V_P(e) = E_P V_M(e) + E_P \Delta_M^2(e) - V_M(Y),$$

$$\text{where } \Delta_M(e) = E_M(e) - E_M(Y),$$

so that $E_M V_P(e) \geq E_P V_M(e^*) - V_M(Y)$, [using (2.3.31)]

$$\begin{aligned}
 &= E_P \left[\sum_{j \in S} \frac{\sigma_j^2 + \sigma_j'^2 - 2 \sigma_j^{*2}}{Q_j^2} + \sum_{j \in S'} \frac{\sigma_j'^2}{P_j^2} \right. \\
 &\quad \left. + 2 \sum_{j \in S} \frac{\sigma_j^{*2}}{Q_j P_j} - 2 \sum_{j \in S} \frac{\sigma_j'^2}{Q_j P_j} \right] - \sum_{j=1}^N \sigma_j^2
 \end{aligned}$$

$$= \sum_{j=1}^N \sigma_j^2 \left(\frac{1}{Q_j} - 1 \right) + \sum_{j=1}^N \sigma_j'^2 \left(\frac{1}{Q_j} - \frac{1}{P_j} \right) \\ + 2 \sum_{j=1}^N \sigma_j''^2 \left(\frac{1}{P_j} - \frac{1}{Q_j} \right)$$

The theorem is proved.

Under a simplified form of the model we have the following theorem of practical importance.

Theorem 2.3.14. For strategies in class C^* with the additional restrictions that $p(\bar{s}) > 0 \implies n(s) = n$, $n(s') = n'$, $P_j = \frac{n'W_j}{W}$, $Q_j = \frac{nW_j}{W}$, and when the model M_4 holds with the restrictions $\theta_j = \beta W_j$, $\theta_j' = \beta' W_j$ for β, β' some unknown positive constants (M_4 may then be called M_5), we have

$$E_M V_p(e) \geq E_M V_p(e^*)$$

Proof. Writing $E_M V_p(e^*) = E_p V_M(e^*) + E_p \Delta_M^2(e^*) - V_M(Y)$ and noting that $\Delta_M(e^*) = 0$ in this case, we have

$$E_M V_p(e) - E_M V_p(e^*) \geq E_p \Delta_M^2(e), \text{ proving}$$

the theorem.

Next following Godambe/Thompson (1973) we establish one more property for e_{bs}^* in the class of all unbiased estimators for designs in the class $D(n', n)$. For this we consider the following model.

Let a_j and b_j , $j=1, \dots, N$ be some given numbers such that $0 < a_j < 1$, $0 < b_j < 1$ and $\sum_{j=1}^N a_j = n'$, $\sum_{j=1}^N b_j = n$ where n' and n are positive integers as before.

Let F be the prior joint distribution for y 's and x 's such that writing $t_j = \frac{x_j}{b_j}$, $r_j = \frac{y_j}{b_j}$, $t'_j = \frac{x_j}{a_j}$,

F stipulates the selection of one at random out of the N points

$$((t_{j_1}, r_{j_1}, t'_{j_1}), \dots, (t_{j_1}, r_{j_1}, t'_{j_1}), \dots, (t_{j_N}, r_{j_N}, t'_{j_N}))$$

obtainable on permuting the integers $j_1, \dots, j_1, \dots, j_N$

which are the integers $1, \dots, N$ in a certain order,

keeping a_j 's and b_j 's fixed.

Let $D_{PF}(n', n)$ be a class of sampling designs $p(\bar{s})$ such that

$$D_{PF}(n', n) = \left\{ p = p(\bar{s}) : \begin{array}{l} \text{(i) } \sum_1 p_1(s') = P_j = a_j, j=1, \dots, N \\ \text{(ii) } \sum_1 \sum_2 p_2(s) p_1(s') = Q_j = b_j, j=1, \dots, N \end{array} \right\}$$

where $P_j, Q_j, j=1, \dots, N$ as before denote inclusion probabilities.

When a sample $\bar{s}$ is chosen according to $D_{PF}(n', n)$

$$e_{b\bar{s}}^* = \left(\sum_{j \in s} \frac{y_j - x_j}{Q_j} + \sum_{j \in s'} \frac{x_j}{P_j} \right) \text{ reduces to}$$

$$\sum_{j \in s} \frac{y_j - x_j}{b_j} + \sum_{j \in s'} \frac{x_j}{a_j} \quad (2.3.32)$$

and is obviously unbiased for Y with respect to p in $D_{PF}(n', n)$.

Let us also define a wider class of sampling designs $D_F(n', n)$ by allowing the inclusion probabilities P_j and Q_j , $j=1, \dots, N$ to vary, holding the sample sizes fixed, so that

$$D_F(n', n) = \left\{ p = p(\bar{s}) : \begin{array}{l} \text{(i) } P_j > 0, \quad Q_j > 0 \\ \text{(ii) } \sum_{j=1}^N P_j = \sum_{j=1}^N a_j \quad (= n' \text{ fixed}) \\ \sum_{j=1}^N Q_j = \sum_{j=1}^N b_j \quad (= n \text{ fixed}) \end{array} \right\}$$

Obviously $D_{PF}(n', n) \subset D_F(n', n)$.

Theorem 2.3.15. (i) For every p in $D_{PF}(n', n)$, we have a common value for $E_F V_p(e_{b\bar{s}}^*)$, which common value will be written as $E_F V_D(e_{b\bar{s}}^*)$;

$$(ii) \quad E_F V_D(e_{b\bar{s}}^*) \leq E_F V_p(e_{\bar{s}}) \quad (2.3.33)$$

whatever p may be in $D_F(n', n)$.

Proof. We prove (ii) first and then (i).

Proof of (ii): Let $e_{\bar{s}}$ be an unbiased (with respect to every p in $D_F(n', n)$) estimator of the population total Y , of the following form,

$$\begin{aligned} e_{\bar{s}} &= e(\bar{s}, y_j, x_j, | j \in s, j' \in s') \\ &= e_1(s) - e_2(s) + e_3(s') \\ &= e_1(s, y_j | j \in s) - e_2(s, x_j | j \in s) + e_3(s', x_j, | j' \in s') \text{ (say)} \end{aligned}$$

such that $E_p(e_1) = Y$, $E_p(e_2) = X$ and $E_p(e_3) = X$.

$$\text{Then } \sum_{\bar{s}} e(\bar{s}, y_j, x_j, | j \in s, j' \in s') p(\bar{s}) = \sum_{j=1}^N y_j.$$

Writing

$$t_j = \frac{x_j}{b_j}, \quad r_j = \frac{y_j}{b_j} \quad \text{and} \quad t'_j = x_j / a_j, \quad j=1, \dots, N$$

$$e_{\bar{s}} = e(\bar{s}, y_j, x_j, | j \in s, j' \in s')$$

$$= f(\bar{s}, r_j, t_j, t'_j, | j \in s, j' \in s') \text{ and}$$

$$e_1(s, y_j | j \in s) = f_1(s, r_j | j \in s) \text{ we get}$$

$$\begin{aligned} \sum_{\bar{s}} f(\bar{s}) p(\bar{s}) &= E [e_1(s, y_j)] = E [f_1(s, r_j)] \\ &= \sum_{j=1}^N y_j = \sum_{j=1}^N b_j r_j. \end{aligned} \quad (2.3.34)$$

Now, $E_F V_P (e_{\bar{s}})$

$$\begin{aligned}
 &= \frac{1}{N!} \sum_P \left[\sum_{\bar{s}} p(\bar{s}) f^2(\bar{s}, r_{P(j)}, t_{P(j)}, t'_{P(j')}, |j \in s, j' \in s'|) \right. \\
 &\quad \left. - \left(\sum_{j=1}^N b_j r_j \right)^2 \right] \\
 &= \frac{1}{N!} \sum_P \left[\sum_{\bar{s}} p(\bar{s}) f^2(\bar{s}, r_{P(j)}, t_{P(j)}, t'_{P(j')}, |j \in s, j' \in s'|) \right] \\
 &\quad - \frac{1}{N!} \sum_P \left(\sum_{j=1}^N b_j r_j \right)^2 \quad (2.3.35)
 \end{aligned}$$

$\sum_P$ = sum over all possible permutations P of $(1, \dots, N)$,
 a permutation P changing j to $P(j)$, $j=1, \dots, N$.

Taking the same kind of average on both sides of (2.3.34) we have

$$\begin{aligned}
 &\frac{1}{N!} \sum_P \sum_{\bar{s}} p(\bar{s}) f(\bar{s}, r_{P(j)}, t_{P(j)}, t'_{P(j')}, |j \in s, j' \in s'|) \\
 &= \frac{1}{N!} \sum_P \left(\sum_{j=1}^N b_j r_{P(j)} \right) = \frac{(N-1)!}{N!} \sum_{j=1}^N b_j \sum_{j=1}^N r_j \quad (2.3.36)
 \end{aligned}$$

Noting that $\sum_{j=1}^N b_j = n$ and putting $\bar{r} = \frac{1}{N} \sum_{j=1}^N r_j$ we may

write (2.3.36) as

$$\sum_{P, \bar{s}} \frac{p(\bar{s})}{N!} f(\bar{s}, r_{P(j)}, t_{P(j)}, t'_{P(j')}, |j \in s, j' \in s'|) = n \bar{r} \quad (2.3.37)$$

Now from (2.3.37) we see that $f(\bar{s}, r_{P(j)}, t_{P(j)}, t'_{P(j')} | j \in s, j' \in s')$ is an unbiased 'estimator' of $n\bar{r}$ with respect to a sampling design which can be realized on drawing the sample $\bar{s} = (s', s)$ by $p(\bar{s})$ and observing $(r_{P(j)}, t_{P(j)}, t'_{P(j')} | j \in s, j' \in s')$ for a randomly chosen permutation P of the vector $V = \{(t_1, r_1, t'_1), \dots, (t_N, r_N, t'_N)\}$.

We may denote (as Godambe/Thompson (1973)) this sampling design by $(p(\bar{s}), P) = p_P$ (say). The variance of the estimator is evidently

$$\sum_{P, \bar{s}} \frac{p(\bar{s})}{N!} r^2(\bar{s}, r_{P(j)}, t_{P(j)}, t'_{P(j')} | j \in s, j' \in s') - (n\bar{r})^2 \quad (2.3.38)$$

Let $r(d) = \{r_{P(j)}, t_{P(j)} : j \in s, t'_{P(j')} : j' \in s'\}$ denote the reduced data with correspondence between values and labels suppressed.

Let, $\sum_P^{\bar{s}, r(d)}$ = sum over permutation P 's for which $\bar{s}, r(d)$ are as observed,

$\sum_{\bar{s}}^{r(d)}$ = sum over $\bar{s}$'s yielding $r(d)$ so that

$$\sum_{\bar{s}}^{r(d)} \sum_P^{\bar{s}, r(d)} = \sum_{\bar{s}} \sum_P (P | r(d)) ,$$

where $\sum_{r(d)}$ is sum over all permutations for which $r(d)$ is realized for a particular $\bar{s}$.

Then, $g(r(d))$ = the conditional expectation (with respect to p_p) of $f(\bar{s}, r_{P(j)}, t_{P(j)}, t'_{P(j')} | j \in s, j' \in s')$ and is

$$g(r(d)) = \frac{\sum_{\bar{s}} \sum_P^{r(d)} \frac{p(\bar{s})}{N!} f(\bar{s}, r_{P(j)}, t_{P(j)}, t'_{P(j')} | j \in s, j' \in s')}{\sum_{\bar{s}} \sum_P^{r(d)} \frac{p(\bar{s})}{N!}} \quad (2.3.39)$$

Certainly $g(r(d))$ is an unbiased estimator of $n\bar{r}$ with respect to p_p .

By Rao - Blackwellisation, then

$$\sum_{P, \bar{s}} \frac{p(\bar{s})}{N!} g^2(r(d)) \leq \sum_{P, \bar{s}} \frac{p(\bar{s})}{N!} f^2(\bar{s}, r_{P(j)}, t_{P(j)}, t'_{P(j')} | j \in s, j' \in s') \quad (2.3.40)$$

Moreover $r(d)$ is the 'order-statistic' corresponding to $(\bar{s}, r_{P(j)}, t_{P(j)}, t'_{P(j')} | j \in s, j' \in s')$ and hence is 'complete' so that if (2.3.34) is to hold, then $g(r(d))$ must in fact be

$$g(r(d)) = \sum_{j \in s} (r_{P(j)} - t_{P(j)}) + \sum_{j \in s'} t'_{P(j)}$$

This is so because of the following reason :

Since $g(r(d))$ is an unbiased estimator of

$$n\bar{r} = \frac{1}{N!} \sum_P \left[\sum_{j=1}^N b_j r_{P(j)} \right]$$

with respect to p_P ,

$$\begin{aligned} \sum_P \sum_s \frac{p(\bar{s})}{N!} g(r(d)) &= \frac{1}{N!} \sum_P \left[\sum_{j=1}^N b_j r_{P(j)} \right] \\ &= \frac{1}{N!} \sum_P \left[\sum_{j=1}^N b_j r_{P(j)} + x - x \right] \\ &= \frac{1}{N!} \sum_P \left[\sum_{j=1}^N b_j r_{P(j)} + \sum_{j=1}^N a_j t'_{P(j)} - \sum_{j=1}^N b_j t_{P(j)} \right] \\ &= \frac{1}{N!} \sum_P \left[\sum_{j=1}^N b_j \{r_{P(j)} - t_{P(j)}\} + \sum_{j=1}^N a_j t'_{P(j)} \right] \\ &= \frac{(N-1)!}{N!} \left[\sum_{j=1}^N b_j \sum_{j=1}^N (r_j - t_j) + \sum_{j=1}^N a_j \sum_{j=1}^N t'_j \right] \\ &= \frac{(N-1)!}{N!} \left[\sum_{j=1}^N Q_j \sum_{j=1}^N (r_j - t_j) + \sum_{j=1}^N P_j \sum_{j=1}^N t'_j \right] \\ &= \frac{1}{N!} \sum_P \left[\sum_{j=1}^N Q_j \{r_{P(j)} - t_{P(j)}\} + \sum_{j=1}^N P_j t'_{P(j)} \right] \\ &= \frac{1}{N!} \sum_P \sum_{j=1}^N \left[\sum_1 \sum_2 p_2(s) p_1(s') \{r_{P(j)} - t_{P(j)}\} + \sum_1 p_1(s') t'_{P(j)} \right] \end{aligned}$$

$$= \frac{1}{N!} \sum_P \sum_{\bar{s}} \left[\sum_{j \in s} \{r_{P(j)} - t_{P(j)}\} + \sum_{j \in s'} t'_{P(j)} \right] p(\bar{s})$$

(Here $\sum_{\bar{s}} = \sum_{s'} \sum_{s \subset s'}$).

Thus we have,

$$\sum_P \sum_{\bar{s}} \frac{p(\bar{s})}{N!} \left[g(r(d)) - \left\{ \sum_{j \in s} (r_{P(j)} - t_{P(j)}) + \sum_{j \in s'} t'_{P(j)} \right\} \right] = 0$$

$$\Rightarrow g(r(d)) = \sum_{j \in s} (r_{P(j)} - t_{P(j)}) + \sum_{j \in s'} t'_{P(j)}$$

because of "completeness of the order statistic $r(d)$ ".

Hence,

$$\begin{aligned} & \sum_{P, \bar{s}} \frac{p(\bar{s})}{N!} \left[\sum_{j \in s} (r_{P(j)} - t_{P(j)}) + \sum_{j \in s'} t'_{P(j)} \right]^2 \\ & \leq \sum_{P, \bar{s}} \frac{p(\bar{s})}{N!} f^2(\bar{s}, r_{P(j)}, t_{P(j)}, t'_{P(j)}, j \in s, j' \in s') \end{aligned} \quad (2.3.41)$$

But the left-hand side of (2.3.41) may be replaced by

$$\sum_{P, \bar{s}} \frac{p'(\bar{s})}{N!} \left[\sum_{j \in s} (r_{P(j)} - t_{P(j)}) + \sum_{j \in s'} t'_{P(j)} \right]^2 \quad (2.3.42)$$

for any sampling design $p'(\bar{s})$ in $D_P(n', n)$ as is clear from

the analysis above. In particular, this is so, if p' is in $D_{PF}(n', n)$.

Subtracting $\frac{1}{N!} \sum_P \left(\sum_{j=1}^N b_j r_j \right)^2$ from each side of (2.3.41) and using (2.3.35) we get

$$\sum_{P, \bar{s}} \frac{p'(\bar{s})}{N!} \left[\sum_{j \in s} (r_{P(j)} - t_{P(j)}) + \sum_{j \in s} t'_{P(j)} \right]^2 - \frac{1}{N!} \sum_P \left(\sum_{j=1}^N b_j r_j \right)^2 \leq E_F V_P(e_{\bar{s}}^-) \quad (2.3.43)$$

Finally, the left-hand side of (2.3.43) is just $E_F V_{p'}(e_{b\bar{s}}^*)$ for any design $p' \in D_{PF}(n', n)$; explicitly, $E_F V_{p'}(e_{b\bar{s}}^*)$ for every p' in $D_{PF}(n', n)$ works out as

$$\begin{aligned} & \left[\frac{n}{N} \sum_{j=1}^N u_j'^2 + \frac{n(n-1)}{N(N-1)} \sum_{j \neq k=1}^N \sum_{j \neq k=1}^N u_j' u_k' + \frac{n'}{N} \sum_{j=1}^N t_j'^2 \right. \\ & \quad + \frac{n'(n'-1)}{N(N-1)} \sum_{j \neq k=1}^N \sum_{j \neq k=1}^N t_j' t_k' + \frac{2n}{N} \sum_{j=1}^N t_j' u_j' \\ & \quad \left. + \frac{2n(n'-1)}{N(N-1)} \sum_{j \neq k=1}^N \sum_{j \neq k=1}^N u_j' t_k' \right] - Y^2 \end{aligned}$$

where $u_j' = \frac{u_j}{b_j} = \frac{y_j - x_j}{b_j} = r_j - t_j$ and $t_j' = \frac{x_j}{a_j}$.

$$\begin{aligned} \sum_{\vec{s}} p(\vec{s}) \sum_{j \in s} u_{P(j)}'^2 &= \sum_{j=1}^N u_{P(j)}'^2 \sum_1 \sum_2 p_2(s) r_1(s') \\ &= \sum_{j=1}^N u_{P(j)}'^2 b_j \quad (\text{since for every} \end{aligned}$$

$$p(\vec{s}) \in D_{PF}(n', n), \quad Q_j = b_j),$$

$$\text{and } \sum_P \frac{1}{N!} \sum_{j=1}^N u_{P(j)}'^2 b_j = \frac{(N-1)!}{N!} \sum_{j=1}^N b_j \sum_{j=1}^N u_j'^2 = \frac{n}{N} \sum_{j=1}^N u_j'^2.$$

Similarly,

$$\begin{aligned} \sum_{\vec{s}} p(\vec{s}) \sum_{j \neq k \in s} u_{P(j)}' u_{P(k)}' &= \sum_{j \neq k=1}^N \sum u_{P(j)}' u_{P(k)}' \sum_1 \sum_2 p_2(s) p_1(s') \\ &= \sum_{j \neq k=1}^N \sum u_{P(j)}' u_{P(k)}' Q_{jk}, \end{aligned}$$

$$\begin{aligned} \sum_P \frac{1}{N!} \sum_{j \neq k=1}^N u_{P(j)}' u_{P(k)}' Q_{jk} &= \frac{(N-2)!}{N!} \sum_{j \neq k=1}^N \sum C_{jk} \sum_{j \neq k=1}^N u_j' u_k' \\ &= \frac{n(n-1)}{N(N-1)} \sum_{j \neq k=1}^N \sum u_j' u_k', \end{aligned}$$

$$\begin{aligned} \sum_{\vec{s}} p(\vec{s}) \sum_{j \in s} t_{P(j)}'^2 &= \sum_{j=1}^N t_{P(j)}'^2 \sum_1 p_1(s') \\ &= \sum_{j=1}^N t_{P(j)}'^2 p_j = \sum_{j=1}^N t_{P(j)}'^2 a_j, \end{aligned}$$

$$\begin{aligned} \text{and } \sum_P \frac{1}{N!} \sum_{j=1}^N t_{P(j)}'^2 a_j &= \frac{(N-1)!}{N!} \sum_{j=1}^N a_j \sum_{j=1}^N t_j'^2 \\ &= \frac{n'}{N} \sum_{j=1}^N t_j'^2. \end{aligned}$$

Similarly,

$$\sum_P \sum_S \frac{p(\bar{s})}{N!} \left[\sum_{j \neq k \in S} t_{P(j)}' t_{P(k)}' \right] = \frac{n'(n'-1)}{N(N-1)} \sum_{j \neq k=1}^N t_j' t_k',$$

$$\begin{aligned} \sum_S p(\bar{s}) \sum_{j \in S} u_{P(j)}' t_{P(j)}' &= \sum_{j=1}^N u_{P(j)}' t_{P(j)}' \sum_1 \sum_2 p_2(s) p_1(s') \\ &= \sum_{j=1}^N u_{P(j)}' t_{P(j)}' q_j \\ &= \sum_{j=1}^N u_{P(j)}' t_{P(j)}' b_j, \end{aligned}$$

$$\begin{aligned} \sum_P \frac{1}{N!} \sum_{j=1}^N u_{P(j)}' t_{P(j)}' b_j &= \frac{(N-1)!}{N!} \sum_{j=1}^N b_j \sum_{j=1}^N u_j' t_j' \\ &= \frac{n}{N} \sum_{j=1}^N u_j' t_j', \end{aligned}$$

$$\begin{aligned} \sum_S p(\bar{s}) \sum_{\substack{j \neq k \\ j \in S, k \in S'}} u_{P(j)}' t_{P(k)}' &= \sum_{j \neq k=1}^N u_{P(j)}' t_{P(k)}' \sum_3 \sum_2 p_2(s) p_1(s') \\ &= \sum_{j \neq k=1}^N u_{P(j)}' t_{P(k)}' R_{jk}, \end{aligned}$$

$$\begin{aligned}
 \text{and } E_P \frac{1}{N!} \sum_{j \neq k=1}^N \sum_{j \neq k=1}^N u'_{P(j)} t'_{P(k)} R_{jk} \\
 = \frac{(N-2)!}{N!} \sum_{j \neq k=1}^N \sum_{j \neq k=1}^N R_{jk} \sum_{j \neq k=1}^N u'_j t'_k \\
 = \frac{n(n'-1)}{N(N-1)} \sum_{j \neq k=1}^N \sum_{j \neq k=1}^N u'_j t'_k
 \end{aligned}$$

$$\text{So, } E_F V_P (e^*_{bs}) = E_F V_D (e^*_{bs})$$

$$\begin{aligned}
 = \frac{n}{N} \sum_{j=1}^N u_j'^2 + \frac{n(n-1)}{N(N-1)} \sum_{j \neq k=1}^N \sum_{j \neq k=1}^N u'_j u'_k + \frac{n'}{N} \sum_{j=1}^N t_j'^2 \\
 + \frac{n'(n'-1)}{N(N-1)} \sum_{j \neq k=1}^N \sum_{j \neq k=1}^N t'_j t'_k + \frac{2n}{N} \sum_{j=1}^N u'_j t'_j \\
 + \frac{2n(n'-1)}{N(N-1)} \sum_{j \neq k=1}^N \sum_{j \neq k=1}^N u'_j t'_k - \left(\sum_{j=1}^N y_j \right)^2
 \end{aligned}$$

$$\text{where } u'_j = \frac{y_j - x_j}{b_j} = r_j - t_j \text{ and } t'_j = \frac{x_j}{a_j}.$$

Hence the theorem is proved. We have proved the result assuming a random permutation model without specifying any particular distribution followed by the fixed vector $V = \{(t_j, r_j, t'_j) | j=1, \dots, N\}$. So the result extends to the case when V has any exchangeable prior distribution as in Godambe/Thompson (1973).

Finally we make a truly Bayesian approach. Following Godambe (1969) and Chaudhuri (1978), we get another uniformly admissible double sampling strategy following the conventional Bayesian approach where post-survey data-analysis is independent of a design. Given a prior F for $\underline{Y}$, whatever p may be, the Bayes estimator for Y , with loss as the square error, is the posterior expectation of Y given the data

$d = (\bar{s}, x_1, y_j \mid i \in s', j \in s)$, which is

$$\begin{aligned} e_1 &= e_1(\bar{s}, \underline{Y}, \underline{X}) = E_F(Y \mid d) \\ &= \sum_{j \in s} y_j + E_F \left[\sum_{j \in s' - s} y_j + \sum_{j \notin s'} y_j \mid d \right]. \end{aligned}$$

Then, with an obvious notation for prior expectation (with respect to F) it follows that for any other estimator $e = e(\bar{s}, \underline{Y}, \underline{X})$ one has

$$E_F(e_1 - Y)^2 \leq E_F(e - Y)^2 \text{ for any } \bar{s} \text{ actually}$$

observed and surveyed and hence

$$\sum_{s'} \sum_s E_F(e_1 - Y)^2 p_2(s) p_1(s') \leq \sum_{s'} \sum_s E_F(e - Y)^2 p_2(s) p_1(s')$$

remembering that a prior on $\underline{Y}$ is postulated posterior to first phase sampling (hence using data on x_1 for $i \in s'$) and prior to second phase sampling so that F depends on $(s', x_1 \mid i \in s') = \underline{x}(s')$, say, and hence we may denote F by $F(\underline{x}(s'))$ - (to

save space often we will use F only).

Let us define two classes of priors as follows:

Definition 1. Let Ω be a class of discrete prior distributions $F = F(\underline{x}(s'))$, given first phase sample data $\underline{x}(s')$, for $\underline{y}$ defined on R^N such that for any point $\underline{y}_0$ in R^N there exists a $F_0 \in \Omega$ such that $F_0(\underline{y}_0) > 0$, this F_0 too depending on $\underline{x}(s')$ assumed already observed.

Definition 2. Let Ω' be the class of all priors $F = F(\underline{x}(s'))$ for $\underline{y}$ such that

- (i) F is discrete,
- (ii) When F -distributed, $y_1, \dots, y_N$ are independent,
- (iii) $E_F(y_i) = \begin{cases} \lambda_i & \text{if } i \notin s' \\ \mu_i(\underline{x}(s')) & \text{if } i \in s' \\ \mu_i(\text{ay}) & \text{if } i \in s' \end{cases}$

with λ_i 's and μ_i 's as some pre-assigned numbers,

$$(iv) E_F(y_i - \lambda_i)^2 = \sigma^2(>0) \text{ for } i \notin s',$$

$$(v) E_F(y_i - \mu_i)^2 = \sigma_1^2(>0) \text{ for } i \in s' (\text{independent of } \underline{x}(s')).$$

Then, we have,

Lemma 1. If $e_0 = e_0(d)$ is a Bayes estimator for Y for every prior distribution $F = F(\underline{x}(s'))$ in Ω (given $\underline{x}(s')$ assumed observed in first phase sample) then for any sampling design

$p(\bar{s})$, e_0 is admissible.

Proof. Let, if possible, another estimator $e = e(d)$ be better. Then,

$$\sum_{s'} \sum_s (e-Y)^2 p_2(s) p_1(s') \leq \sum_{s'} \sum_s (e_0-Y)^2 p_2(s) p_1(s')$$

$\forall \underline{Y} \in R^N$, with strict inequality for at least one $\underline{Y}$ in R^N .

This would hold only if

$$\sum_s (e-Y)^2 p_2(s) < \sum_s (e_0-Y)^2 p_2(s)$$

for at least one s' with $p_1(s') > 0$ and at least one $\underline{Y}$ in R^N , say, for $\underline{Y}_0$ in R^N .

Let s' be a particular first phase sample with $p_1(s') > 0$ and $F_0 = F_0(\underline{x}(s'))$ be a prior in Ω (given $\underline{x}(s')$ assumed already observed) such that $F_0(\underline{Y}_0) > 0$. Then it would follow that

$$\sum_{s'} E_{F_0} \sum_s (e-Y)^2 p_2(s) p_1(s') < \sum_{s'} E_{F_0} \sum_s (e_0-Y)^2 p_2(s) p_1(s')$$

$$\text{i.e. } \sum_{s'} \sum_s E_{F_0} (e-Y)^2 p_2(s) p_1(s') < \sum_{s'} \sum_s E_{F_0} (e_0-Y)^2 p_2(s) p_1(s')$$

since $p_2(s)$ is free of y_1 's. But this contradicts that e_0 is a Bayes estimator for Y with respect to F_0 . Hence the lemma is proved.

Lemma 2. For any point, say, $\underline{Y}_0'$ in R^N there exists a $F = F(\underline{x}(s'))$, say, $F_0 = F_0(\underline{x}(s'))$ in Ω' (given $\underline{x}(s')$ assumed already observed) such that $F_0(\underline{Y}_0') > 0$.

Proof. Let $\underline{y}'_0 = (y'_{10}, \dots, y'_{i0}, \dots, y'_{N0})$ be any point in R^N . Let, for a set of numbers $y_{i0}, y_{i1}, \dots, y_{iM}$ ($i=1, \dots, N$), with M as a suitable positive integer, the following hold, namely,

$$\begin{aligned} \text{(i)} \quad \sum_{j=0}^M y_{ij} &= (M+1) \lambda_i \quad \text{if } i \notin s' \\ &= (M+1) \mu_i \quad \text{if } i \in s' \\ \text{(ii)} \quad \sum_{j=0}^M y_{ij}^2 &= (M+1) (\sigma^2 + \lambda_i^2) \quad \text{if } i \notin s' \\ &= (M+1) (\sigma_1^2 + \mu_i^2) \quad \text{if } i \in s' \end{aligned}$$

Such a choice is possible provided σ^2 and σ_1^2 are sufficiently large so that $(M+1) (\sigma^2 + \lambda_i^2) - y_{i0}^2$ and $(M+1) (\sigma_1^2 + \mu_i^2) - y_{i0}^2$ are positive for every $i \notin s'$ and every $i \in s'$ respectively. Then it is possible to think of $(M+1)^N$ points $\underline{y}$ such that in $\underline{y} = (y_1, \dots, y_i, \dots, y_N)$ each y_i is assigned one of the values $y_{i0}, \dots, y_{iM}$ ($i=1, \dots, N$) as above. Also it is possible to choose a probability distribution F_0 on R^N such that $F_0(\underline{y}) = \frac{1}{(M+1)^N} > 0$ for each of the above-noted $(M+1)^N$ points $\underline{y}$

in R^N . Let such a prior be assigned to the earlier fixed point $\underline{Y}'_0 = (y'_{10}, \dots, y'_{i0}, \dots, y'_{N0})$. Then, $F_0(\underline{Y}'_0)$ is such that the probability that y'_{i0} assumes a value y_{ij} as above under P_{F_0} is $P_{F_0}(y'_{i0} = y_{ij}) = \frac{1}{M+1}$ for $j=0, \dots, M$ for each $i=1, \dots, N$.

Now, it follows that with respect to this particular distribution F_0 , the conditions (ii) - (v) of Definition 2 are realized and F_0 belongs to Ω' . Q.E.D.

From Lemmas 1 and 2 we have

Theorem 2.3.16. Given any sampling design p ,

$$e_1 = e_1(\bar{s}, \underline{Y}, \underline{X}) = \sum_{j \in s} y_j + \sum_{j \in s' - s} \mu_j(\underline{x}(s')) + \sum_{j \notin s'} \lambda_j$$

for arbitrarily chosen λ_j 's and μ_j 's is admissible for Y .

Proof. Since e_1 is a Bayes estimator for Y with respect to every prior $F \in \Omega'$, the theorem follows by Lemma 1.

Theorem 2.3.17. The strategy (e_1, p^*) with p^* in D^* is uniformly admissible.

Proof. Whatever design p^* in D^n may be, e_1 based on data collected through a sample $\bar{s} = (s', s)$ chosen according to p^* is a Bayes estimator for Y with respect to every prior $F \in \Omega$.

If possible, let another strategy (e', p') with p' in D^n , be uniformly better than (e_1, p^*) . That would require that

$$\sum_{s'} \sum_s (e' - Y)^2 p_2'(s) p_1'(s') \leq \sum_{s'} \sum_s (e_1 - Y)^2 p_2^*(s) p_1^*(s') \quad \forall Y \in R^N,$$

with strict inequality for at least one $\underline{Y}$ in R^N .

This would hold only if

$$(e' - Y)^2 p'(\bar{s}) < (e_1 - Y)^2 p^*(\bar{s})$$

for at least one $\bar{s}$ with $p'(\bar{s}) \geq 0$, $p^*(\bar{s}) > 0$ and at least one $\underline{Y}$ in R^N , say, for $\underline{Y}_0$ in R^N .

Now corresponding to $\underline{Y}_0$ in R^N , there exists a $F_0 \in \Omega'$ such that $F_0(\underline{Y}_0) > 0$ (by Lemma 2) for a particular s' , supposed drawn actually, and $\underline{x}(s')$ actually observed such that $p'(\bar{s}) \geq 0$ and $p^*(\bar{s}) > 0$.

Then it would follow that

$$\sum_{s'} \sum_s E_{F_0} (e' - Y)^2 p'(\bar{s}) < \sum_{s'} \sum_s E_{F_0} (e_1 - Y)^2 p^*(\bar{s}).$$

Now whatever p in D^n may be according to which a sample $\bar{s} = (s', s)$ is chosen yielding the statistic e_1 , we have

$$E_{F_0} (e_1 - Y)^2 = E_{F_0} (Y - e_1)^2$$

$$= E_{F_0} \left[\sum_{j \in S' - S} y_j + \sum_{j \notin S'} y_j - \sum_{j \in S' - S} \mu_j - \sum_{j \notin S'} \lambda_j \right]^2$$

$$= V_{F_0} \left[\sum_{j \in S' - S} y_j + \sum_{j \notin S'} y_j \right] \quad (V_{F_0} \text{ denoting prior variance with respect to } F_0)$$

$$= [n(s') - n(s)] \sigma_1^2 + [N - n(s')] \sigma^2.$$

$$\text{Then, } \sum_{s'} \sum_s E_{F_0} (e_1 - Y)^2 p^*(\bar{s})$$

$$= \sum_{s'} \sum_s E_{F_0} (e_1 - Y)^2 p(\bar{s}) \quad (\text{for every } p \in D'')$$

$$= \sum_{s'} \sum_s E_{F_0} (e_1 - Y)^2 p'(\bar{s}), \quad \text{in particular for } p = p'.$$

Hence we would have

$$\sum_{s'} \sum_s E_{F_0} (e_1 - Y)^2 p'(\bar{s}) < \sum_{s'} \sum_s E_{F_0} (e_1 - Y)^2 p'(\bar{s})$$

contradicting that e_1 is a Bayes estimator for Y with respect to F_0 when it is based on data collected through a sample chosen according to any design in D'' .

This proves the theorem.

CONCLUDING REMARKS CONCERNING CHOICE OF TWO-PHASE STRATEGIES

In respect of two-phase sampling we have considered certain 'optimal' strategies available under certain circumstances. But in practice it is difficult to check when these circumstances are plausible. So, an alternative strategy may fare better than a so-called 'optimal' one in a situation not favouring the latter. So, various alternatives 'may' and 'should' be tried out in actual surveys. It may be of interest to note the following (applicable to double sampling):

Suppose in the first phase an SRSWOR, say, s' is chosen and one contemplates to use the first-phase survey data on x_j for $j \in s'$ in choosing the second-phase sub-sample s from s' and estimating Y directly from second-phase sample using y_j values for j in s . For example, three alternatives may be on using the normed size-measures $q_j = x_j / \sum_{j \in s'} x_j$ ($\forall j \in s'$) to

(1) take an IPPS (inclusion-probability proportional to q_j 's for second-phase sampling design p_2) sample s and use the Horvitz/Thompson (1952), (HT, say) estimator

$$e_1 = \frac{N}{n'} \cdot \frac{1}{n} \sum_{j \in s} \frac{y_j}{q_j} \text{ or}$$

(2) take s using q_j 's and employing Rao/Hartley/Cochran (1962), (RHC, say) scheme and employing their estimator viz. $e_2 = \frac{N}{n} \sum_s \frac{y_j}{q_j} S_j$, (where S_j 's are sums of q_j -values over units falling in particular groups formed in applying RHC scheme or (3) apply Midzuno/Sen (1952,1953), (MS, say) scheme using q_j 's to choose s and employ Lahiri's (1951) procedure (LMS, in brief) to use the estimator

$$e_3 = \frac{N}{n} \cdot \frac{\sum_s y_j}{\sum_s q_j}.$$

We postulate a linear (through the origin) regression model for y on x as follows:

$$\left. \begin{aligned} E_M(y_i) &= \rho x_i, \quad E_M(y_i - \rho x_i)(y_j - \rho x_j) = \sigma^2 x_i^2 \quad \text{if } i=j \\ &= 0 \quad \text{if } i \neq j \end{aligned} \right\}$$

where ρ , σ are any positive numbers and $0 \leq g \leq 2$, as usual. We may recall that under the above model the model-expected variances of HT estimator based on IPPS design, RHC estimator and LMS estimator each based on common size-measures and a common sample size in uni-phase sampling, denoted as M_1 , M_2 , M_3 (say, respectively) satisfy the relations

$$\left. \begin{array}{lll} M_1 < M_2 < M_3 & \text{if } g > 1 \\ = & = & \text{if } g = 1 \\ > & > & \text{if } g < 1 \end{array} \right\}$$

(Vide Chaudhuri/Arnab (1979)).

Noting that the variances of the above three estimators e_i ($i = 1, 2, 3$) can be written as

$$\begin{aligned} V_{p_1}(e_i) &= E_{p_1} V_{p_2}(e_i) + V_{p_1} E_{p_2}(e_i) \\ &= E_{p_1} V_{p_2}(e_i) + V_{p_1} \left(\frac{N}{n}, \sum_{j \in S} y_j \right) \end{aligned}$$

where the second term is common to all the variances of the estimators e_i ($i = 1, 2, 3$) and E_{p_1} is the same operator on $V_{p_2}(e_i)$, $i = 1, 2, 3$, the model-expected variances denoted as M_1^*, M_2^*, M_3^* (say, respectively) satisfy the relations

$$\left. \begin{array}{lll} M_1^* < M_2^* < M_3^* & \text{if } g > 1 \\ = & = & \text{if } g = 1 \\ > & > & \text{if } g < 1. \end{array} \right\}$$

$$\begin{aligned} \text{Here } M_1^* &= \sigma^2 \left[\frac{x}{n} \cdot \frac{N(n'-1)}{n'(N-1)} \sum_{i=1}^N x_i^{g-1} - \frac{N}{\sum_{i=1}^N x_i^g} \right] \\ &+ \frac{N}{n'} \cdot \frac{N-n'}{N-1} \cdot \frac{\sigma^2}{n} \sum_{i=1}^N x_i^g \\ &+ N^2 \left(\frac{1}{n'} - \frac{1}{N} \right) \rho^2 S_x^2, \end{aligned}$$

$$\begin{aligned} M_2^* &= \frac{N}{n'} \cdot \frac{n'-n}{N-1} \frac{\sigma^2}{n} \left[x \sum_{i=1}^N x_i^{g-1} - \frac{N}{\sum_{i=1}^N x_i^g} \right] \\ &+ \frac{N-n'}{n'} \sigma^2 \sum_{i=1}^N x_i^g + N^2 \left(\frac{1}{n'} - \frac{1}{N} \right) \rho^2 S_x^2, \end{aligned}$$

and

$$\begin{aligned} M_3^* &= \frac{N^2}{n'^2} \sigma^2 E_{p_1} \left[\sum_{j \in s'} x_j \cdot \frac{1}{\left(\frac{n'-1}{n} \right)} \sum_{s \subset s'} \frac{\sum_{i \in s} x_i^g}{\sum_{i \in s} x_i} \right] \\ &- \sigma^2 \sum_{i=1}^N x_i^g + N^2 \left(\frac{1}{n'} - \frac{1}{N} \right) \rho^2 S_x^2 \end{aligned}$$

$$\text{where } S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})^2, \quad \bar{X} = \frac{1}{N} \sum_{i=1}^N x_i.$$

Thus we get an exactly analogous relative standings for e_1, e_2 and e_3 . However, it may be conceded that in case the first-phase sample is chosen with varying probabilities we are unable to suggest concrete but simple double-sampling

strategies with second-phase sampling design utilizing the realized x_1 -values in the first-phase sample.

2.5 UNBIASED ESTIMATION OF VARIANCES FOR ESTIMATORS IN TWO-PHASE SAMPLING-

Writing $\pi_j = \Sigma_2 p_2(s) = nq_j$ and $\pi_{jk} = \Sigma_4 p_2(s)$ we get an unbiased estimator for $V_p(e_1)$ as follows:

$$\hat{V}_p(e_1) = e_1^2 - \frac{N}{n} \left[\Sigma_{j \in s} \frac{y_j^2}{\pi_j} + \frac{N-1}{n-1} \Sigma_{j \neq k \in s} \frac{y_j y_k}{\pi_{jk}} \right].$$

Assuming that the groups formed at random for applying RHC scheme are of constant size viz. n'/n (assumed to be an integer) we have an unbiased estimator for $V_p(e_2)$ as follows:

$$\begin{aligned} \hat{V}_p(e_2) = & \frac{N}{n'(n'-1)} \left[\left(1 - \frac{n}{n'}\right) \cdot \frac{N-1}{n-1} \Sigma_{j \in s} \frac{y_j^2}{q_j} S_j \right. \\ & \left. + (N-n') \Sigma_{j \in s} \frac{y_j^2}{q_j} S_j \right] - \frac{(N-n)}{N(n-1)} e_2^2. \end{aligned}$$

Finally, following Raj (1954), we get an unbiased estimator for $V_p(e_3)$ as follows:

$$\hat{V}_p(e_3) = e_3^2 - \frac{N}{n' \Sigma_{j \in s} q_j} \left[\Sigma_{j \in s} y_j^2 + \frac{N-1}{n-1} \Sigma_{j \neq k \in s} y_j y_k \right]$$

These estimators may take negative values; but we are unable to suggest uniformly non-negative alternatives.

CHAPTER - 3

ON THE PERFORMANCES OF SOME SINGLE-PHASE

SAMPLING STRATEGIES

1 SUMMARY. In uni-phase sampling we may observe the following concerning the choice of strategies. Some of the popular strategies are : (i) Horvitz/Thompson (1952) estimator based on a fixed sample-size IPPS design; (ii) Rao/Hartley/Cochran (1962) strategy; (iii, iv) the ratio estimator based on SRSWOR or on Midzuno/Sen (1952,1953) scheme of sampling, (v) sample mean of ratios of y_i 's for SRSWOR, (vi) Hansen/Hurwitz (1943) strategy, (vii, viii) Raj's (1956) and Murthy's (1957) estimators based on PPSWOR sampling. Analytical comparisons of the first six with the last two seem difficult because the variances of the latter are complicated. So, we will consider the first six only and later we will consider some variants of them derived on certain modifications. First we consider linear regression (not necessarily through the origin) models with identically independently gamma-distributed regressors and uncorrelated errors with zero means but conditional variances as functions of the regressors. The model-expected biases and mean square errors (MSE) of the above six strategies

are then compared - analytically and numerically. Allowing more generality on the distribution of regressor variates but assuming existence of second order moments certain asymptotic results also have been derived under assumption of large population and/or sample sizes. In these derivations law of large numbers has been applied on the x_1 's but not on the errors. Because, the latter are not assumed independently or identically distributed and about the conditional error variances no assumptions are made to ensure validity of law of large numbers (LLN). However in case LLN is also assumed to hold for the errors certain modifications on the relative standings among the strategies are also noted. Comparatively, the Lahiri/Midzuno/Sen (1951, 1952, 1953) (LMS) strategy is observed to stand out as a very promising one under several circumstances not favourable to the HT strategy studded with many optimal properties. So, next we consider this strategy vis-a-vis the strategies involving selection by SRSWOR method and using ratio estimator and a few derivatives of it based on half-sampling. Under similar models again, LMS strategy figures impressively in this comparative study too. In this investigation the status of regression estimator based on SRSWOR is difficult to ascertain because it does not admit a simple model-

expected MSE expression except when errors are homoscedastic.

Singh/Srivastava (1980) have given a sampling scheme yielding an unbiased regression estimator and an unbiased regression-type strategy. By model sampling from bivariate normal populations they have demonstrated superiority of their unbiased regression strategy to biased regression estimation through SRSWOR sampling. But, considering more general models involving linear regression with homoscedasticity of errors we identify situations giving varying levels of efficiencies for them under different circumstances.

In case the postulated regression line passes through the origin assuming x_1 's as fixed constants, results concerning relative efficiencies of various strategies are available in the literature. If, however, the regression line has a non-zero intercept α (say) with an unknown magnitude then if one may have some idea about its magnitude, it seems plausible to modify the available strategies using origin-shifted y 's rather than the y 's themselves along with other necessary adjustments. That this yields higher efficiencies on suitable occasions is illustrated with reference to some standard strategies and other modifications. In addition, relative efficiencies of several popular

strategies modified with variate transformations, are worked out with postulated identically independently gamma distributed regressors.

Remark. From our discussions in the previous Chapter it was clear that appropriateness of an over-all double sampling strategy depends on those of the uni-phase ones at the two separate phases. Intuitively too this should be obvious. So, in double sampling context it is important to pay heed to the question of choosing good single phase strategies.

3.2.1 RELATIVE EFFICIENCIES OF SEVERAL STRATEGIES FOR ESTIMATING A FINITE POPULATION MEAN

Following Chakrabarti (1973) and Chaudhuri (1979) we compare the relative efficiencies of six well-known sampling strategies under models more general than those considered by them. Exact (small sample) as well as asymptotic results are presented along with numerical ones.

Pursuing with Chakrabarti's (1973) and Chaudhuri's (1979) works we shall consider the following generalization of their models for estimating the mean $\bar{Y} = \frac{1}{N} \sum_{i=1}^N y_i$

of a finite population on the i th unit of which a real variate y assumes the value y_i ($i = 1, \dots, N$):

$$y_i = \alpha + \beta x_i + u_i \quad (i = 1, \dots, N) \quad (3.2.1)$$

where $\alpha \geq 0$, $\beta \geq 0$ and u_i 's are random variables with conditional expectations

$$\varepsilon(u_i | x_i) = 0, \quad \varepsilon(u_i^2 | x_i) = \delta x_i^g \quad \forall i,$$

$$(\text{with } 0 < \delta < \infty, \quad 0 \leq g \leq 2),$$

$$\varepsilon(u_i u_j | x_i, x_j) = 0 \quad \forall i \neq j,$$

where x_i 's are known positive quantities which are assumed to be realizations on random variables (also denoted as x_i 's), with incompletely specified distributions. Sometimes (as e.g. in section 3.2.2) we assume the auxiliary variables x_i 's to be identically and independently gamma-distributed with the common density

$$f(v) = \frac{1}{\Gamma(m)} e^{-v} v^{m-1}, \quad v > 0,$$

with a single parameter namely the mean m (supposed to exceed 2) which we take $\lfloor$ on the strength, if needed for logicality, of the law of large numbers which may be

supposed to be applicable provided we are ready to assume N to be large, as we are, following Chakrabarti (1973)] as equal to $\bar{X} = \frac{1}{N} \sum_{i=1}^N x_i$ which is known for the given finite population. This model will be denoted as I. In section 3.2.3 the forms of the distributions of x_i 's are left unspecified except that they are assumed to be distributed independently and identically with a common mean $m = \bar{X}$ as above and their common moments μ_r' of orders r exist for r upto g ($0 \leq g \leq 2$). The model in this case will be denoted as II. Our objective here is to compare the biases and mean square errors of six well-known sampling strategies under the models stipulated above, which are slight generalizations of those due to Chakrabarti (1973) and Chaudhuri (1979) in the sense that unlike them we allow g to be non-zero. Following Chaudhuri (1979), in section 3.2.3 we present a few asymptotic results assuming both the sample and the population sizes to be large. Our results differ from similar others available in the literature because of the peculiarities following from the assumption $m = \bar{X}$ which introduces a novelty suggested initially by Chakrabarti (1973).

Our notations are as in Chaudhuri (1979): by s is meant a sample (typically, the sample-size being taken throughout

as a fixed positive integer n), its selection - probability is $p(s)$ for a design p (generically), and the expectation - operator for the design is E . By ϵ we denote conditional expectation over the error term u (with fixed x) in the model (3.2.1), ϵ_x denotes expectation over the distribution of x (standing for x_i , $i = 1, \dots, N$); also $\bar{\epsilon} = \epsilon_x \epsilon$ is the two-step expectation (for u with x fixed and then over x) and $e = \bar{\epsilon} E = \epsilon_x \epsilon E$ is the three-step (including the one over the sampling design p) expectation - operator.

The six strategies we consider will be numbered consecutively as 1, ..., 6 which respectively involve the estimators $t_1 = \bar{X} \cdot \frac{\bar{Y}}{\bar{x}}$ ($\bar{y}$, $\bar{x}$ denoting the sample means for y and x), $t_2 = \bar{X} \frac{1}{n} \sum_{i \in s} \frac{y_i}{x_i}$, both based on the SRSWOR scheme, $t_3 = t_1$ based on Midzuno/Sen/Lahiri (1952) scheme, the Horvitz/Thompson (1952) estimator $t_4 = \frac{1}{N} \sum_{i \in s} \frac{y_i}{\pi_i}$, based on a π ps design with inclusion - probability $\pi_i = n p_i \forall i$ (say), ($p_i = \frac{x_i}{N\bar{x}}$), and t_5, t_6 the usual Rao/Hartley/Cochran (RHC, in brief) (1962), (for this we assume $\frac{N}{n} = k$ to be an integer and each group formed in applying this scheme is of size k), and Hansen/Hurwitz (1943) estimators (HHE), in brief, both involving normed size - measures p_i 's.

The expectations, biases and mean square errors of the estimators will be denoted respectively as

$$e_i = e(t_i) = \bar{e} E(t_i) = e_x e E(t_i)$$

$$B_i = e(t_i - \bar{y}) = e_x e [E(t_i) - \bar{y}],$$

$$M_i = \bar{e} [E(t_i - \bar{y})^2] = e_x e [E\{E(t_i - \bar{y})^2\}], i=1, \dots, 6$$

(in case $\alpha = 0$ in the model (3.2.1) we shall write M'_i for M_i , $i = 1, \dots, 6$).

2.2 EXACT (SMALL SAMPLE) FORMULAE CONCERNING RELATIVE EFFICIENCIES OF SAMPLING STRATEGIES UNDER MODEL I WITH GAMMA-DISTRIBUTED AUXILIARY VARIATES.

Omitting easily verifiable algebraic steps, to save space, and using Rao and Webster's (1966) lemma concerning properties of gamma distributions we get

$$B_1 = \frac{\alpha}{n m - 1}, \quad B_2 = \frac{\alpha}{m - 1}, \quad \text{so that } B_2 > B_1;$$

of course $B_i = 0$ for $i = 3, \dots, 6$.

However if m is not assumed to be equal to $\bar{x}$, then writing B'_1 and B'_2 for biases of t_1 and t_2 respectively

we have

$$B_1' = (1 - \frac{n}{N}) \frac{\alpha}{nm-1} \quad \text{and} \quad B_2' = (1 - \frac{1}{N}) \frac{\alpha}{m-1}.$$

Then, we have,

$$M_1 = \alpha^2 \cdot \frac{(nm+2)}{(nm-1)(nm-2)} + \delta nm^2 \cdot \frac{\Gamma(m+g)}{\Gamma(m)} \cdot \frac{1}{(nm+g-1)(nm+g-2)} \\ + \frac{\delta}{N} \cdot \Gamma \cdot \frac{\Gamma(m+g)}{\Gamma(m)} \left\{ 1 - 2 \frac{nm}{nm+g-1} \right\} \rfloor$$

$$M_2 = \alpha^2 \cdot \Gamma \cdot \frac{m^2}{n} \left\{ \frac{1}{(m-1)(m-2)} + \frac{n-1}{(m-1)^2} \right\} - \frac{2m}{m-1} + 1 \rfloor \\ + \frac{\delta m^2}{n} \cdot \frac{\Gamma(m+g-2)}{\Gamma(m)} + \frac{\delta}{N} \cdot \Gamma \cdot \frac{\Gamma(m+g-1)}{\Gamma(m)} (g-m-1) \rfloor$$

$$M_3 = \frac{\alpha^2}{nm-1} + \delta m \cdot \frac{\Gamma(m+g)}{\Gamma(m)} \cdot \frac{1}{(nm+g-1)} - \frac{\delta}{N} \cdot \frac{\Gamma(m+g)}{\Gamma(m)}.$$

In calculating M_4 we neglect the term $\alpha^2 \sum_{i \neq j=1}^N \sum \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right)$

and thus get a conservative expression on assuming

$$\pi_{ij} < \pi_i \pi_j \quad \forall i, j$$

[which is a desirable condition on π ps designs to ensure existence of non-negative variance estimator for t_4 and which is satisfied for many sampling schemes, vide Lanke(1975), e.g.]

$$M_4 = \alpha^2 \left[\frac{m}{n(m-1)} - \frac{1}{N} \right] + M_4', \text{ where}$$

$$M_4' = \frac{\delta m}{n} \frac{\Gamma(m+g-1)}{\Gamma(m)} - \frac{\delta}{N} \cdot \frac{\Gamma(m+g)}{\Gamma(m)},$$

$$M_5 = \frac{N-n}{N-1} \cdot \frac{1}{n} \left[\frac{\alpha^2}{m-1} + \delta m \frac{\Gamma(m+g-1)}{\Gamma(m)} - \frac{\delta}{N} \cdot \frac{\Gamma(m+g)}{\Gamma(m)} \right]$$

$$M_6 = \frac{N-1}{N-n} M_5.$$

Noting the complicated form of the co-efficient of α^2 in M_2 we assume m to be large and neglecting terms $O(1/m^2)$ for the sake of sheer simplicity in it (only) we approximate this co-efficient by $\frac{(m+2)}{(m-1)(m-2)}$ and the quantity M_2 is approximated by

$$\begin{aligned} \bar{M}_2 = \alpha^2 \cdot \frac{(m+2)}{(m-1)(m-2)} + \frac{\delta}{n} m^2 \frac{\Gamma(m+g-2)}{\Gamma(m)} \\ + \frac{\delta}{N} \cdot \frac{1}{\Gamma(m)} \left[\Gamma(m+g) - 2m \Gamma(m+g-1) \right] \end{aligned}$$

3 ASYMPTOTIC FORMULAE CONCERNING RELATIVE EFFICIENCIES OF STRATEGIES UNDER MODEL II

As in Cochran (1963), in what follows, we shall assume N and n large to such an extent that we may neglect the error

in writing $\bar{x} = \bar{X}$ for every sample s with $p(s) > 0$ for the strategies 1 and 3. For other strategies also N and n will be taken to be large whenever necessary (as discussed below) in comparing the efficiencies. For the model II, we have the following formulae on omitting details of calculations, viz.

$$M_1 = \frac{N-n}{Nn} \cdot \frac{1}{N-1} \left[N \sigma^2 \mu_2 + \delta (N-1) \mu_g' \right],$$

$$\text{where } \mu_2 = \mu_2' - \mu_1'^2,$$

$$M_2 = \alpha^2 \left[\frac{m^2}{n} \left\{ \varepsilon_x \left(\frac{1}{x^2} \right) + (n-1) \varepsilon_x^2 \left(\frac{1}{x} \right) \right\} - 2m \cdot \varepsilon_x \left(\frac{1}{x} \right) + 1 \right] \\ + \frac{\delta m^2}{n} \mu_{g-2}' + \frac{\delta}{N} \mu_g' - 2\delta \frac{m}{N} \mu_{g-1}',$$

$$M_3 = \delta \left(\frac{1}{n} - \frac{1}{N} \right) \mu_g'$$

On neglecting the term $\alpha^2 \sum_{i \neq j} \sum \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right)$ as in section 3.2.2,

$$M_4 = \alpha^2 \left[\frac{m}{n} \varepsilon_x \left(\frac{1}{x} \right) - \frac{1}{N} \right] + M_4',$$

$$\text{where } M_4' = \delta \left[\frac{m}{n} \mu_{g-1}' - \frac{1}{N} \mu_g' \right],$$

$$M_5 = \frac{N-n}{N-1} \cdot \frac{1}{n} \left[\alpha^{2(m \cdot \varepsilon_x (\frac{1}{x}) - 1)} + \delta m \mu_{g-1}' - \frac{\delta}{N} \mu_g' \right]$$

$$M_6 = \frac{N-1}{N-n} M_5.$$

RELATIVE EFFICIENCIES OF THE STRATEGIES

With simple algebra (suppressing details to save space) noting inter alia that $\text{cov}(x, x^{g-1}) \geq 0$ as $g \geq 1$, $\text{cov}(x, x^{g-2}) \leq 0$ since $g \leq 2$, $\mu_1'^2 > \mu_1'$, since $\mu_1' = m > 2$, writing $\frac{1}{N-1} \simeq \frac{1}{N}$, (when necessary) etc. we can derive the results presented in Table 3.2.4.1. Using formulae in section 3.2.2 we also present results of numerical investigations in Table 3.2.4.2.

Table 3.2.4.1: showing relative performances of the strategies with variation in g (section I of the table represents results under Model I and section II those for Model II)

$g < \frac{1}{2}$	$g \leq 1$	$g \leq g_0$	$1 \leq g \leq 1.5$
$M_3 < M_1$	$M_1' \leq M_4'$		
$M_1 < M_4$	$M_3 \leq M_4$		
	$M_5 \leq M_4$	$M_1' \leq M_6'$	$M_1 < \bar{M}_2$
	$M_3' \leq M_5' \leq M_4'$	(for large N)	
	$M_3 \leq M_5$	$M_1 \leq M_6$	
	(if $N > nm$)	(if $n \geq 5$)	
	$M_3 \leq M_6$		
	$M_3' \leq M_1'$		
	$\bar{M}_2 \leq M_6'$		
	$M_3 \leq \bar{M}_2$		

I

$g \leq g_0$	$g \geq 1$	$0 \leq g \leq 2$
$M_6' \leq M_1'$	$M_4' \leq M_1'$	$M_4' \leq M_2'$
(for large N)	$M_4' \leq M_2'$	(if $N > 2n$)
	$M_4' \leq M_5' \leq M_3'$	$M_4' \leq M_6'$
	$M_6' \leq M_3'$	
	(for large N)	
	$M_6' < \bar{M}_2'$	
	$\bar{M}_2 \leq M_3$	

Table 3.2.4.1 (continued).

	$g \leq 1$	$g \geq 1$	$0 \leq g \leq 2$
	$M_6' \leq M_1'$	$M_3 \leq M_1$	$M_6' \leq M_2'$
	(neglecting terms $O(1/N)$)	$M_4' \leq M_1'$	(neglecting terms $O(1/N)$)
	$M_3' \leq M_2'$	$M_5' \leq M_1'$	$M_4' \leq M_6'$
II	(if $N > 2n$)	$M_5' \leq M_2'$	
	$M_3' \leq M_5' \leq M_4'$	(neglecting terms $O(1/N^2)$)	
	$M_3 \leq M_4$	$M_4' \leq M_2'$	
	$M_3 \leq M_5$	$M_4' \leq M_5' \leq M_3'$	
	(if $\frac{n-1}{N-1} \approx 0$)		
	$M_3 \leq M_6$		

where g_0 is a root in $[0, 2]$ of $g^2 - (n^2m - 2nm + 3)g + (n^2m - 3nm + 2) = 0$. If g less than 0 is permitted, then $M_1' \leq M_2'$ if $\rho = 0$.

Table 3.2.4.2. showing relative efficiencies of strategies for certain parametric combinations using formulae in section 3.2.2 (writing $E_1=100 \frac{\bar{M}_2}{\bar{M}_1}$ or $100 \frac{M_2'}{M_1'}$ in case $\alpha = 0$)

			E_1	E_2	E_3	E_4	E_5	E_6
I	$\alpha = 0.5$	$g = 0.0$	211	100	236	176	200	168
	$\delta = 2.0$	$g = 0.5$	162	100	176	152	163	137
	$m = 5$	$g = 1.0$	129	100	138	134	138	116
	$n = 4$	$g = 1.5$	107	100	112	121	120	101
	$N = 20$	$g = 2.0$	92	100	94	111	107	90
II	$\alpha = 1.5$	$g = 0.0$	213	100	229	111	209	193
	$\delta = 2.0$	$g = 0.5$	155	100	164	114	156	144
	$m = 8$	$g = 1.0$	124	100	128	112	128	118
	$n = 4$	$g = 1.5$	106	100	108	108	113	104
	$N = 40$	$g = 2.0$	94	100	95	103	103	95
III	$\alpha = 0$	$g = 0.0$	192	100	206	157	180	100
	$\delta = 2.5$	$g = 0.5$	156	100	167	146	157	87
	$m = 6$	$g = 1.0$	131	100	140	140	140	78
	$n = 5$	$g = 1.5$	113	100	121	138	128	71
	$N = 10$	$g = 2.0$	100	100	107	140	119	66

			E_1	E_2	E_3	E_4	E_5	E_6
IV	$\alpha = 0.5$	$g = 0.0$	204	100	227	172	192	177
	$\delta = 2.0$	$g = 0.5$	157	100	170	148	157	145
	$m = 5$	$g = 1.0$	126	100	133	130	133	123
	$n = 4$	$g = 1.5$	105	100	108	116	116	107
	$N = 40$	$g = 2.0$	90	100	91	105	104	96
V	$\alpha = 0$	$g = 0.0$	167	100	179	151	154	138
	$\delta = 2.5$	$g = 0.5$	139	100	146	135	136	122
	$m = 6$	$g = 1.0$	118	100	123	123	123	110
	$n = 5$	$g = 1.5$	102	100	105	113	112	100
	$N = 40$	$g = 2.0$	90	100	91	105	103	92

Remarks.

- I. From the numerical data we see that $M_1 > M_3 \forall g$ and $\bar{M}_2 > M_3$ unless $g = 2$. But we are unable to establish general results about M_1 versus M_2 versus M_3 .
- II. Relative performances of these strategies under other alternative models are well-known and a relevant recent reference is Chaudhuri (1979).
- III. One may, in practice, choose among these six strategies with reference to the above according to situations believed to be in hand.

IV. From the Table 3.2.4.2 presenting numerical values one may gain some insight into the relative performances of the strategies under certain particular combinations of the values of the parameters for the Model I; e.g. for $g < 1$, strategy 3 is found to fare best amongst all, for $g = 1$, strategy 5 is as good and strategy 4 is also on a par if $\alpha = 0$ - consistently with theories developed here and elsewhere.

V. In the derivations of mean square errors law of large numbers has been applied on the x_1 's (on the strength of which we have equated $\bar{X}$ to m) but not on the errors u_1 's because the latter are not assumed independently or identically distributed and about the conditional error variances no assumption was made to ensure validity of law of large numbers. However in case LLN is also assumed to hold for the errors u_1 's so that we can equate $\bar{U} = \frac{1}{N} \sum_{i=1}^N u_i$ to 0 (and hence $\bar{Y}$ to $\alpha + \beta m$) certain modifications on the mean square errors of the estimators (assuming N very large, practically infinite) are noted (writing M_i^* for $E(t_i - \bar{Y})^2$ in the present situation) as follows:

$$1^* = \alpha^2 \cdot \frac{nm+2}{(nm-1)(nm-2)} + \delta m^2 \frac{\Gamma(m+g)}{\Gamma(m)} \cdot \frac{1}{(nm+g-1)(nm+g-2)},$$

$$2^* = \alpha^2 \left[\frac{m^2}{n} \left\{ \frac{1}{(m-1)(m-2)} + \frac{n-1}{(m-1)^2} \right\} - \frac{2m}{m-1} + 1 \right] \\ + \frac{\delta m^2}{n} \frac{\Gamma(m+g-2)}{\Gamma(m)},$$

$$3^* = \frac{\alpha^2}{nm-1} + \delta m \frac{\Gamma(m+g)}{\Gamma(m)} \cdot \frac{1}{(nm+g-1)},$$

in neglecting the term $\alpha^2 \sum_{i=j} \sum \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right)$ as earlier,

$$4^* = \alpha^2 \cdot \frac{m}{n(m-1)} + \frac{\delta m}{n} \frac{\Gamma(m+g-1)}{\Gamma(m)}$$

$$5^* = \frac{\alpha^2}{n(m-1)} + \frac{\delta m}{n} \frac{\Gamma(m+g-1)}{\Gamma(m)}$$

$$2^* = \alpha^2 \cdot \frac{(m+2)}{(m-1)(m-2)} + \frac{\delta}{n} m^2 \frac{\Gamma(m+g-2)}{\Gamma(m)}$$

(in case $\alpha = 0$, we shall write M_i^* for M_i^* , $i = 1, \dots, 5$)

The Table 3.2.4.3 below shows certain modifications on the relative standings of the strategies when LLN is also assumed to hold on the errors u_i 's. Using the above formulae we also present results of numerical investigation in Table 3.2.4.4.

Table 3.2.4.3. showing relative performances of the strategies with variation in g (when LLN is also applied on the u_i 's)

$g \leq \frac{1}{2}$	$g \leq 1$	$g \geq 1$	$0 \leq g \leq 2$
$M_1^* \leq M_4^*$	$M_3^* \leq M_4^*$	$M_5^{**'} \leq M_3^{**'}$	$M_3^* \leq M_1^*$
(if $n \geq 5$)	$M_3^{**'} \leq M_4^{**'}$	$M_4^{**'} \leq M_1^{**'}$	$M_3^{**'} \leq M_1^{**'}$
$M_1^{**'} \leq M_4^{**'}$	$M_3^* \leq M_5^*$		$M_5^* \leq M_4^*$
(if $n \geq 4$)	$M_3^{**'} \leq M_5^{**'}$		$M_5^{**'} = M_4^{**'}$
	$M_3^* \leq \bar{M}_2^*$		$M_4^{**'} \leq M_2^{**'}$
	$M_3^{**'} \leq M_2^{**'}$		
	$M_1^* \leq \bar{M}_2^*$		
	$M_1^{**'} \leq M_2^{**'}$		

Remark. Thus, strategy 3 is again good in relevant situations but the relative positions for 3,4,5 cannot be made explicit in this case.

Table 3.2.4.4. showing relative efficiencies of strategies under certain parametric combinations (writing $E_i^* = 100 \frac{\bar{M}_2^*}{M_i^*}$ or $100 \frac{M_2^*}{M_i^*}$ in case $\alpha = 0$).

			E_1^*	E_2^*	E_3^*	E_4^*	E_5^*
I	$\alpha = 0.5$	$g = 0.0$	198	100	220	169	185
	$\delta = 2.0$	$g = 0.5$	153	100	166	146	153
	$m = 5$	$g = 1.0$	124	100	130	127	130
	$n = 4$	$g = 1.5$	103	100	106	112	113
		$g = 2.0$	88	100	88	100	101
II	$\alpha = 1.5$	$g = 0.0$	205	100	220	107	199
	$\delta = 2.0$	$g = 0.5$	151	100	159	111	151
	$m = 8$	$g = 1.0$	121	100	125	110	125
	$n = 4$	$g = 1.5$	104	100	106	105	111
		$g = 2.0$	93	100	93	100	101
III	$\alpha = 0$	$g = 0.0$	162	100	174	150	150
	$\delta = 2.5$	$g = 0.5$	136	100	143	133	133
	$m = 6$	$g = 1.0$	116	100	120	120	120
	$n = 5$	$g = 1.5$	101	100	102	109	109
		$g = 2.0$	89	100	89	100	100

			E_1^*	E_2^*	E_3^*	E_4^*	E_5^*
IV	$\alpha = 2.5$	$g = 0.0$	475	100	548	182	462
	$\delta = 2.0$	$g = 0.5$	334	100	372	170	336
	$m = 5$	$g = 1.0$	223	100	239	152	234
	$n = 4$	$g = 1.5$	152	100	157	131	166
		$g = 2.0$	110	100	110	112	125
V	$\alpha = 0$	$g = 0.0$	178	100	198	167	167
	$\delta = 3.5$	$g = 0.5$	143	100	155	143	143
	$m = 5$	$g = 1.0$	119	100	125	125	125
	$n = 4$	$g = 1.5$	101	100	104	111	111
		$g = 2.0$	87	100	87	100	100

Remark. The numerical values presented in Table 3.2.4.4. conform to the algebraic results presented in Table 3.2.4.3. We find that $M_1^* > M_3^* \forall g$ and $\bar{M}_2^* > M_3^*$ unless $g = 2$; if $g = 2$, then $\bar{M}_2^* < M_3^*$. For $g < 1$; strategy 3 is found to fare best amongst all and for $g = 1$, the strategies 3, 4 and 5 are equivalent if $\alpha = 0$ - consistently with Chaudhuri/ Arnab's (1979) findings. For $g > 1$, (unlike Chaudhuri/ Arnab (1979)), the strategy 5 is found to fare better than strategy 4 which is again better than strategy 3.

3.3.1 ON THE EFFICIENCY OF MIDZUNO/SEN STRATEGY RELATIVE TO SEVERAL RATIO-TYPE ESTIMATORS UNDER A MODEL

Efficiency of Midzuno/Sen strategy relative to a few estimators based on SRSWOR is studied under a postulated linear regression model with uncorrelated error and a gamma-distributed auxiliary character.

Let y_i 's and x_i 's ($>0 \forall i$) be the values of positively correlated variables y and x defined on a finite population of size N . We consider estimating the population mean $\bar{Y}$ using the ratio-estimator based on the well known Midzuno/Sen (1952, 1953) scheme (to be called MSS - the Midzuno/Sen strategy) and comparing its efficiency relative to the ratio and regression estimators based on SRSWOR and a few ratio-type estimators based on random half-sampling of an SRSWOR (sample size n in each case is taken as an even number). In facilitating a comparison, following our predecessors [e.g., Durbin (1959), Rao (1969a), Rao (1969b), Rao/Rao (1971), Rao (1974), Chakrabarti (1973), Rao (1979), Cassel/Särndal/Wretman (1977) and many others] we postulate a linear regression model (described below) and consider their biases and mean square errors (in terms of sampling design and assumed model). Yet analytical results, even

under assumption of large (practically infinite) population-size, are found inadequate to settle their relative standings. However, certain numerical investigations as carried out and reported here, point to the superiority of the MSS in most of the cases illustrated.

3.3.2 A MODEL AND THE STRATEGIES

The values of y_i 's and x_i 's fixed for a given finite population are supposed (under a model, say M, as postulated below) to be realizations of random variates Y_i 's and X_i 's ($i = 1, \dots, N$) so jointly distributed that we may write

$$Y_i = \alpha + \beta X_i + U_i, \quad i = 1, \dots, N.$$

Here α, β are unknown real constants, U_i 's are random errors distributed with $E_C(U_i | X_i) = 0$, $E_C(U_i^2 | X_i) = \delta X_i^g$, ($0 < \delta < \infty$, $0 \leq g \leq 2$), $E_C(U_i U_j | X_i, X_j) = 0 \quad \forall i \neq j$, the X_i 's are supposed to be independently, identically gamma-distributed with a common density

$$f(x) = \frac{1}{\Gamma(m)} e^{-x} x^{m-1}, \quad (x > 0, 2 < m < \infty);$$

here E_C denotes conditional (given X_i 's) expectation

(- operator). We will write E_X for expectation(-operator) with respect to the common distribution of X_i 's and $E_M = E_X E_c$ as the over-all model - expectation (-operator). By p we will denote a design and by E_p the design - expectation ($p(s)$ denoting the selection probability of a sample s). We may note that for the SRSWOR we have $E_M E_p = E_p E_M$ but not so necessarily for varying probability sampling. In the later case (i.e. with varying probability sampling schemes) only $E_M E_p$ (and not $E_p E_M$) is relevant because p involves x_i 's (usually).

For MSS, $p(s) = \frac{1}{M_1} \Sigma' x_i / X$ where $X = N \bar{X} = \Sigma x_i$,

$\Sigma' =$ sum over sample and $\Sigma =$ sum over population,

$M_1 = \binom{N-1}{n-1}$; this p will be denoted as p_M and the p corresponding to SRSWOR and independent half-sampling therefrom as p_e and p_h respectively. Let $\bar{y}$, $\bar{x}$, $\bar{r}$ denote the sample means for y_i 's, x_i 's and $r_i = \frac{y_i}{x_i}$'s (and $\bar{y}_1$, $\bar{y}_2$, $\bar{x}_1$, $\bar{x}_2$ the half-sample means).

Let $r^* = 2 \cdot \frac{\bar{y}}{\bar{x}} - \frac{1}{2} \left(\frac{\bar{y}_1}{\bar{x}_1} + \frac{\bar{y}_2}{\bar{x}_2} \right)$ and

$$b = \frac{\Sigma' (y_i - \bar{y})(x_i - \bar{x})}{\Sigma' (x_i - \bar{x})^2}.$$

We will consider the following strategies :

For p_M , the estimator is $t_1 = \bar{X} \cdot \frac{\bar{Y}}{\bar{X}}$,

For p_e , the estimators are $t_2 = \bar{X} \cdot \frac{\bar{Y}}{\bar{X}}$, $t_3 = \bar{y} + b(\bar{X} - \bar{x})$,

For p_h , the estimators are $t_4 = \bar{X} r^*$, $t_5 = \bar{y} + (\bar{X} - \bar{x}) r^*$,

$t_6 = \bar{X} \frac{\bar{y}}{\bar{x}} + \frac{1}{n-1} (\bar{y} - \bar{r} \bar{x})$, $t_7 = \bar{X} r^* + \frac{1}{n-1} (\bar{y} - \bar{r} \bar{x})$,

$t_8 = \bar{X} r^* + \frac{\bar{x}}{n-1} (r^* - \bar{r})$, $t_9 = \bar{X} r^* + \frac{\bar{X}(r^* - \bar{r})}{n-1}$,

t_4 is due to Durbin (1959), t_5 due to Rao(1979), t_6 due to Pascual (1961).

3.3.3 THE BIASES AND MEAN SQUARE ERRORS

Let the biases and mean square errors (MSE) of the estimators be $B_i = E_M E_p (t_i - \bar{Y})$ and $M_i = E_M E_p (t_i - \bar{Y})^2$,
 $i = 1, \dots, 9$.

We find $B_1 = 0$, $B_2 = (1 - \frac{n}{N}) \frac{\alpha}{nm-1}$, $B_3 = 0$,

$B_4 = - \frac{2N + n(nm + 3)}{N(nm-1)(nm-2)} \alpha$, $B_5 = \frac{(1 - n/N)(nm - 3)}{(nm-1)(nm-2)} \alpha$;

$B_6 = - \frac{N(n-1) + n^2(m-1)}{Nn(m-1)(nm-1)} \alpha$,

$$B_7 = - \frac{N(n^2m^2 - nm - 2n + 2) + n^2(m-1)(nm+3)}{Nn(m-1)(nm-1)(nm-2)} \alpha ,$$

$$B_8 = - \frac{N(n^3m^2 - 2n^2m - 2n^2 + 5n - 2) + n^2(n-1)(m-1)(nm+3)}{Nn(n-1)(m-1)(nm-1)(nm-2)} \alpha ,$$

$$B_9 = - \frac{N(n^2m^2 - nm - 2n + 2) + n^2(m-1)(nm-3) - (nm-1)(nm-2)}{N(n-1)(m-1)(nm-1)(nm-2)} \alpha .$$

However, M_i 's are found to be very complicated and hence are not reported here. Instead, we assume N to be very large so that on the strength of the law of large numbers (LLN) we neglect the error in equating $\bar{X}$ to m . So doing, the biases and mean square errors are written as B_i' and M_i' ($i = 1, \dots, 9$).

Ignoring terms in B_i 's involving $1/N$, B_i' 's are immediately available and are as follows:

$$B_1' = 0, \quad B_2' = \frac{\alpha}{nm-1}, \quad B_3' = 0, \quad B_4' = - \frac{2\alpha}{(nm-1)(nm-2)},$$

$$B_5' = \frac{(nm-3)}{(nm-1)(nm-2)} \alpha, \quad B_6' = - \frac{(n-1)\alpha}{n(m-1)(nm-1)},$$

$$B_7' = - \frac{(n^2m^2 - nm - 2n + 2)}{n(m-1)(nm-1)(nm-2)} \alpha ,$$

$$B'_8 = - \frac{(n^3 m^2 - 2n^2 m - 2n^2 + 5n - 2)}{n(n-1)(m-1)(nm-1)(nm-2)} \alpha ,$$

$$B'_9 = - \frac{(n^2 m^2 - nm - 2n + 2)}{(n-1)(m-1)(nm-1)(nm-2)} \alpha .$$

It follows now that

$$B'_9 < B'_8 < B'_7 < B'_6 < B'_4 < B'_5 < B'_2$$

$$\text{and } |B'_9| > |B'_8| > |B'_7| > |B'_2| > |B'_5| > |B'_6| > |B'_4|.$$

(Of course we have $|B'_6| > |B'_4|$ only for $n > 2$).

In deriving the expressions for the various mean square errors (MSE) that follow use has been freely made of the following briefly stated lemmas :

If $X_1, X_2, \dots, X_n$ be independently and identically distributed gamma variates each with a density

$$h(x) = \frac{1}{\Gamma(m)} e^{-x} x^{m-1}, \quad x > 0, \quad m > 0,$$

to be called a density $G(m)$, then it follows that

Lemma 1 For any $f(X_1, \dots, X_n)$ and any q, t, u (with $m > u > 0$),

$$E \left[f(X_1, \dots, X_n) \cdot \frac{x_1^q x_j^t}{x_k^u} \right] = \frac{\Gamma(m+q) \Gamma(m+t) \Gamma(m-u)}{\{\Gamma(m)\}^3} \\ \times E \left[f(X_1, \dots, X_n) | C \right]$$

where $E \left[f(\cdot) | C \right]$ denotes expectation taken with respect to the joint distribution of $X_1, \dots, X_n$, such that the marginal distributions of $X_1, \dots, X_n$ have densities $G(m+q)$ for X_1 , $G(m+t)$ for X_j , $G(m-u)$ for X_k and $G(m)$ for each of X_l ($l = 1, \dots, n$
 $\neq i, j, k$)

Proof. Easy and hence omitted.

Lemma 2 For any non-negative integers a, b and c ,

$$E \left[\frac{x_1^a x_j^b}{\left(\sum_{i=1}^n x_i \right)^c} \right] = \frac{\Gamma(m+a) \Gamma(m+b)}{\{\Gamma(m)\}^2} - \frac{1}{\prod_{i=1}^c (nm+a+b-i)}$$

Proof. Given by Rao/Webster (1966) and hence not reproduced here.

Lemma 3. If X_1 and X_2 are independently gamma distributed with densities $G(a)$ and $G(b)$ respectively, then

$$E \left[\frac{1}{X_1(X_1 + X_2)} \right] = \frac{1}{(a-1)(a+b-2)}$$

Proof. Given by Rao/Webster (1966) and hence not reproduced here.

On calculations we find

$$M_1' = \frac{\alpha^2}{nm-1} + \frac{\delta \Gamma(m+g)}{\Gamma(m)} \cdot \frac{m}{nm+g-1} - \frac{\delta}{N} \cdot \frac{\Gamma(m+g)}{\Gamma(m)}$$

$$M_2' = \frac{\alpha^2 \cdot (nm+2)}{(nm-1)(nm-2)} + \frac{\delta \Gamma(m+g)}{\Gamma(m)} \cdot \frac{nm^2}{(nm+g-1)(nm+g-2)} \\ + \frac{\delta}{N} \cdot \frac{\Gamma(m+g)}{\Gamma(m)} \left[1 - \frac{2mn}{nm+g-1} \right]$$

$$M_4' = \frac{\alpha^2 (n^3 m^3 - 5n^2 m^2 + 12nm + 16)}{(nm-1)(nm-2)^2(nm-4)} \\ + \frac{\delta \Gamma(m+g)}{\Gamma(m)} \cdot \frac{nm^2 (n^2 m^2 + 6nm g - 7nm + 9g^2 - 27g + 18)}{(nm+g-1)(nm+g-2)(nm+2g-2)(nm+2g-4)} \\ - \frac{\delta}{N} \cdot \frac{\Gamma(m+g)}{\Gamma(m)} \left[\frac{nm(nm+3g-3) - 2(g-1)^2}{(nm+g-1)(nm+2g-2)} \right]$$

$$M_5' = \frac{\alpha^2 (n^3 m^3 - 6n^2 m^2 + 3nm + 38)}{(nm-1)(nm-2)^2(nm-4)}$$

$$+ \frac{\delta \Gamma(m+g)}{\Gamma(m)} \cdot \frac{n^4 m^4 + 2(2g-3)n^3 m^3 + (2g^2 - 6g + 5)n^2 m^2 - 2(g-1)(g-2)(2g-5)nm + (g-1)^2(g-2)^2}{n(nm+g-1)(nm+g-2)(nm+2g-2)(nm+2g-4)}$$

$$- \frac{\delta}{N} \cdot \frac{\Gamma(m+g)}{\Gamma(m)} \cdot \frac{n^2 m^2 + (g-1)nm - 4(g-1)^2}{(nm+g-1)(nm+2g-2)}$$

$$M_6' = \alpha^2 \int \frac{n^2 m^2}{(nm-1)(nm-2)} + \frac{(n-2)^2}{(n-1)^2} + \frac{(nm-2)(nm-1)}{n^3(n-1)^2(m-1)(m-2)}$$

$$+ \frac{(nm-2)(nm-1)}{n^3(n-1)(m-1)^2} - \frac{2nm(n-2)}{(n-1)(nm-1)} - \frac{2m}{(n-1)(m-1)}$$

$$+ \frac{2(n-2)(nm-1)}{n(n-1)^2(m-1)} \int$$

$$+ \frac{\delta \Gamma(m+g-2)}{\Gamma(m)} \int \frac{nm^2(m+g-2)(m+g-1)}{(nm+g-1)(nm+g-2)} + \frac{2m(m+g-2)(m+g-1)}{(n-1)(nm+g-1)}$$

$$+ \frac{(m+g-2)(m+g-1) + m^2}{n^3} - \frac{2m(n^2 + n-1)(m+g-2) - m}{n^3(n-1)} \int$$

$$+ \frac{\delta}{N} \cdot \frac{\Gamma(m+g)}{\Gamma(m)} \int 1 - 2\left(\frac{nm}{nm+g-1} + \frac{1}{n-1}\right) + \frac{2(nm+g-1)}{n(n-1)(m+g-1)} \int$$

$$\begin{aligned}
 M_7' = & \alpha^2 \left[\frac{n^2 m^2 (n^2 m^2 - 8nm + 19)}{(nm-1)(nm-2)^2(nm-4)} + \frac{(n-2)^2}{(n-1)^2} + \frac{(nm-1)(nm-2)}{n^3(n-1)^2(m+1)(m-2)} \right. \\
 & + \frac{(nm-1)(nm-2)}{n^3(n-1)(m-1)^2} + \frac{2(n-2)nm(nm-3)}{(n-1)(nm-1)(nm-2)} \\
 & + \frac{2(n-2)}{n(n-1)^2} \cdot \frac{(nm-1)}{(n-1)} - \frac{2m(nm-5)}{(n-1)(m-1)(nm-4)} \Big] \\
 & + \frac{\delta \Gamma(m+g)}{\Gamma(m)} \left[\frac{4nm^2}{(nm+g-1)(nm+g-2)} + \frac{1}{n(n-1)^2} + \frac{4m}{(n-1)(nm+g-1)} \right. \\
 & + \frac{nm^2}{(nm+2g-2)(nm+2g-4)} + \frac{(nm+g-2)(nm+g-1)}{(m+g-1)(m+g-2)n^3(n-1)^2} \\
 & + \frac{m}{n(n-1)} \cdot \frac{1}{(m+g-1)} \left(1 + \frac{nm}{nm+2g-4} \right) \\
 & - \frac{2m}{(nm+2g-2)} \left(\frac{2nm}{nm+g-2} + \frac{1}{n-1} \right) - \frac{2}{(m+g-1)n^2(n-1)} \left(2nm + \frac{nm+g-1}{n-1} \right) \Big] \\
 & + 2 \frac{\delta}{N} \frac{\Gamma(m+g)}{\Gamma(m)} \left[\frac{nm}{(nm+2g-2)} + \frac{nm+g-1}{(m+g-1)n(n-1)} - \left(\frac{2mn}{nm+g-1} + \frac{1}{n-1} \right) \right] \\
 & + \frac{\delta}{N} \cdot \frac{\Gamma(m+g)}{\Gamma(m)}
 \end{aligned}$$

$$\begin{aligned}
 M'_8 = & \alpha^2 \int \frac{(n^2 m^2 - 8nm + 19)(n^4 m^2 - 4n^2 m + nm + 2)}{(n-1)^2 (nm-1)(nm-2)(nm-4)} \\
 & + \frac{(nm-2)(nm-1)}{n^3 (n-1)^2 (m-1)(m-2)} + \frac{(nm-1)(nm-2)}{n^3 (n-1)(m-1)^2} + 1 \\
 & - \frac{2(nm-5)(n^2 m-1)}{n(n-1)^2 (m-1)(nm-4)} + \frac{2(nm-1)}{n(n-1)(m-1)} - \frac{2(nm-3)(n^2 m-1)}{(n-1)(nm-1)(nm-2)} \int \\
 & + \frac{\delta \Gamma(m+g)}{\Gamma(m)} \int \frac{4nm^2}{(nm+g-1)(nm+g-2)} + \frac{4}{n(n-1)^2} + \frac{8m}{(n-1)(nm+g-1)} \\
 & + \frac{nm^2}{(nm+2g-2)(nm+2g-4)} + \frac{1}{4n(n-1)^2} + \frac{m(nm+2)}{4(n-1)^2 (nm+2g-2)(nm+2g-4)} \\
 & + \frac{m}{2(n-1)^2 (nm+2g-2)} + \frac{m}{(n-1)(nm+2g-2)} \\
 & + \frac{nm^2}{(n-1)(nm+2g-2)(nm+2g-4)} - \frac{4nm^2}{(nm+2g-2)(nm+g-2)} \\
 & - \frac{2}{n(n-1)^2} - \frac{2m}{(n-1)^2 (nm+2g-2)} - \frac{8m}{(n-1)(nm+2g-2)} \\
 & + \frac{(nm+g-2)(nm+g-1)}{(m+g-2)(m+g-1)n^3 (n-1)^2} - \frac{2}{n(n-1)} \cdot \frac{1}{(m+g-1)} \left\{ 2m \right.
 \end{aligned}$$

$$+ \frac{2(nm+g-1)}{n(n-1)} - m \left(\frac{1}{2} + \frac{nm}{2(nm+2g-4)} \right)$$

$$- \frac{1}{2(n-1)} \left(\frac{nm+2g-2}{2n} + \frac{m(nm+2)}{2(nm+2g-4)} + m \right) J$$

$$+ \frac{\delta}{N} \cdot \frac{\Gamma(m+g)}{\Gamma(m)} \int_1^2 \left\{ \frac{2nm}{nm+g-1} + \frac{2}{n-1} - \frac{nm}{nm+2g-2} \right.$$

$$\left. - \frac{1}{2(n-1)} \left(1 + \frac{nm}{nm+2g-2} \right) \right\} + \frac{2(nm+g-1)}{(m+g-1)n(n-1)} J$$

$$M'_9 = \alpha^2 \int \frac{n^2}{(n-1)^2} + \frac{n^2 m^2 (n^2 m^2 - 8nm + 19)}{(nm-1)(nm-2)^2(nm-4)}$$

$$+ \frac{m^2(nm-2n+1)}{n(n-1)^2(m-1)^2(m-2)} + 1 - \frac{2n^2}{(n-1)^2} \cdot \frac{m^2(nm-3)}{(m-1)(nm-2)(nm-4)}$$

$$+ \frac{2m}{(n-1)(m-1)} - \frac{2n^2 m}{(n-1)} \cdot \frac{(nm-3)}{(nm-1)(nm-2)} J$$

$$+ \delta \frac{\Gamma(m+g)}{\Gamma(m)} \int \frac{n^2}{(n-1)^2} \cdot nm^2 \cdot \frac{n^2 m^2 + 6nmg - 7nm + 9g^2 - 27g + 18}{(nm+g-1)(nm+g-2)(nm+2g-2)(nm-4)}$$

$$+ \frac{m^2}{(m+g-1)(m+g-2)n(n-1)^2} - \frac{2m^2}{(n-1)^2} \cdot \frac{n}{(m+g-1)} \left(\frac{2}{nm+g-2} - \frac{1}{nm+2g-4} \right) J$$

$$+ \frac{\delta}{N} \cdot \frac{\Gamma(m+g)}{\Gamma(m)} \int_1^2 - \frac{2n^2 m}{n-1} \left(\frac{2}{nm+g-1} - \frac{1}{nm+2g-2} \right) + \frac{2m}{(m+g-1)(n-1)} J$$

The expressions for M_i' 's are complicated and it is difficult to compare the relative performances of the strategies. So, for a few selected parametric combinations namely $\alpha = 0.5$, $\delta = 2.0$, $m = 5.0$, $g = 0.0, 0.5, 1.0, 1.5, 2.0$, and $n = 6, N = 20$ (and also $n = 10, N = 100$ - corresponding values are given in parentheses below those for $n = 6, N = 20$) the values of $|B_i'|$'s, M_i' 's and $B_i'^2$ are presented in Table 3.3.3.1.

Defining relative efficiencies as $E_i' = 100 \frac{M_9'}{M_i'}$ ($i = 1, 2, 4, \dots, 9$) we present their values in Table 3.3.3.2, the values for $n = 10, N = 100$ are given in parentheses below those for $n = 6, N = 20$. From these two tables it appears that MSS fares most promisingly in most of the cases.

Table 3.3.3.1. Giving the biases (B_1^1) and MSE's (M_1^1) of strategies under various parametric combinations. $\alpha = 0.5$, $\delta = 2.0$, $m = 5.0$, $n = 6$, $N = 20$ (values for $n = 10$, $N = 100$ are given in parentheses below those for $n = 6$, $N = 20$).

Estimator \ g	0.0	0.5	1.0	1.5	2.0	$ B_1^1 $	B_1^{12}
t_1	0.2534 (0.1891)	0.5298 (0.4021)	1.1686 (0.9051)	2.7418 (2.1403)	6.6860 (5.2875)	0.0000 (0.0000)	0.000000 (0.000000)
t_2	0.2725 (0.1973)	0.5625 (0.4151)	1.2341 (0.9259)	2.8493 (2.1697)	6.8809 (5.3115)	0.0172 (0.0102)	0.000296 (0.000104)
t_4	0.2695 (0.1945)	0.5739 (0.4169)	1.2955 (0.9471)	3.0313 (2.2597)	7.6063 (5.6281)	0.0012 (0.0004)	0.000001 (0.000000)
t_5	0.2708 (0.1969)	0.5616 (0.3925)	1.2354 (0.9263)	2.8568 (2.1709)	6.8964 (5.3139)	0.0166 (0.0100)	0.000276 (0.000100)
t_6	0.2581 (0.1893)	0.5519 (0.4081)	1.2479 (0.9287)	2.9577 (2.2135)	7.3077 (5.4987)	0.0036 (0.0023)	0.000013 (0.000005)
t_7	0.2638 (0.1787)	0.5780 (0.4145)	1.3348 (0.9597)	3.2378 (2.3211)	8.1604 (5.8483)	0.0221 (0.0129)	0.000488 (0.000166)
t_8	0.2696 (0.1878)	0.5892 (0.4132)	1.3666 (0.9594)	3.3238 (2.3282)	8.4178 (5.8906)	0.0256 (0.0141)	0.000655 (0.000199)
t_9	0.2852 (0.2564)	0.5890 (0.4106)	1.3674 (0.9054)	3.3226 (2.3286)	7.4144 (7.4672)	0.0265 (0.0144)	0.000702 (0.000207)

Table 3.3.3.2. showing efficiencies (E_1') of strategies under various parametric combinations:
 $\alpha = 0.5$, $\delta = 2.0$, $m = 5.0$, $n = 6$, $N = 20$
 (values for $n = 10$, $N = 100$ are given in parentheses below those for $n=6$, $N = 20$)

Estimator \ g	0.0	0.5	1.0	1.5	2.0
t_1	112 (136)	111 (102)	117 (100)	121 (109)	111 (141)
t_2	104 (130)	105 (99)	111 (98)	117 (107)	108 (141)
t_4	106 (132)	103 (98)	106 (96)	110 (103)	97 (133)
t_5	105 (130)	105 (105)	111 (98)	116 (107)	108 (141)
t_6	110 (135)	107 (101)	110 (97)	112 (105)	101 (136)
t_7	108 (143)	102 (99)	102 (94)	103 (100)	91 (128)
t_8	106 (137)	100 (99)	100 (94)	100 (100)	88 (127)
t_9	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)

Remarks. In these derivations law of large numbers has been applied on the X_i 's (on the strength of which we have equated $\bar{X}$ to m) but not on the errors U_i 's because the latter are not assumed independently or identically distributed and about the conditional error variances no assumption was made to ensure validity of law of large numbers (LLN). However in case LLN is also assumed to hold for the errors U_i 's we can equate $\bar{U} = \frac{1}{N} \sum_{i=1}^N U_i$ to 0, $\bar{Y}$ to $\alpha + \beta m = \mu$ and also $S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})^2$ to $E_x(S_x^2) = m$. So we take up a comparative study assuming N to be very large (practically infinite) and neglecting the error in equating $\bar{X}$ to m , $\bar{U}$ to 0, $\bar{Y}$ to $\mu = \alpha + \beta m$ and S_x^2 to m . So doing, the approximate biases and mean square errors will be written as B_i'' and M_i'' ($i = 1, \dots, 9$). Incidentally we note that $B_i' = B_i''$, $i = 1, \dots, 9$. Noting that M_i' can be written in the form $M_i' = A_i(n) \alpha^2 + B_i(n) \delta + C_i(N) \delta$ for $i = 1, 2, 4, \dots, 9$ with $A_i(n)$, $B_i(n)$ as functions of n, m and $C_i(N)$ of N, n, m it may be checked that M_i'' 's are available from M_i' in the

form $M_i'' = A_i(n) \alpha^2 + B_i(n) \delta$, omitting therefrom additional terms of the form $C_i(N) \delta$. The expressions for M_i'' 's are also complicated and it is difficult to assess the relative efficiencies of these strategies. So for a few selected parametric combinations namely $\alpha = 0.5$, $\delta = 2.0$, $m = 5.0$, $g = 0.0, 0.5, 1.0, 1.5, 2.0$, $n = 6$ and $n = 10$ the values of M_i'' 's are presented in Table 3.3.3.3. From this and the Table 3.3.3.4 where relative efficiencies (efficiency of a strategy involving t_i is taken as $E_i'' = 100 \frac{M_9''}{M_i''}$ $i = 1, 2, 4, \dots, 9$) are shown it appears that MSS fares most promisingly in most of the cases.

Remarks. In these derivations law of large numbers has been applied on the X_i 's (on the strength of which we have equated $\bar{X}$ to m) but not on the errors U_i 's because the latter are not assumed independently or identically distributed and about the conditional error variances no assumption was made to ensure validity of law of large numbers (LLN). However in case LLN is also assumed to hold for the errors U_i 's we can equate $\bar{U} = \frac{1}{N} \sum_{i=1}^N U_i$ to 0, $\bar{Y}$ to $\alpha + \beta m = \mu$ and also $S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})^2$ to $E_x(S_x^2) = m$. So we take up a comparative study assuming N to be very large (practically infinite) and neglecting the error in equating $\bar{X}$ to m , $\bar{U}$ to 0, $\bar{Y}$ to $\mu = \alpha + \beta m$ and S_x^2 to m . So doing, the approximate biases and mean square errors will be written as B_i'' and M_i'' ($i = 1, \dots, 9$). Incidentally we note that $B_i' = B_i''$, $i = 1, \dots, 9$. Noting that M_i' can be written in the form $M_i' = A_i(n) \alpha^2 + B_i(n) \delta + C_i(N) \delta$ for $i = 1, 2, 4, \dots, 9$ with $A_i(n)$, $B_i(n)$ as functions of n, m and $C_i(N)$ of N, n, m it may be checked that M_i'' 's are available from M_i' in the

form $M_i'' = A_i(n) \alpha^2 + B_i(n) \delta$, omitting therefrom additional terms of the form $C_i(N) \delta$. The expressions for M_i'' 's are also complicated and it is difficult to assess the relative efficiencies of these strategies. So for a few selected parametric combinations namely $\alpha = 0.5$, $\delta = 2.0$, $m = 5.0$, $g = 0.0, 0.5, 1.0, 1.5, 2.0$, $n = 6$ and $n = 10$ the values of M_i'' 's are presented in Table 3.3.3.3. From this and the Table 3.3.3.4 where relative efficiencies (efficiency of a strategy involving t_i is taken as $E_i'' = 100 \frac{M_9''}{M_i''}$ $i = 1, 2, 4, \dots, 9$) are shown it appears that MSS fares most promisingly in most of the cases.

Table 3.3.3.3. giving MSE's (M_1^n) of strategies under various parametric combinations :

$\alpha = 0.5$, $\delta = 2.0$, $m = 5.0$, $n = 6$, 10

(values in parentheses are for $n = 10$).

Estimator \ δ	0.0	0.5	1.0	1.5	2.0
t_1	0.3534 (0.2091)	0.7478 (0.4457)	1.6686 (1.0051)	3.9414 (2.3803)	9.6860 (5.8875)
t_2	0.3793 (0.2181)	0.7881 (0.4597)	1.7341 (1.0259)	4.0093 (2.4047)	9.6873 (5.8879)
t_4	0.3691 (0.2143)	0.7917 (0.4605)	1.7955 (1.0471)	4.2703 (2.4995)	10.5943 (6.2271)
t_5	0.3773 (0.2177)	0.7872 (0.4369)	1.7354 (1.0263)	4.0156 (2.4061)	9.6968 (5.8899)
t_6	0.3567 (0.2091)	0.7695 (0.4517)	1.7479 (1.0287)	4.1539 (2.4529)	10.2805 (6.0951)
t_7	0.3550 (0.1977)	0.7878 (0.4571)	1.8348 (1.0597)	4.4728 (2.5655)	11.3150 (6.4673)
t_8	0.3594 (0.2068)	0.7974 (0.4558)	1.8666 (1.0594)	4.5660 (2.5726)	11.6096 (6.5120)
t_9	0.3746 (0.2752)	0.7972 (0.4530)	1.8674 (1.0054)	4.5642 (2.5730)	10.6000 (8.0886)

Table 3.3.3.4. showing efficiencies (E_1^*) of strategies under various parametric combinations :
 $\alpha = 0.5$, $\delta = 2.0$, $m = 5.0$, $n = 6, 10$
 (values in parentheses are for $n = 10$).

Estimator \ g	0.0	0.5	1.0	1.5	2.0
t_1	106 (132)	107 (102)	112 (100)	116 (108)	109 (137)
t_2	99 (126)	101 (98)	108 (98)	114 (107)	109 (137)
t_4	101 (128)	101 (98)	104 (96)	107 (103)	100 (130)
t_5	99 (126)	101 (104)	108 (98)	114 (107)	109 (137)
t_6	105 (132)	104 (100)	107 (98)	110 (105)	103 (133)
t_7	106 (139)	101 (99)	102 (95)	102 (100)	94 (125)
t_8	104 (133)	100 (99)	102 (95)	100 (100)	91 (124)
t_9	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)

3.3.4 APPROXIMATE DESIGN - EXPECTATIONS OF SQUARE ERRORS

An approximate formula for $E_{p_e} (t_2 - \bar{Y})^2$ is available in the literature [vide Cochran (1963)] and is given by

$$E_{p_e} (t_2 - \bar{Y})^2 = \frac{N-n}{Nn} (S_y^2 + R^2 S_x^2 - 2R\rho S_x S_y) = A, \text{ say,}$$

$$\text{where } R = \frac{\bar{Y}}{\bar{X}}, \quad S_y^2 = \frac{1}{N-1} \sum (y_1 - \bar{Y})^2, \quad \rho = \frac{\sum (x_1 - \bar{X})(y_1 - \bar{Y})}{S_x S_y}.$$

Writing M_2^{II} for $E_M(A)$, when LLN is applied on the distribution of X_1 's, we have

$$M_2^{\text{II}} = \left(\frac{1}{n} - \frac{1}{N}\right) \left[\frac{N}{N-1} \frac{\alpha^2}{m} + \frac{\delta \Gamma(m+g)}{\Gamma(m)} \left\{ \frac{N}{N-1} + \frac{m+1}{Nm} \right. \right. \\ \left. \left. + \frac{(m+g)}{N(N-1)m} \left(\frac{m+g+1}{m} - 2N \right) \right\} \right].$$

However if LLN is also applied on the errors U_1 's we should equate $\bar{U}$ to 0 (and hence $\bar{Y}$ to $\alpha + \beta m = \mu$) and also S_x^2 to m . So doing, $E_M(A)$ will now be written as $M_2^{(iv)}$ and given by (assuming N very large, practically infinite)

$$M_2^{(iv)} = \frac{\alpha^2}{nm} + \frac{\delta}{n} \cdot \frac{\Gamma(m+g)}{\Gamma(m)}.$$

We note importantly that since for SRSWOR, $E_M E_p$
 $\equiv E_p E_M$, $E_M E_{p_c} (t_2 - \bar{Y})^2$ agrees with $E_{p_e} E_M (t_2 - \bar{Y})^2$.
 Simple formulae for the latter under various models are
 available directly but not for the former. To get the
 former if one uses approximations (as we noted above)
 misleading conclusions may be reached; e.g. using the
 above - noted Cochran's (1963) approximate formula for
 $E_{p_e} (t_2 - \bar{Y})^2$ and further neglecting the error in
 equating $\bar{x}$ to $\bar{X}$ for every sample s with $p(s) > 0$
 Avadhani/Srivastava (1972) observed the strategies
 (p_M, t_1) and (p_e, t_2) as approximately equally efficient
 and also, following the same approach, Arnab (1979)
 observed, under a different model, (p_M, t_1) to fare
 worse than (p_e, t_2) . So, it is suggested that if the
 model M seems appropriate in a situation, it seems
 improper to use approximate formula for $E_{p_e} (t_2 - \bar{Y})^2$
 because in calculating $E_M (t_2 - \bar{Y})^2$ the stringent
 assumption that $\bar{x} \simeq \bar{X}$ for every s with $p(s) > 0$
 may be dispensed with, noting further that it is

constant (for every s) so that $E_{p_e} E_M (t_2 - \bar{Y})^2$ is then just in hand and this should be directly obtained and used.

The situation in respect of the regression estimator is not, however, just as simple. Unfortunately an exact expression in a closed form is not available for $E_M E_{p_e} (t_3 - \bar{Y})^2$. An approximate formula for $E_{p_e} (t_3 - \bar{Y})^2$ is of course available in the literature [vide Cochran (1963)] and is given by

$$E_{p_e} (t_3 - \bar{Y})^2 = \frac{N-n}{Nn} S_y^2 (1 - \rho^2) = B, \text{ say.}$$

Writing M_3^{**} for $E_M(B)$, when LLN is applied on the distribution of X_i 's, we have

$$M_3^{**} = \frac{N-2}{N-1} \cdot \frac{N-n}{Nn} \delta \quad \text{for } g = 0.$$

However if LLN is also assumed to hold on the errors U_i 's, so that assuming N very large (practically infinite) and equating $\bar{U}$ to 0, ($\bar{Y}$ to $\alpha + \rho m = \mu$) and S_x^2 to m we have an approximate expression for $E_M(B)$ (which we will denote by $M_3^{(iv)}$) given by

$$M_3^{(iv)} = \delta/n \quad \text{for } g = 0$$

(general formula for $g \neq 0$ is too complicated for presentation).

However, even when $g = 0$, a simple formula for $E_{p_e} E_M (t_3 - \bar{Y})^2$ is not available under the approximations mentioned in section 3.3.3. (i.e. $\bar{X} \simeq m$ for which $E_{p_e} E_M (t_3 - \bar{Y})^2$ may be denoted by M_3' and $\bar{X} \simeq m$, $\bar{U} \simeq 0$ for which $E_{p_e} E_M (t_3 - \bar{Y})^2$ may be denoted by M_3'').

But if, in addition, n is also supposed large so that $\frac{1}{n-1} \sum (x_i - \bar{x})^2$ is approximated by $\frac{1}{N-1} \sum (x_i - \bar{X})^2$ which is approximated by m , then we have an approximate expression for $E_{p_e} E_M (t_3 - \bar{Y})^2$ which will be denoted by $M_3^{(v)}$ and $M_3^{(vi)}$ in case LLN is applied on the distribution of X_1 's, and X_1 's and U_1 's both respectively.

We have

$$M_3^{(v)} = \delta \left(\frac{1}{n-1} - \frac{1}{N} \right), \quad M_3^{(vi)} = \delta/n-1.$$

3.3.5 A FEW ANALYTICAL BUT APPROXIMATE RESULTS ON MSE'S

Assuming n and m large and neglecting terms $O(\frac{1}{n})$, $O(\frac{1}{m})$ whenever necessary we may note the

following asymptotic results about the MSE's when $g=0$,
1 or 2 :

$$M_1' = M_2' = M_2'' = M_4' = M_5' = \frac{\alpha^2}{nm} + \delta m^g \left(\frac{1}{n} - \frac{1}{N} \right);$$

$$M_6' - \frac{\delta m^{g-1}}{n^4} = M_7' = M_8' = \alpha^2 \left(\frac{1}{n^3} + \frac{1}{n^2} \right) + \delta m^g \left(\frac{1}{n} - \frac{1}{N} \right);$$

$$M_9' = \frac{\alpha^2}{n^2} + \delta m^g \left[\left(\frac{1}{n} - \frac{1}{N} \right) \left(1 - \frac{2}{n} \right) + \frac{1}{n^3} \right].$$

$$\text{For } g=0, \quad M_3'' = \frac{N-2}{N-1} \left(\frac{1}{n} - \frac{1}{N} \right) \delta, \quad M_3^{(v)} = \left(\frac{1}{n-1} - \frac{1}{N} \right) \delta.$$

$$M_1'' = M_2'' = M_2^{(iv)} = M_4'' = M_5'' = \frac{\alpha^2}{nm} + \frac{\delta m^g}{n};$$

$$M_6'' - \frac{\delta m^{g-1}}{n^4} = M_7'' = M_8'' = \alpha^2 \left(\frac{1}{n^3} + \frac{1}{n^2} \right) + \frac{\delta m^g}{n};$$

$$M_9'' = \frac{\alpha^2}{n^2} + \delta m^g \left[\frac{1}{n} \left(1 - \frac{2}{n} \right) + \frac{1}{n^3} \right].$$

$$\text{For } g=0, \quad M_3^{(iv)} = \frac{\delta}{n}, \quad M_3^{(vi)} = \frac{\delta}{n-1}.$$

Thus, in case $g=0$, the regression estimator seems preferable to most of the others; but if, for example $\frac{1}{n^3}$ is negligible and n is large, the strategy 9 may fare even better.

3.4.1 ON THE EFFICIENCIES OF UNBIASED REGRESSION AND REGRESSION - TYPE ESTIMATORS IN FINITE POPULATIONS

Recently Singh/Srivastava (1980) suggested two sampling schemes respectively ensuring design-unbiasedness of a 'regression' and a 'regression-type' estimator for a finite population mean. Our purpose here is to investigate the efficiency of each relative to that of the design-biased 'regression' estimator based on SRSWOR. For this we postulate a usual 'linear regression model' in a super-population set-up and find their relative performances to 'vary with circumstances'.

3.4.2 FORMULATION OF THE PROBLEM AND THE RESULTS

Let y and x be real variates with values Y_i, X_i ($i = 1, \dots, N$) for a finite population of N units with means $\bar{Y}, \bar{X}$. For any sample s (generically, each supposed to be of fixed size n) the regression estimator for $\bar{Y}$ is

$$t = \bar{y} + b (\bar{X} - \bar{x})$$

where $\bar{y}, \bar{x}$ are sample means and $b = s_{xy}/s_x^2$ where $s_x^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$, $s_{xy} = \frac{1}{n-1} \sum (x_i - \bar{x}) y_i$ where $\sum$ is sum for a sample, y_i, x_i 's are sampled values.

Writing $x'_i = x_i - \bar{X}$, $X'_i = X_i - \bar{X}$, $y'_i = y_i - \bar{Y}$,
 $Y'_i = Y_i - \bar{Y}$, $\bar{x}' = \bar{x} - \bar{X}$, $\bar{y}' = \bar{y} - \bar{Y}$ and $S_x^2 = \frac{1}{N-1} \sum X_i'^2$,

where $\sum$ is sum over the population, Singh/Srivastava(1980) gave a scheme for which selection-probability for s is

$$p(s) = \frac{1}{M_0} \cdot \frac{S_x^2}{S_x^2}$$

where $M_r = \binom{N-r}{n-r}$, $r = 0, 1, \dots, n$. For this $E_p(t) = \bar{Y}$ but $E_{p_e}(t) \neq \bar{Y}$, writing E_p, E_{p_e} as expectation-operators with respect to designs p and p_e (i.e. equal probability sampling (SRSWOR)) respectively.

For their another scheme the selection probability is

$$q(s) = \frac{1}{M_1} \sum X_i'^2 / \sum X_i'^2.$$

Their proposed regression-type estimator is

$$t_1 = \frac{n(N-1)}{N(n-1)} \left[\bar{y} - \bar{x}' \sum y_i x_i' / \sum x_i'^2 \right]$$

for which $E_{q_1}(t) = \bar{Y}$ i.e. t_1 is design(q) - unbiased for $\bar{Y}$.

Writing V_p, V_{p_e}, V_q respectively as variance-operators with respect to designs p, SRSWOR and q it may be checked that

$$V_p(t) = \frac{1}{S_x^2} E_{p_e} \left[\bar{y}'^2 s_x^2 - 2 \bar{y}' \bar{x}' s_{xy} + \bar{x}'^2 \frac{s_{xy}^2}{s_x^2} \right]$$

$$V_q(t_1) = \frac{n}{N} \cdot \frac{(N-1)^2}{(n-1)^2} \cdot \frac{1}{\Sigma X_1^2} E_{p_e} \left[\left\{ \bar{y} - \bar{x}' \frac{\Sigma y_1 x_1'}{\Sigma x_1'^2} \right\}^2 \Sigma x_1'^2 \right] - \bar{y}^2$$

Let us, as usual, postulate a model enabling us to write

$$Y_1 = \alpha + \beta X_1 + e_1$$

where α, β are unknown regression-parameters and e_1 is a random variate with conditional (given X_1 values) expectation and variance (with respect to an implied super-population) as 0 and δX_1^g with $\delta > 0$ and $0 \leq g \leq 2$. Writing E_M as the expectation-operator under this model (assuming X_1 's as non-stochastic) and incorporating this model into the system we work out the following:

$$E_M V_p(t) = \frac{\delta}{S_x^2} E_{p_e} \left[\frac{1}{n^2} \Sigma x_1^g - \frac{2}{Nn} \Sigma x_1^g + \frac{1}{N^2} \Sigma X_1^g \right] s_x^2 \\ - 2 \left(\frac{1}{n} - \frac{1}{N} \right) \bar{x}' \cdot \frac{1}{n-1} \Sigma x_1^g (x_1 - \bar{x})$$

$$+ \frac{1}{(n-1)^2} \cdot \frac{\bar{x}'^2}{s_x^2} \sum x_1^g (x_1 - \bar{x})^2 J$$

$$= \frac{n}{n-1} \left(\frac{1}{n} - \frac{1}{N} \right) \delta, \quad (\text{in case } g=0) = I \text{ (say)}$$

$$E_{M^V_q}(t_1) = \frac{n}{N} \cdot \frac{(N-1)^2}{(n-1)^2} \cdot \frac{1}{\sum x_1'^2} E_{P_e} \left[(\alpha + \beta \bar{x})^2 \sum x_1'^2 \right.$$

$$+ \frac{\delta}{n^2} \sum x_1^g \sum x_1'^2 - 2(\alpha + \beta \bar{x}) \bar{x}' \sum x_1' (\alpha + \beta x_1)$$

$$- \frac{2\delta}{n} \bar{x}' \sum x_1' x_1^g + \frac{\bar{x}'^2 \left\{ \sum x_1' (\alpha + \beta x_1) \right\}^2 + \delta \sum x_1'^2 x_1^g}{\sum x_1'^2} J$$

$$- (\alpha + \beta \bar{x})^2 - \frac{\delta}{N^2} \sum x_1^g$$

$$= (\alpha + \beta \bar{x})^2 \left[\frac{n}{n-1} \cdot \frac{N-1}{N(n-1)} \cdot \frac{1}{s_x^2} E_{P_e} \left\{ \frac{(\sum x_1'^2 - n\bar{x}')^2}{\sum x_1'^2} \right\} - 1 \right]$$

$$+ \frac{n}{n-1} \left(\frac{1}{n} - \frac{1}{N} \right) \delta \quad (\text{in case } g=0)$$

$$= (\alpha + \beta \bar{x})^2 (\psi - 1) + \frac{n}{n-1} \left(\frac{1}{n} - \frac{1}{N} \right) \delta, \quad (\text{say})$$

$$= II \text{ (say).}$$

For t based on SRSWOR, an asymptotic expression for $V_{pe}(t)$ is well-known [vide Cochran (1963)] as $\frac{N-n}{Nn} S_y^2 (1-\rho^2)$ where $S_y^2 = \frac{1}{N-1} \sum Y_i'^2$, $\rho = \frac{S_{xy}}{S_x S_y}$ where $S_{xy} = \frac{1}{N-1} \sum Y_i' X_i'$ and $E_M V_{pe}(t) = (\frac{1}{n} - \frac{1}{N}) \delta \left[\frac{1}{N} \sum X_i^g - \frac{1}{S_x^2} \cdot \frac{1}{(N-1)^2} \sum X_i^g X_i'^2 \right]$.

In case $g=0$, we have $E_M V_{pe}(t) = \frac{N-2}{N-1} (\frac{1}{n} - \frac{1}{N}) \delta = III$ (say).

For a comparison of strategies we will consider throughout the case $g=0$ (the homoscedastic case). Singh/Srivastava (1980) also illustrated a special (normal) case of this.

By Cauchy-Schwarz's inequality it follows that

$$\Psi \geq 1$$

and hence we have

Result 1 $II > I$.

For t based on SRSWOR, we also work out an exact expression for $E_M V_{pe}(t)$ as follows:

$$E_M V_{pe}(t) = \delta E_{pe} \left[\frac{1}{n^2} \sum x_1^g - \frac{2}{Nn} \sum x_1^g + \frac{1}{N^2} \sum x_1^g \right. \\ \left. - 2\bar{x}' \left(\frac{1}{n} - \frac{1}{N} \right) \frac{\sum x_1^g (x_1 - \bar{x})}{\sum (x_1 - \bar{x})^2} + \frac{\bar{x}'^2 \sum x_1^g (x_1 - \bar{x})^2}{\left\{ \sum (x_1 - \bar{x})^2 \right\}^2} \right]$$

For t based on SRSWOR, an asymptotic expression for $\dot{V}_{p_e}(t)$ is well-known [vide Cochran (1963)] as $\frac{N-n}{Nn} S_y^2 (1-\rho^2)$ where $S_y^2 = \frac{1}{N-1} \sum Y_i'^2$, $\rho = \frac{S_{xy}}{S_x S_y}$ where $S_{xy} = \frac{1}{N-1} \sum Y_i' X_i'$ and $E_M V_{p_e}(t) = (\frac{1}{n} - \frac{1}{N}) \delta \left[\frac{1}{N} \sum X_i^g - \frac{1}{S_x^2} \cdot \frac{1}{(N-1)^2} \sum X_i^g X_i'^2 \right]$.

In case $g=0$, we have $E_M V_{p_e}(t) = \frac{N-2}{N-1} (\frac{1}{n} - \frac{1}{N}) \delta = III$ (say).

For a comparison of strategies we will consider throughout the case $g=0$ (the homoscedastic case). Singh/Srivastava (1980) also illustrated a special (normal) case of this.

By Cauchy-Schwarz's inequality it follows that

$$\Psi \geq 1$$

and hence we have

Result 1 $II > I$.

For t based on SRSWOR, we also work out an exact expression for $E_M V_{p_e}(t)$ as follows:

$$\begin{aligned} E_M V_{p_e}(t) = & \delta E_{p_e} \left[\frac{1}{n^2} \sum x_1^g - \frac{2}{Nn} \sum x_1^g + \frac{1}{N^2} \sum x_1^g \right. \\ & \left. - 2\bar{x}' \left(\frac{1}{n} - \frac{1}{N} \right) \frac{\sum x_1^g (x_1 - \bar{x})}{\sum (x_1 - \bar{x})^2} + \frac{\bar{x}'^2 \sum x_1^g (x_1 - \bar{x})^2}{\left\{ \sum (x_1 - \bar{x})^2 \right\}^2} \right] \end{aligned}$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right)\delta + \delta E_{p_e} \left[\frac{\bar{x}'^2}{\sum (x_i - \bar{x})^2} \right] \text{ (in case } g=0\text{)}.$$

If following Avadhani and Srivastava (1972) we assume n and N so large that for each s we may ignore the error in neglecting $\bar{x}'$, then we have approximately

$$E_M V_q(t_1) \simeq \frac{N-n}{N(n-1)} (\alpha + \beta \bar{X})^2 + \frac{n}{n-1} \left(\frac{1}{n} - \frac{1}{N}\right)\delta = IV \text{ (say).}$$

With this assumption

$$E_M V_{p_e}(t) \simeq \left(\frac{1}{n} - \frac{1}{N}\right)\delta = V \text{ (say)}$$

and hence we have

Result 2. $IV > I > V$.

Thus design-unbiasedness may lead to higher average mean square error and there should be little surprise for the price one has to pay.

If we further assume that X_i 's are random variates distributed independently identically normally with a mean m and variance σ^2 (say), then denoting the model-expectation (for variation of e_1 given X_1 and over X_1) by E_M' and

assuming (on the strength of the law of large numbers) $\bar{X}$ (for the given population with N large) to equal m , we have

$$E'_M V_{p_e}(t) \simeq \delta \left[\left(\frac{1}{n} - \frac{1}{N} \right) + \frac{1}{n(n-3)} \right] = VI \quad (\text{say}),$$

$$E'_M V_q(t_1) \simeq \frac{2(n+1)}{(n-1)^2(n-2)} (\alpha + \beta m)^2 + \frac{n}{n-1} \left(\frac{1}{n} - \frac{1}{N} \right) \delta = VII(\text{say})$$

on neglecting the error in writing 1 for $\frac{N-1}{N}$ and $\Sigma X_1'^2 \simeq N\sigma^2$ (using the fact that $\frac{\Sigma X_1'^2}{\sigma^2} = \frac{\Sigma (X_1 - m)^2}{\sigma^2}$ is a χ_N^2 - variate i.e. a chi-square variate with N degrees of freedom).

So we have

Result 3. $VI > I$;

$$\text{but } VII - VI = \frac{2(n+1)}{(n-1)^2(n-2)} (\alpha + \beta m)^2 - \frac{\delta}{n-1} \left[\frac{2}{n(n-3)} + \frac{1}{N} \right]$$

which may be either positive or negative.

Incidentally we note that $E_M E_{p_e}(t - \bar{Y}) = 0$ although $E_{p_e}(t - \bar{Y}) \neq 0$ i.e. t is 'model-design-unbiased' though it is 'design-biased'.

So we conclude that the two strategies due to Singh/Srivastava (1980) may often fare worse than the classical strategy of regression estimator with SRSWOR scheme; however the latter also does not always beat the former two.

We may recall that a similar situation prevails in respect of ratio estimators based on SRSWOR and Lahiri/Midzuno/Sen (LMS, say, in brief, 1951, 1952, 1953) scheme - the latter is design-unbiased, the former is design-biased but in terms of efficiency neither is uniformly better than the other [as may be checked from Arnab (1979), Avadhani and Srivastava (1972) and our results in section 3.3.3.].

If we assume $\alpha=0$, $g=0$ in the above model and in addition assume X_1 's to be identically, independently distributed with a common density

$$f(x) = \frac{1}{\Gamma(m)} e^{-x} x^{m-1}$$

with $x > 0$ and $m (> 2)$ taken to equal $\bar{X}$ (using law of large numbers, assuming N large) we have (writing E_M^* for the corresponding model-expectation-operator - over distribution of e_1 given X_1 and over X_1)

$$E_M^* V_r(t_2) = \left(\frac{m}{nm-1} - \frac{1}{N} \right) \delta = \text{VIII (say)}$$

(writing V_r as operator for variance with respect to LMS design and t_2 the ratio estimator viz. $t_2 = \frac{\bar{Y}}{\bar{X}} \cdot \bar{X}$).

Then we have the

Result 4. $\text{II} > \text{I} > \text{VIII}$: ..., (assuming $N > 2n-1$).

In case $\alpha \neq 0$, $g \neq 0$ we have using the notation E_M^* for model - expectation - operator in the above [vide Chaudhuri/Adhikary (1982)]

$$E_M^* V_r(t_2) = \frac{\alpha^2}{nm-1} + \frac{\delta \Gamma(m+g)}{\Gamma(m)} \cdot \frac{m}{nm+g-1} - \frac{\delta}{N} \cdot \frac{\Gamma(m+g)}{\Gamma(m)}$$

(which is β -free), which for $g=0$ reads as

$$E_M^* V_r(t_2) = \frac{\alpha^2}{nm-1} + \delta \left(\frac{m}{nm-1} - \frac{1}{N} \right) = \text{IX (say)}$$

which is to be compared with I, II and III and at least we have

Result 5. $\text{IX} > \text{III}$. (Dropped).

If, in addition, n is also supposed large so that $\frac{1}{n-1} \sum (x_i - \bar{x})^2$ is approximated by $\frac{1}{N-1} \sum (X_i - \bar{X})^2$ which is approximated by m (under assumption of gamma density and by law of large numbers), then we have approximately

$$E_M^* V_{p_e}(t) \simeq \frac{6}{n-1} = X \text{ (say), when } g=0.$$

Then we have

Result 6. $X \simeq I < II$ and if we neglect $\frac{1}{m}$, then

$$X \simeq I \simeq VIII < IX.$$

3.4.3 VARIANCE ESTIMATION

Incidentally it may be noted that the unbiased estimators for $V_p(t)$ and $V_q(t_1)$ proposed by Singh/Srivastava (1980) tacitly following Raj (1954) should often turn out negative because it is known that Raj's variance-estimator (for ratio estimator based on LMS - scheme) is negative for samples with high selection - probabilities. Following Rao(1979) and noting that $V_p(t)$ and $V_q(t_1)$ equal 0, when $Y_i = CX_i$ and $Y_i = CX_i'$, $i = 1, \dots, N$ respectively (for some $C \neq 0$), they can be written as

$$V_p(t) = -\frac{1}{N^2} \sum_{i < j=1}^N \sum d_{ij} x_i x_j \left(\frac{y_i}{x_i} - \frac{y_j}{x_j} \right)^2$$

$$\text{and } V_q(t_1) = -\frac{1}{N^2} \sum_{i < j=1}^N \sum d'_{ij} x'_i x'_j \left(\frac{y_i}{x'_i} - \frac{y_j}{x'_j} \right)^2$$

$$\text{where } d_{ij} = \sum_{s \rightarrow i, j} d_{is} d_{js} p(s) - 1 = e_{ij} - 1, \text{ say,}$$

$$d'_{ij} = \sum_{s \rightarrow i, j} d'_{is} d'_{js} q(s) - 1 = e'_{ij} - 1, \text{ say,}$$

$$\text{where } d_{is} = 0 \text{ if } s \not\rightarrow i$$

$$= \frac{N}{n} \left[1 - n\bar{x}' \cdot \frac{(x_i - \bar{x})}{\sum' (x_i - \bar{x})^2} \right] \text{ if } s \rightarrow i \quad \left. \vphantom{\frac{N}{n}} \right\}$$

$$\text{and } d'_{is} = 0 \text{ if } s \not\rightarrow i$$

$$= \frac{N-1}{n-1} \left[1 - n\bar{x}' \cdot \frac{x'_i}{\sum' x'^2_i} \right] \text{ if } s \rightarrow i \quad \left. \vphantom{\frac{N-1}{n-1}} \right\}$$

Then, non-negative quadratic unbiased estimators of $V_p(t)$ and $V_q(t_1)$ are necessarily of the form, respectively,

$$V_p(t) = -\frac{1}{N^2} \sum_{i < j \in S} \sum d_{ij}(s) x_i x_j \left(\frac{y_i}{x_i} - \frac{y_j}{x_j} \right)^2$$

$$V_q(t_1) = -\frac{1}{N^2} \sum_{i < j \in S} \sum d'_{ij}(s) x'_i x'_j \left(\frac{y_i}{x'_i} - \frac{y_j}{x'_j} \right)^2$$

where $d_{ij}(s) = 0$ if $s \neq i, j$, $\sum_s d_{ij}(s)p(s) = d_{ij}$, $i < j$
and $d'_{ij}(s) = 0$ if $s \neq i, j$, $\sum_s d'_{ij}(s)p(s) = d'_{ij}$, $i < j$.

We may consider three classes of estimators for each of $V_p(t)$ and $V_q(t_1)$ of the above-noted form as follows [c.f. Chaudhuri (1981)] :

$$C_1 : d_{ij}(s) = e_{ij} \left[\frac{\alpha}{\pi_{ij}} + \frac{1-\alpha}{M_{2p}(s)} \right] - \left[\frac{\beta}{\pi_{ij}} + \frac{1-\beta}{M_{2p}(s)} \right]$$

$$C_2 : d_{ij}(s) = d_{ij} \left[\frac{\alpha}{\pi_{ij}} + \frac{1-\alpha}{M_{2p}(s)} \right]$$

$$C_3 : d_{ij}(s) = d_{is} d_{js} - \left[\frac{\lambda}{\pi_{ij}} + \frac{1-\lambda}{M_{2p}(s)} \right]$$

where $0 \leq \alpha, \beta, \lambda \leq 1$ and $\pi_{ij} = \sum_{s \neq i, j} p(s)$ - for $V_p(t)$

$$C'_1 : d'_{ij}(s) = e'_{ij} \left[\frac{\alpha'}{\pi'_{ij}} + \frac{1-\alpha'}{M_{2q}(s)} \right] - \left[\frac{\beta'}{\pi'_{ij}} + \frac{1-\beta'}{M_{2q}(s)} \right]$$

$$C'_2 : d'_{ij}(s) = d'_{ij} \left[\frac{\alpha'}{\pi'_{ij}} + \frac{1-\alpha'}{M_{2q}(s)} \right]$$

$$C'_3 : d'_{ij}(s) = d'_{is} d'_{js} - \left[\frac{\lambda'}{\pi'_{ij}} + \frac{1-\lambda'}{M_{2q}(s)} \right]$$

where $0 \leq \alpha', \beta', \lambda' \leq 1$ and $\pi'_{ij} = \sum_{s \neq i, j} q(s)$ - for $V_q(t_1)$.

the inclusion - probabilities of first two orders for the sampling schemes $p(s)$ and $q(s)$ are respectively obtained as

$$\pi_i = \sum_{s \ni i} p(s) = \frac{Nn-N-n}{N(N-2)} + \frac{N-n}{N-2} \cdot \frac{X_i'^2}{\sum X_i'^2}$$

$$\begin{aligned} \pi_{ij} = \sum_{s \ni i, j} p(s) &= \frac{(n-2)(Nn-N-2n)}{N(N-2)(N-3)} + \frac{(X_j - X_i)^2}{N \sum X_i'^2} \cdot \frac{(N-n)(N-n-1)}{(N-2)(N-3)} \\ &+ \frac{X_i'^2 + X_j'^2}{\sum X_i'^2} \cdot \frac{(n-2)(N-n)}{(N-2)(N-3)} \end{aligned}$$

$$\pi'_i = \sum_{s \ni i} q(s) = \frac{n-1}{N-1} + \frac{N-n}{N-1} \cdot \frac{X_i'^2}{\sum X_i'^2}$$

and

$$\pi'_{ij} = \sum_{s \ni i, j} q(s) = \frac{n-1}{N-1} \cdot \frac{n-2}{N-2} + \frac{n-1}{N-1} \cdot \frac{N-n}{N-2} \cdot \frac{X_i'^2 + X_j'^2}{\sum X_i'^2}$$

In any particular situation, values of $\alpha, \beta, \lambda; \alpha', \beta', \lambda'$ are to be appropriately hit upon to render the variance-estimators non-negative, other parameters like e_{ij} 's, e'_{ij} 's being fully known. Whether such choices are possible seem difficult to investigate.

3.5.1 A STUDY OF THE EFFECT OF VARIATE-TRANSFORMATIONS ON STRATEGIES OF SAMPLING FINITE POPULATIONS

Following Srivenkataramana (1978) we examine the use of variate transformations in improving on efficiencies of several standard sampling strategies involving selections with varying probabilities. Regression models with and without specified forms of distributions of auxiliary variates are considered and exact as well as asymptotic results are derived. Among the strategies studied the one due to Midzuno/Sen (1952,1953) with transformed variate-values appears to be the most promising one under various situations.

In a recent article Srivenkataramana (1978) following Mohanty/Das (1971) suggested a procedure of utilizing the available knowledge about the variate - value for a particular population unit in improving upon usual estimators for a finite population mean $\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$ of a variate y assumed to have a linear regression (in a 'finite population' sense) on an auxiliary variate x whose values x_i 's for all the units of the population $U = (1, \dots, i, \dots, N)$ of size N are positive and known. They, however, consider only SRSWOR sampling designs

and assume the variate - values y_i 's to be non-negative. Here we examine the use of their technique in improving upon other well-known sampling strategies involving selection with varying probabilities. In particular we consider methods of improving upon the following strategies viz. (i) Midzuno/Sen (1952, 1953) strategy; (ii) Horvitz/Thompson (1952) estimators (HTE) based on any IPPS sampling scheme; (iii) Hansen/Hurwitz (1943) strategy; (iv) Rao/Hartley/Cochran (1962) (RHC), strategy and (v,vi) the ratio estimator and the regression estimator both based on SRSWOR sampling scheme. Here the sample-size (number of draws in (iii)) is throughout n ; the common normed size-measures are $p_i = x_i/X$'s and in (iv) the groups are each of size $\frac{N}{n} = k$ assumed to be an integer. By shifting the y -origin to a suitable point we modify the estimators involved in these strategies and find (i) that under certain conditions the modified strategies have smaller expected (under suitable models) variances than the corresponding original ones and (ii) compare the relative efficiencies in terms of expected variances and mean square errors of the different modified estimators including Srivenkataramana's (1978) "modified ratio"

estimator. The models as detailed below are the well-known ones usual in survey-sampling literature [vide, e.g. Cochran (1963), Durbin (1959), Hanurav (1967), Rao (1966), Rao/Webster (1966), Rao/Ramachandran (1974), Rao (1968), Rao/Rao (1971), Rao (1967), among many others].

In the sequel we find that the modification on the Midzuno/Sen (1952, 1953) strategy leads to best results in most situations.

3.5.2 MODIFIED STRATEGIES AND CERTAIN MODELS

Following usual practices we postulate the linear regression set-up with the following model, viz. M (say), for which

$$y_i = \alpha + \beta x_i + u_i, \quad i = 1, \dots, N$$

where $\alpha > 0$, $\beta \geq 0$, (both unknown otherwise),
 $\epsilon(u_i | x_i) = 0 \quad \forall i$, ϵ denoting the operator for conditional expectation given x_i 's in respect of a super-population from which the given finite population is supposed to be a random sample. Further, x_i 's are taken as realized known and positive values on random

variables (also denoted as x_i 's) with incompletely specified common distributions with a mean $m (> 2)$ taken as before through law of large numbers, as equal to the known magnitude of $\bar{X}$. The expectation operators as used earlier will be

$$\bar{\epsilon} = \epsilon_x \epsilon \text{ and } e = \bar{\epsilon}E = \epsilon_x \epsilon E.$$

For M_1 , we will consider two specific cases:

- (i) M_1 , for which x_i 's will be taken to be identically and independently gamma distributed with a common density $f(v) = \frac{1}{\Gamma(m)} e^{-v} v^{m-1}$, $v > 0$, $m > 2$; and
- (ii) M_2 for which x_i 's are independently and identically distributed for which the raw-moments $\mu_r' = \epsilon_x (x^r)$ exist for orders upto to g such that for either model we have $\epsilon(u_i^2 | x_i) = \delta x_i^g \forall i$, $\epsilon(u_i u_j | x_i, x_j) = 0 \forall i \neq j$ with $0 < \delta < \infty$ and $0 \leq g \leq 2$.

Our results will differ from those due to Srivenkataramana (1978) and Mohanty/Das (1971) because they did not adequately characterize and utilize the role of the error term (u_i) in their models and they considered a finite population model as opposed to our super-population model.

By t_i ($i = 1, \dots, 5$) we shall denote the standard estimators (for the first five situations mentioned above) based on the respective sampling schemes mentioned above (the respective strategies being denoted as D_i) and their modifications through variate-transformations (in fact, translations) by t'_i and D'_i respectively. Thus,

$$t_1 = t_1(\underline{y}) = \frac{\bar{y}}{\bar{x}} \cdot \bar{X}, \quad t'_1 = t_1(\underline{z}) + \theta$$

$$t_2 = t_2(\underline{y}) = \frac{1}{N} \sum_{i \in S} \frac{y_i}{\pi_i}, \quad t'_2 = t_2(\underline{z}) + \theta$$

$$t_3 = t_3(\underline{y}) = \frac{1}{N} \cdot \frac{1}{n} \sum_{r=1}^n \frac{y_r}{p_r}, \quad t'_3 = t_3(\underline{z}) + \theta$$

$$t_4 = t_4(\underline{y}) = \frac{1}{N} \sum_{i \in S} \frac{y_i}{p_i} \cdot P_i, \quad t'_4 = t_4(\underline{z}) + \theta$$

$$t_5 = t_5(\underline{y}) = \frac{\bar{y}}{\bar{x}} \cdot \bar{X}, \quad t'_5 = t_5(\underline{z}) + \theta$$

In the above $\bar{y}$, $\bar{x}$, $\bar{z}$ are the sample means,

$\underline{r} = (r_1, \dots, r_i, \dots, r_N)$ for $r = x, y, z$, $z_i = y_i - \theta$,

y'_r is the value of y_i for the unit chosen on the r th

draw and p'_r the corresponding value of p_i ; P_i is the

sum of the p_j 's over the units falling in the i th ($i = 1, \dots, n$) group formed in applying RHC scheme. Finally, θ is supposed to be a quantity at the sampler's choice. The regression estimator based on SRSWOR is

$$t_R = \bar{y} + b(m - \bar{x}), \quad \text{where } b = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

For varying probability sampling schemes, generalised regression estimators are also available in the literature [vide Cassel/Särndal/Wretman (1976)] but we will not treat them here to avoid complicated formulae.

3.5.3 COMPARISON OF MODIFIED STRATEGIES WITH THE CORRESPONDING ORIGINAL ONES

For any homogeneous linear estimator $t = t(\underline{y}) = \sum_{i \in S} a_{is} y_i$ satisfying the condition $t(\underline{x}) = \bar{X}$, it may be checked that for the above super-population model (in fact for a more general one with uncorrelated errors and with common distribution of x_i 's being of any arbitrary form) if we consider the competing estimator $t' = t'(\underline{z}) = t(\underline{z}) + \theta$, then follows (as shown below)

Theorem 3.5.3.

$$\epsilon E (t' - \bar{Y})^2 < \epsilon E (t - \bar{Y})^2$$

if $0 < \theta < 2\alpha$ (3.5.3.1)

Proof.

$$\begin{aligned} & \epsilon E (t' - \bar{Y})^2 - \epsilon E (t - \bar{Y})^2 \\ &= \epsilon E \left[t(y) + \theta \left(1 - \sum_{i \in S} a_{is} \right) - \bar{Y} \right]^2 - \epsilon E \left[t(y) - \bar{Y} \right]^2 \\ &= \epsilon E \left[\theta^2 \left(1 - \sum_{i \in S} a_{is} \right)^2 + 2\theta \left(1 - \sum_{i \in S} a_{is} \right) \{ t(y) - \bar{Y} \} \right] \\ &= \epsilon E \left[\theta^2 \left(1 - \sum_{i \in S} a_{is} \right)^2 \right] + 2\theta \epsilon E \left[\left(1 - \sum_{i \in S} a_{is} \right) \{ t(y) - \bar{Y} \} \right] \\ &= \epsilon E \left[\theta^2 \left(1 - \sum_{i \in S} a_{is} \right)^2 \right] + 2\theta \epsilon \sum_s \left(1 - \sum_{i \in S} a_{is} \right) \left[\sum_{i \in S} a_{is} y_i - \bar{Y} \right] p(s) \\ &= \epsilon E \left[\theta^2 \left(1 - \sum_{i \in S} a_{is} \right)^2 \right] + 2\alpha \sum_s \left(1 - \sum_{i \in S} a_{is} \right) \left[\sum_{i \in S} a_{is} (\alpha + \beta x_i) \right. \\ & \quad \left. - \alpha - \beta \bar{X} \right] p(s) \\ &= \epsilon E \left[\theta^2 \left(1 - \sum_{i \in S} a_{is} \right)^2 \right] - 2\alpha \sum_s \left(1 - \sum_{i \in S} a_{is} \right)^2 p(s) \\ &= (\theta^2 - 2\alpha) \sum_s \left(1 - \sum_{i \in S} a_{is} \right)^2 p(s) < 0 \text{ if } 0 < \theta < 2\alpha, \end{aligned}$$

implying thereby that the modified estimator $t'(\underline{z})$ is better than $t(\underline{y})$ if $0 < \theta < 2\alpha$.

The estimators t and t' for the above-noted strategies are of this form and hence this result applies to them, giving a rule to choose among t_i and t'_i ($i = 1, \dots, 5$) if we can guess about the condition (3.5.3.1).

3.5.4 COMPARISON OF RELATIVE EFFICIENCIES OF STRATEGIES UNDER MODEL M_1 .

On writing $V_1 = \bar{\epsilon} E (t'_i - \bar{Y})^2$, $i = 1, \dots, 5$, we get on calculations (details omitted to save space),

$$V_1 = \frac{(\alpha - \theta)^2}{(nm-1)} + \delta \left[m \cdot \frac{\Gamma(m+g)}{\Gamma(m)} \cdot \frac{1}{nm+g-1} - \frac{1}{N} \cdot \frac{\Gamma(m+g)}{\Gamma(m)} \right]$$

In calculating V_2 we neglect the term

$$(\alpha - \theta)^2 \sum_{i \neq j} \sum \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) \text{ and thus get a conservative}$$

expression on assuming $\pi_{ij} < \pi_i \pi_j \quad \forall i, j$

(otherwise no useful formula for V_2 is available).

$$V_2 = (\alpha - \theta)^2 \left[\frac{m}{n(m-1)} - \frac{1}{N} \right] + \delta \left[\frac{m}{n} \cdot \frac{\Gamma(m+g-1)}{\Gamma(m)} - \frac{1}{N} \cdot \frac{\Gamma(m+g)}{\Gamma(m)} \right]$$

$$V_3 = \frac{(\alpha - \theta)^2}{n(m-1)} + \delta \left[\frac{m}{n} \cdot \frac{\Gamma(m+g-1)}{\Gamma(m)} - \frac{1}{n} \cdot \frac{1}{N} \cdot \frac{\Gamma(m+g)}{\Gamma(m)} \right]$$

$$V_4 = \frac{N-n}{N-1} + V_3$$

$$V_5 = (\alpha - \theta)^2 \cdot \frac{(nm+2)}{(nm-1)(nm-2)} + \delta \int \frac{m^2 n \cdot \Gamma(m+g)}{\Gamma(m)} \cdot \frac{1}{(nm+g-1)(nm+g-2)}$$

$$+ \frac{1}{N} \cdot \frac{\Gamma(m+g)}{\Gamma(m)} - \frac{2mn}{N} \cdot \frac{\Gamma(m+g)}{\Gamma(m)} \cdot \frac{1}{(nm+g-1)} \int.$$

Unfortunately such an exact expression in a closed form is not available for

$$\bar{e}E (t_R - \bar{Y})^2 = V_R \text{ (say)}$$

In fact, we have

$$V_R = \delta \left(\frac{1}{n} - \frac{1}{N} \right) \frac{\Gamma(m+g)}{\Gamma(m)} + \delta E \epsilon_X \int (m - \bar{x})^2 \cdot \frac{\sum_1^n x_1^g (x_1 - \bar{x})^2}{\left\{ \sum_1^n (x_1 - \bar{x})^2 \right\}^2}$$

$$+ 2 \left(\frac{1}{n} - \frac{1}{N} \right) (m - \bar{x}) \frac{\sum_1^n x_1^g (x_1 - \bar{x})}{\sum_1^n (x_1 - \bar{x})^2} \int \text{ for } g \neq 0$$

$$= \delta E \epsilon_X \int \left(\frac{1}{n} - \frac{1}{N} \right) + \frac{(m - \bar{x})^2}{\sum_1^n (x_1 - \bar{x})^2} \int \text{ for } g = 0$$

Further simplifications are obviously difficult to achieve and hence it is not easy to study the exact efficiency of t_R under the present model. If, however, we use the usual asymptotic formula namely

$$E(t_R - \bar{Y})^2 \simeq \left(\frac{1}{n} - \frac{1}{N}\right) S_y^2 (1 - \rho^2)$$

$$\text{[with } S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{Y})^2, \quad S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})^2,$$

$$S_{xy} = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{Y})(x_i - \bar{X}) \text{ and } \rho = S_{xy} / S_x S_y \text{]}$$

available for n and N large then we get

[on algebraic manipulations, with details omitted]

$$E(t_R - \bar{Y})^2 \simeq \frac{N-2}{N-1} \cdot \frac{N-n}{Nn} \sigma^2 = \bar{V}_R, \text{ (say), if } g = 0$$

[general formula for $g \neq 0$ is too complicated for presentation].

On further calculations we may observe the following results in particular

$$(i) \quad V_5 > V_1 > \bar{V}_R > 0, \text{ if } g = 0,$$

$$(ii) \quad V_2 > V_1, \text{ if } g < 1,$$

- (iii) $V_3 > V_1$, if $g < 1$,
- (iv) $V_3 > V_4 \forall g$,
- (v) $V_5 > V_1$, if $1 < g < 2$,
- (vi) $V_4 > V_1$ if $nm < N(1-g)$ when $0 < g < 1$,
- (vii) $V_2 > V_3$ and $V_2 > V_4$, if terms $O(1/N)$ are neglected,
- (viii) $V_2 > V_5$, if terms $O(1/N)$ are neglected and if $g < 1$,
- (ix) $V_3 > V_5$, if $n \geq 5$ and if $g \leq g_0$ where g_0 is a root in $[0, 2]$ of $g^2 - (n^2m - 2nm + 3)g + (n^2m - 3nm + 2) = 0$.

3.5.5 COMPARISON OF RELATIVE EFFICIENCIES OF STRATEGIES UNDER MODEL M_2

In this subsection we shall assume that the model M_2 holds. Also, we shall suppose the sample and population sizes to be large when considering the strategies (i) and (v) and in respect of them we shall neglect the error in equating $\bar{x}$ with $\bar{X} = m$ for every sample s with $p(s) > 0$. The expected (asymptotic or exact) variances of the estimators t_i' ($i = 1, \dots, 5$) under the above situations

will now be written as W_i ($i = 1, \dots, 5$). While comparing the relative efficiencies in what follows n and N will be assumed to be suitably large.

On calculating with simplifications we get

$$W_1 = \delta \left(\frac{1}{n} - \frac{1}{N} \right) \mu'_g.$$

On neglecting the term $(\alpha - \theta)^2 \sum_{i+j} \sum \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right)$ as in section 3.5.4.

$$W_2 = (\alpha - \theta)^2 \left[\frac{m}{n} \varepsilon_x \left(\frac{1}{x} \right) - \frac{1}{N} \right] + \delta \left[\frac{m}{n} \mu'_{g-1} - \frac{\mu'_g}{N} \right].$$

$$W_3 = (\alpha - \theta)^2 \cdot \frac{1}{n} \left[m \varepsilon_x \left(\frac{1}{x} \right) - 1 \right] + \frac{\delta}{n} \left(m \mu'_{g-1} - \frac{\mu'_g}{N} \right)$$

$$W_4 = \frac{N-n}{N-1} W_3.$$

$$W_5 = \left(\frac{1}{n} - \frac{1}{N} \right) \frac{N}{N-1} \rho^2 \mu_2 + \delta \left(\frac{1}{n} - \frac{1}{N} \right) \mu'_g$$

$$(\text{where } \mu_2 = \mu_2' - \mu_1'^2).$$

On further calculations we observe the following results:

$$(i) \quad W_2 > W_1 \quad \text{if } g < 1$$

- (ii) $W_3 > W_1$ if $g < 1$
- (iii) $W_3 > W_4 \forall g$
- (iv) $W_5 > W_1 \forall g$
- (v) $W_4 > W_1$, if terms $O(1/N)$ are neglected and
if $g > 1$
- (vi) $W_2 > W_3$
- (vii) $W_2 > W_4$, if terms $O(1/N)$ are neglected
- (viii) $W_2 > W_5$, if $g > 1$ and $\beta = 0$
- (ix) $W_3 > W_5$ if $g > 1$ and $\beta = 0$
- (x) $W_4 > W_5$ if $g > 1$ and $\beta = 0$.

3.5.6 NUMERICAL VALUES OF RELATIVE EFFICIENCIES OF STRATEGIES

Defining the efficiencies of the strategies as
 $E_i = 100 \frac{V_2}{V_i}$ for $i = 1, \dots, 6$ and writing $V_6 = \bar{V}_R$
 we present below values of the relative efficiencies of
 these strategies for a few combinations of the parametric
 values under the Model M_1 considered above.

We consider the following cases respectively denoted
 as I-VI in Table 3.5.6.1 below and present the values

of E_1 for $\theta = 0.1, 0.5, 0.8$ in the order from top to bottom in cases I-IV and for $\theta = 1.1, 1.5, 1.8$ for case V and $\theta = 0.5, 1.5$ and 2.0 in case VI.

Case I. $\alpha = 0.5, \delta = 2.0, m = 3, n = 5, N = 20$

Case II. $\alpha = 0.5, \delta = 2.0, m = 5, n = 5, N = 20$

Case III. $\alpha = 0.5, \delta = 2.0, m = 5, n = 6, N = 90$

Case IV. $\alpha = 0.5, \delta = 2.0, m = 8, n = 4, N = 100$

Case V. $\alpha = 1.5, \delta = 2.0, m = 8, n = 4, N = 40$

Case VI. $\alpha = 1.5, \delta = 2.0, m = 8, n = 4, N = 100$

Table 3.5.6.1. giving relative efficiencies(E_i) of strategies.

Parametric
Combinations

Case I.	$\alpha = 0.5$	$g = 0.0$	159	100	91	115	137	190
	$\delta = 2.0$		152	100	86	109	131	176
	$m = 3$		156	100	89	112	134	184
	$n = 5$	$g = 0.5$	126	100	86	109	112	
	$N = 20$		121	100	83	105	108	
			124	100	84	107	110	
		$g = 1.0$	103	100	81	103	94	
			100	100	79	100	91	
			102	100	80	102	93	
		$g = 1.5$	86	100	77	97	81	
			85	100	75	95	79	
			86	100	76	96	80	
		$g = 2.0$	74	100	72	92	70	
			73	100	71	90	69	
			73	100	72	91	70	
Case II.	$\alpha = 0.5$	$g = 0.0$	134	100	88	112	123	152
	$\delta = 2.0$		126	100	83	106	116	141
	$m = 5$		130	100	86	109	120	147
	$n = 5$	$g = 0.5$	115	100	84	106	108	
	$= 20$		112	100	81	103	104	
			114	100	83	105	106	
		$g = 1.0$	102	100	80	102	96	
			100	100	79	100	95	
			101	100	80	101	96	
		$g = 1.5$	91	100	77	98	87	
			90	100	77	97	87	
			91	100	77	98	87	
		$g = 2.0$	82	100	75	95	80	
			82	100	74	94	80	
			82	100	75	95	80	

Table 3.5.6.1 (Contd.)

		E_1	E_2	E_3	E_4	E_5	E_6	
Case III.	$\alpha=0.5$ $\delta=2.0$ $m=5$	$g = 0.0$	130	100	102	108	121	138
			122	100	95	101	114	128
			126	100	99	105	118	134
	$n=6$ $N=90$	$g = 0.5$	114	100	98	104	108	
			110	100	95	101	104	
			112	100	97	102	106	
	$g = 1.0$		102	100	96	102	98	
			100	100	94	100	96	
			101	100	95	101	97	
	$g = 1.5$		93	100	95	100	91	
			92	100	94	99	90	
			92	100	94	100	90	
	$g = 2.0$		85	100	94	99	85	
			85	100	93	99	85	
			85	100	93	99	85	
Case IV.	$\alpha=0.5$ $\delta=2.0$ $m=8$	$g = 0.0$	119	100	104	107	111	125
			111	100	97	100	104	116
			115	100	101	104	108	121
	$n=4$ $N=100$	$g = 0.5$	108	100	100	103	103	
			105	100	97	100	100	
			107	100	99	102	102	
	$g = 1.0$		101	100	98	101	98	
			100	100	97	100	97	
			101	100	98	101	97	
	$g = 1.5$		96	100	97	100	94	
			95	100	97	100	94	
			96	100	97	100	94	
	$g = 2.0$		91	100	97	100	91	
			91	100	97	100	91	
			91	100	97	100	91	

Table 3.5.6.1 (Contd.)

		E_1	E_2	E_3	E_4	E_5	E_6	
Case V.	$\alpha=1.5$	$g = 0.0$	119	100	100	108	112	128
	$\delta=2.0$		112	100	93	101	105	119
	$m=8$		116	100	97	105	109	124
	$n=4$	$g = 0.5$	108	100	95	103	103	
	$N=40$		106	100	93	101	100	
			107	100	94	102	102	
		$g = 1.0$	101	100	93	101	97	
			100	100	92	100	97	
			101	100	93	101	97	
		$g = 1.5$	95	100	92	100	93	
			95	100	92	99	93	
			95	100	92	100	103	
		$g = 2.0$	91	100	91	99	90	
			91	100	91	99	90	
			91	100	91	99	90	
Case VI.	$\alpha=1.5$	$g = 0.0$	156	100	137	142	146	174
	$\delta=2.0$		111	100	97	100	104	116
	$m=8$		123	100	108	111	115	131
	$n=4$	$g = 0.5$	123	100	113	117	117	
	$N=100$		105	100	97	100	100	
			110	100	101	104	104	
		$g = 1.0$	106	100	103	106	103	
			100	100	97	100	97	
			102	100	98	102	98	
		$g = 1.5$	98	100	99	102	96	
			95	100	97	100	94	
			96	100	97	100	94	
		$g = 2.0$	92	100	97	100	92	
			91	100	97	100	91	
			91	100	97	100	91	

Remark I. The numerical values presented in Table 3.5.6.1 conform to the algebraic results derived in section 3.5.4. We find that for $g=0$, the regression estimator, of course, is the most efficient among the competitors even for small n and N , but we considered only an asymptotic variance formula for this estimator but exact ones for the rest - (obviously not a fair approach). For $g \neq 0$, the regression estimator is not considered as its asymptotic or exact variance formula is not available and in this case the strategy D_1' fares best in case $g < 1$ but D_2' fares best in case $g > 1$ and D_1' , D_2' and D_4' are equivalent if $g=1$ and $\theta = \alpha$; otherwise D_4' fares mid-way between D_1' and D_2' [consistently with Chaudhuri/Arnab's (1979) findings in case when no variate-translation is made]. Interestingly, D_5' fares poorer compared to D_1' even if N is very large compared to n .

Remark II. In the derivations of mean square errors law of large numbers (LLN) has been applied on the x_1 's (on the strength of which we have equated $\bar{X}$ to m) but not on the errors u_1 's because the latter are not assumed independently or identically distributed and about the conditional error

variances no assumption was made to ensure validity of law of large numbers. However, in case LLN is also assumed to hold for the errors u_i 's so that we can equate

$$\bar{U} = \frac{1}{N} \sum_{i=1}^N u_i \text{ to } 0 \text{ (and hence } \bar{Y} \text{ to } \alpha + \beta m) \text{ certain}$$

modifications on the mean square errors of the estimators (assuming N very large, practically infinite) are noted (writing V_i^* for $E (t_i' - \bar{Y})^2$ in the present situation) as follows :

$$V_1^* = \frac{(\alpha - \theta)^2}{(nm - 1)} + \delta m \frac{\Gamma(m + g)}{\Gamma(m)} \cdot \frac{1}{(nm + g - 1)} .$$

On neglecting the term $(\alpha - \theta)^2 \sum_{i+j} \sum \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right)$ as earlier,

$$V_2^* = (\alpha - \theta)^2 \cdot \frac{m}{n(m - 1)} + \frac{\delta m}{n} \cdot \frac{\Gamma(m + g - 1)}{\Gamma(m)}$$

$$V_4^* = \frac{(\alpha - \theta)^2}{n(m - 1)} + \frac{\delta m}{n} \cdot \frac{\Gamma(m + g - 1)}{\Gamma(m)}$$

$$V_5^* = (\alpha - \theta)^2 \cdot \frac{(nm+2)}{(nm-1)(nm-2)} + \delta \cdot nm^2 \frac{\Gamma(m+g)}{\Gamma(m)} \cdot \frac{1}{(nm+g-1)(nm+g-2)}$$

$$\bar{V}_R^* = \frac{\delta}{n}.$$

On further calculations we may observe the following results in particular

- (i) $V_5^* > V_1^* > \bar{V}_R^*$, if $g = 0$
- (ii) $V_2^* > V_1^*$, if $g < 1$,
- (iii) $V_4^* > V_1^*$, if $g < 1$,
- (iv) $V_5^* > V_1^* \forall g$
- (v) $V_2^* > V_4^* \forall g$
- (vi) $V_2^* > V_5^*$, if $g < 1$.
- (vii) $V_4^* > V_5^*$ if $n \geq 5$ and if $g \leq g_0$ where g_0 is a root in $[0, 2]$ of $g^2 - (n^2m - 2nm + 3)g + (n^2m - 3nm + 2) = 0$
- (viii) $V_2^* = V_4^* \forall g$ if $\theta = \alpha$
- (ix) $V_1^* = V_2^* = V_4^*$ if $g = 1$ and $\theta = \alpha$

$$(x) \quad V_1^* = V_5^* \quad \text{if } g = 2 \quad \text{and} \quad \theta = \alpha.$$

Using the above formulae for mean square errors and defining the efficiencies of the strategies as $E_i^* = 100 \frac{V_2^*}{V_i^*}$ for $i = 1, 2, 4, 5, 6$ and writing $V_6^* = \bar{V}_R^*$ we present values of the relative efficiencies of these strategies for a few selected combinations of the parametric values in Table 3.5.6.2.

We consider the following cases respectively denoted as I - VI in Table 3.5.6.2 below and present the values of E_i^* for $\theta = 0.1, 0.5, 0.8$ in the order from top to bottom in cases I - IV and for $\theta = 1.1, 1.5, 1.8$ for case V and $\theta = 0.5, 1.5$ and 2.0 in case VI.

Case I.	$\alpha = 0.5,$	$\delta = 2.0,$	$m = 3,$	$n = 5$
Case II.	$\alpha = 0.5,$	$\delta = 2.0,$	$m = 5,$	$n = 5$
Case III.	$\alpha = 0.5,$	$\delta = 2.0,$	$m = 5,$	$n = 6$
Case IV.	$\alpha = 0.5,$	$\delta = 2.0,$	$m = 8,$	$n = 4$
Case V.	$\alpha = 1.5,$	$\delta = 2.5,$	$m = 8,$	$n = 4$
Case VI.	$\alpha = 1.5,$	$\delta = 3.5,$	$m = 8,$	$n = 4.$

Table 3.5.6.2 giving relative efficiencies(E_i^*) of strategies

Parametric
Combinations

			E_1^*	E_2^*	E_4^*	E_5^*	E_6^*
Case I.	$\alpha=0.5$ $\delta=2.0$ $m=3$ $n=5$	$g = 0.0$	147	100	105	127	162
			140	100	100	121	150
			144	100	103	125	157
		$g = 0.5$	121	100	104	108	
			116	100	100	104	
			119	100	102	107	
		$g = 1.0$	103	100	103	96	
			100	100	100	93	
			102	100	101	95	
		$g = 1.5$	90	100	101	70	
			88	100	100	68	
			89	100	101	69	
		$g = 2.0$	81	100	101	81	
			80	100	100	80	
			80	100	100	80	
Case II.	$\alpha=0.5$ $\delta=2.0$ $m=5$ $n=5$	$g = 0.0$	127	100	107	117	135
			120	100	100	110	125
			124	100	104	114	131
		$g = 0.5$	112	100	103	106	
			109	100	100	102	
			111	100	102	104	
		$g = 1.0$	102	100	102	98	
			100	100	100	96	
			101	100	101	97	
		$g = 1.5$	93	100	101	92	
			93	100	100	91	
			93	100	100	91	
		$g = 2.0$	87	100	100	87	
			87	100	100	87	
			87	100	100	87	

Table 3.5.6.2 (contd.)

		E_1^*	E_2^*	E_4^*	E_5^*	E_6^*		
Case III.	$\alpha=0.5$ $\delta=2.0$ $m=5$ $n=6$	$g = 0.0$	128	100	106	120	135	
			121	100	100	113	125	
			125	100	104	117	131	
		$g = 0.5$	113	100	103	107		
			109	100	100	104		
			111	100	102	106		
		$g = 1.0$	102	100	102	98		
			100	100	100	97		
			101	100	101	98		
	$g = 1.5$	93	100	101	92			
		92	100	100	91			
		93	100	100	91			
	$g = 2.0$	86	100	100	86			
		86	100	100	86			
		86	100	100	86			
	Case IV.	$\alpha=0.5$ $\delta=2.0$ $m=8$ $n=4$	$g = 0.0$	118	100	107	111	123
				111	100	100	104	114
				115	100	104	108	119
$g = 0.5$			108	100	103	103		
			105	100	100	100		
			107	100	102	102		
$g = 1.0$			101	100	101	98		
			100	100	100	97		
			100	100	100	97		
$g = 1.5$		96	100	100	94			
		96	100	100	94			
		96	100	100	94			
$g = 2.0$		92	100	100	92			
		92	100	100	92			
		92	100	100	92			

Table 3.5.6.2 (contd.)

			E_1^*	E_2^*	E_4^*	E_5^*	E_6^*
Case V.	$\alpha=1.5$ $\delta=2.5$ $m=8$ $n=4$	$g = 0.0$	117	100	106	109	122
			111	100	100	104	114
			88	100	103	107	118
		$g = 0.5$	107	100	102	102	
			105	100	100	100	
			106	100	101	101	
		$g = 1.0$	101	100	101	98	
			100	100	100	97	
			100	100	100	97	
		$g = 1.5$	96	100	100	94	
			96	100	100	94	
			96	100	100	94	
		$g = 2.0$	92	100	100	92	
			92	100	100	92	
			92	100	100	92	
Case VI.	$\alpha=1.5$ $\delta=3.5$ $m=8$ $n=4$	$g = 0.0$	137	100	124	128	147
			111	100	100	104	114
			118	100	106	110	122
		$g = 0.5$	115	100	109	110	
			105	100	100	100	
			108	100	102	102	
		$g = 1.0$	104	100	104	100	
			100	100	100	97	
			101	100	101	98	
		$g = 1.5$	97	100	101	95	
			96	100	100	94	
			96	100	100	94	
		$g = 2.0$	92	100	100	92	
			92	100	100	92	
			92	100	100	92	

Remarks. From the numerical values presented in Table 3.5.6.2 we again find that for $g=0$, the regression estimator is the most efficient among the competitors. For $g \neq 0$, the regression estimator is not considered and in this case the strategy D_1' fares best in case $g < 1$ and D_1' , D_2' and D_4' are equivalent if $g = 1$ and $\theta = \alpha$ [consistently with Chaudhuri/Arnab's (1979) findings in case when no variate - translation is made]. For $g > 1$, however, the strategy D_4' fares better than the strategy D_2' (except when $\theta = \alpha$ in which case D_2' and D_4' are equivalent $\forall g$) which is again better than the strategy D_1' . Interestingly, D_5' fares poorer compared to D_1' excepting when $g = 2$ and $\theta = \alpha$ in which case they are equivalent.

3.6.1 ON A USE OF VARIATE - TRANSFORMATION IN IMPROVING EFFICIENCY IN TWO SPECIFIC SAMPLING STRATEGIES

Let us consider the strategies involving Horvitz/Thompson (1952) estimator based respectively on the sampling schemes due to (1) Goodman/Kish (1950) and (2) Sampford (1967). For them asymptotic variances of the HTE's due respectively to Hartley/Rao (1962) and Asok/Sukhatme (1976) are available. We intend to modify

these strategies by variate-transformation (i.e. translation) following Srivenkataramana (1978) and examine the effect on their efficiencies. Since we have only asymptotic variances for the original ones we also have to be satisfied with asymptotic variances after modification as well. So, it needs to be seen whether results like them extend to this case also. As earlier we consider super-population models for this investigation. We observe then that the modification on the latter yields better results just as the 'original' latter strategy is known to fare better than the 'original' former one.

Srivenkataramana (1978) demonstrated how on shifting the origin in a variate y the ratio estimator for its finite population (of size N) total $T = \sum_{i=1}^N y_i$ based on an SRSWOR (of size n , say) could be improved under certain general conditions. In what follows we report some consequences in extending the technique to more general sampling strategies as mentioned above. For either sampling scheme, on replacing the HTE viz. $t_1 = \sum_{i \in s} \frac{y_i}{\pi_i}$ (π_i 's are the inclusion - probabilities and s is typically a sample, size of every sample being a fixed integer n), by $t_2 = \sum_s \frac{1_i}{\pi_i} + N\theta$ (where θ is a fixed constant to be

chosen suitably at will and $l_i = y_i - \theta$), we show that t_2 has a smaller expected (with respect to a super-population model M described below) 'asymptotic' variance than t_1 under conditions similar to those given by Srivenkataramana (1978) and also that t_2 for Sampford's (1967) scheme has a smaller 'asymptotic' variance than that of t_2 for Goodman/Kish's (1950) scheme.

3.6.2 A SUPER-POPULATION MODEL AND ASYMPTOTIC EFFICIENCIES OF STRATEGIES

Following Durbin (1959) and many others we postulate the following super-population model M for which we write the variate - values as

$$y_i = \alpha + \beta x_i + u_i, \quad i = 1, \dots, N.$$

Here $\alpha, \beta (> 0)$, are unknown constants, x_i 's are known size-measures (with $\bar{X} = \frac{1}{N} \sum_{i=1}^N x_i$, $X = N\bar{X}$) and u_i 's are supposed to be randomly distributed error terms with conditional (for fixed x_i 's) expectations

$$\begin{aligned} \varepsilon(u_i | x_i) &= 0 \quad \forall i, \quad \varepsilon(u_i u_j | x_i, x_j) = 0 \quad \forall i \neq j, \\ \varepsilon(u_i^2 | x_i) &= \delta x_i^g \quad \forall i, \quad \delta > 0, \quad 0 \leq g \leq 2. \end{aligned}$$

We shall throughout take $\pi_i = np_i = \frac{nx_i}{X} \quad \forall i$ (writing $p_i = \frac{x_i}{X} \equiv$ the normed size-measures) and by π_{ij} 's we shall mean the (second-order) inclusion - probabilities of pairs of units (i,j) .

From Hartley/Rao (1962) and Asok/Sukhatme (1976) and using their notations we have the following asymptotic variance (with respect to the sampling designs) expressions:

I For Goodman/Kish (1950) method,

$$\begin{aligned} V_1 = \text{Var}(t_1) = & \sum \pi_i \left(1 - \frac{n-1}{n} \pi_i\right) \left(\frac{y_i}{\pi_i} - \frac{T}{n}\right)^2 \\ & - \frac{n-1}{n^2} \sum \left(2\pi_i^3 - \frac{\pi_i^2}{n} \sum \pi_j^2\right) \left(\frac{y_i}{\pi_i} - \frac{T}{n}\right)^2 \\ & + \frac{2(n-1)}{n^3} \left(\sum \pi_j y_j - \frac{T}{n} \sum \pi_j^2\right) \end{aligned}$$

[correct to $O(N^0)$, taking π_{ij} 's correct to $O(N^{-1/2})$],
which, on writing $z_i = \frac{y_i}{p_i} - T$, simplifies to

$$\begin{aligned} V_1 = & \frac{1}{n} \sum p_i \{1 - (n-1) p_i\} z_i^2 - \frac{n-1}{n} \sum p_i^3 z_i^2 \\ & - \sum p_i^2 \sum p_i^2 z_i^2 - 2 \left(\sum p_i^2 z_i\right)^2 J ; \end{aligned}$$

II For Sampford's (1967) method ,

$$V_2 = \text{Var}(t_1) = \frac{1}{n} \sum p_i \{ 1 - (n-1) p_i \} z_i^2 \\ - \frac{n-1}{n} \left[2 \sum p_i^3 z_i^2 - \sum p_i^2 \sum p_i^2 z_i^2 \right. \\ \left. + (n-2) \left(\sum p_i^2 z_i \right)^2 \right] J,$$

to the same order of approximations as V_1 ;

[of course $V_1 > V_2$, as is easy to verify and as was noted by Asok/Sukhatme (1976) in passing].

Also, with simple calculations we find the following asymptotic variances for the modified estimators for the two sampling schemes mentioned above:

III For Goodman/Kish's (1950) scheme, using Rao/Hartley's (1962) results we have

$$V_3 = V(t_2) = \frac{1}{n} \sum p_i \{ 1 - (n-1)p_i \} w_i^2 \\ - \frac{n-1}{n} \left[2 \sum p_i^3 w_i^2 - \sum p_i^2 \sum p_i^2 w_i^2 \right. \\ \left. - 2 \left(\sum p_i^2 w_i \right)^2 \right] J,$$

$$\begin{aligned}
 \text{where } W_1 &= \frac{y_1 - \theta}{p_1} - (T - N\theta) \\
 &= z_1 + \theta \left(N - \frac{1}{p_1} \right) \\
 &= z_1 + \theta f_1, \text{ writing } f_1 = N - \frac{1}{p_1}.
 \end{aligned}$$

IV For Sampford's (1967) scheme

$$\begin{aligned}
 V_4 = V(t_2) &= \frac{1}{n} \sum p_i \left\{ 1 - (n-1) p_i \right\} W_i^2 \\
 &- \frac{n-1}{n} \sum p_i^3 W_i^2 - \sum p_i^2 \sum p_i^2 W_i^2 \\
 &+ (n-2) \left(\sum p_i^2 W_i \right)^2 J.
 \end{aligned}$$

On calculations, after simplifying, we observe the following results, suppressing detailed algebraic steps, to save space :

$$\begin{aligned}
 \text{(i)} \quad V_3 - V_4 &= (n-1) \left(\sum p_i^2 W_i \right)^2 > 0, \\
 \text{(ii)} \quad V_3 - V_1 &= \frac{1}{n} \sum p_i \left\{ 1 - (n-1) p_i \right\} (W_i^2 - z_i^2) \\
 &- \frac{n-1}{n} \sum 2 p_i^3 (W_i^2 - z_i^2) - \sum p_i^2 \sum p_i^2 (W_i^2 - z_i^2) \\
 &- 2 \left\{ \left(\sum p_i^2 W_i \right)^2 - \left(\sum p_i^2 z_i \right)^2 \right\} J,
 \end{aligned}$$

$$E \left[(w_1^2 - z_1^2) | x_1 \right] = \theta(\theta - 2\alpha) \left(N - \frac{1}{p_1} \right)^2 ;$$

$$E \left[\{ (\sum p_1^2 w_1)^2 - (\sum p_1^2 z_1)^2 \} | x_1 \right] = \theta(\theta - 2\alpha) (\sum p_1^2 f_1)^2.$$

So,

$$E \left[(V_3 - V_1) | x_1 \right] = \theta (\theta - 2\alpha) V_1' < 0,$$

when $0 < \theta < 2\alpha$, (V_1' is V_1 with z_1 replaced by f_1).

Likewise,

$$E \left[(V_4 - V_2) | x_1 \right] = \theta (\theta - 2\alpha) V_2' < 0,$$

when $0 < \theta < 2\alpha$, (V_2' is V_2 with z_1 replaced by f_1).

Remark. If one can guess roughly about the magnitude of α , the y-intercept on the postulated line of regression of y on x, then choosing θ in an appropriate manner, one may, in practice, expect to gain in efficiency over a standard HT strategy through a modification on it as suggested above.

Sri Venkataramana (1978) also observed his modified (with variate-translations) estimator to have a smaller mean square error than that of the usual ratio estimator, both based on SRSWOR, when $0 < \theta < 2\alpha$. His model, however, is different from ours.

CHAPTER-4

SAMPLING STRATEGIES WITH RANDOMIZED RESPONSE TRIALS

4.1 SUMMARY

We consider strategies of sampling with varying probabilities, employing randomized survey techniques, to estimate finite population totals concerning sensitive and stigmatizing characters. Godambe's (1980) approach forms the basis of our study. But we modify it and use Eriksson's (1973) 'given answer' procedure for 'randomization'. Each respondent is given a device with a request to use it unnoticed to divulge his true value with a given probability and to report, with the complementary probability, a value chosen at random from a 'set of numbers' presented to him. Assuming a flow of responses, for various standard sampling designs we present estimation formulae which are readily usable unlike Godambe's (1980) which involves unknown parameters. Under randomized set-up we note modifications on results for existence of 'best' and 'optimal' strategies and indicate changes in criteria for comparison among a few popular ones.

4.2 INTRODUCTION

A real variate y is defined on a finite population labelled $1, \dots, i, \dots, N$ yielding the vector $\underline{Y} = (y_1, \dots, y_1, \dots, y_N)$ of values. The problem is to estimate $Y = \sum_{i=1}^N y_i$ employing a suitable strategy. But y is supposed to stand for a sensitive item of information. So individuals are mostly reluctant to report their y -values on request. In order to promote the rate of response a randomized device is then needed to ensure privacy and confidentiality. For this, essentially following Eriksson (1973) and Godambe (1980) we 'choose' a number (say M) of values $l_1, \dots, l_j, \dots, l_M$ (with $L = \sum_{j=1}^M l_j = 0$) and conceive hypothetically that each individual independently performs a Bernoullian trial with $c(0 < c \leq 1$ - for a direct survey $c=1$) as the probability of 'success' (and $1-c$ as the probability of 'failure'). Our stipulation then is this that in case of a 'success' the i th individual, if selected in a sample, will report his true value y_i and else will report the value of one of the l_j 's chosen at random. Operationally each sampled person will be offered a randomizing device to implement the above unnoticed and co-operate on request. Thus, a vector $\underline{Z} = (z_1, \dots, z_1, \dots, z_N)$ with $Z = \sum_{i=1}^N z_i$, is supposed to be generated

such that z_i is y_i ($i=1, \dots, N$) with probability c and is l_j with probability $\frac{1}{M}(1-c)$ (for $j=1, \dots, M$). This device cannot be effective in ensuring privacy unless l_j 's are cleverly chosen to cover the entire range of relevant y_i -values in each context. So, M may be required to be large in practice even if N is small. Also, prior stratification may be needed to reduce the range of y_i -values. We assume these are cared for in practice. However admittedly, our technique may not claim universal approval. But we may still expect a higher rate of truthful response compared to a 'direct' survey. The condition $\sum_{j=1}^M l_j = 0$ is just to simplify analysis - this may need origin-shifting for y_i 's. Deviations from Godambe's (1980) method should be noted. But this difference enables us to derive practicable results whereas Godambe (1980) gave an optimal estimator which cannot be used as it involves unknown parameters. Eriksson (1973), on the otherhand, did not investigate optimal strategies with unequal probability sampling with 'randomization'.

4.3 CERTAIN 'NON-EXISTENCE', 'ADMISSIBILITY' AND 'OPTIMALITY' RESULTS

Let p be any arbitrary sampling design with $p(s)$ as the probability of selecting a typical sample s . By E_R ,

$E_p, V_p, C_p, E_R, V_R, C_R$ etc. we will denote expectations, variances and covariances with respect to Bernoullian randomization trial (R), design (p) etc. singly or in succession. Based on $(s, z_1 | i \in s)$ an estimator for Y will be written as $e = e(s, \underline{Z})$, understanding that it depends on $\underline{Z}$ only through z_i 's for $i \in s$.

First we consider the class C_H (homogeneous linear unbiased) of estimators of the form $e = e(s, \underline{Z}) =$

$\sum_{i \in s} b_{si} z_i = e_b$, say, such that it is R_p -unbiased for Y i.e.

$E_{R_p}(e) = E_R E_p(e) = E_p E_R(e)$, noting commutativity in

operators E_R, E_p (equivalently "p - unbiased for Z/c", with c. $\sum_{s \rightarrow i} b_{si} p(s) = 1 \forall i$) and its natural extension

to a wider class C_G of p - unbiased estimators $e = e(s, \underline{Z})$ for Z/c i.e. $E_p(e) = Z/c$ (and hence R_p - unbiased for Y).

We will use e's only in C_G . Godambe (1980) considered a class of non-homogeneous linear unbiased estimators - a class wider than C_H but not necessarily narrower than C_G . Now for every e in C_G we have

$$\begin{aligned} V_{R_p}(e) &= E_R E_p(e - Y)^2 = E_R V_p(e) + V_R E_p(e) \\ &= E_R V_p(e) + V_R (Z/c) \end{aligned} \quad (4.3.1)$$

We have the following parallel to Basu's (1971) result about non-existence of 'best' (i.e. UMV, uniformly minimum variance) estimator in 'direct' surveys for any design (not amounting to 'census') for e 's in C_G . Writing $\underline{A} = (A_1, \dots, A_1, \dots, A_N)$, (with $A = \sum_{i=1}^N A_i$), for $\underline{Z}$ when z_i equals known A_i for $i = 1, \dots, N$, if e be the 'best' in C_G then we may take another, say, $\bar{e}$ in C_G as

$$\bar{e} = \bar{e}(s, \underline{Z}) = e(s, \underline{Z}) - e(s, \underline{A}) + \frac{A}{C}.$$

Since $V_{R_p}(\bar{e}) = 0$ if $\underline{Z} = \underline{A}$, and $\underline{A}$ is arbitrary, the result follows. Keeping e within C_H , from Godambe's (1955) work a corresponding non-existence result also easily follows, provided we employ a 'non-uni-cluster' design, also from Hege/Hanurav (1965, 1966) it may be checked that $e^* = \frac{1}{C} \cdot \sum_{i \in S} \frac{z_i}{\pi_i}$, (with $\pi_i = \sum_{s \ni i} p(s)$, assumed positive, throughout, $\forall i$) is UMV for Y in C_H if this design is 'uni-cluster'.

It is also easy to check that e^* is admissible for Y among all e_b 's in C_H as well as among e 's in C_G - extending the 'direct survey' results due to Godambe (1960) and Godambe/Joshi (1965). Also, that the strategy (e^*, p^*)

(p^* being a fixed expected sample-size design) is uniformly admissible among all strategies (e, p) with e in C_G and p in a class of fixed expected sample - size designs may likewise be checked - extending the result due to Ramakrishnan (1975) in 'direct' survey situation.

Writing $\pi_{ij} = \sum_{s \ni i, j} p(s)$ we may note, (for comparison with the 'direct-survey' situations) that

$$\begin{aligned} \text{Var}(e^*) = v_{Rp}(e^*) = & \sum_{i=1}^N y_i^2 \left(\frac{1}{c\pi_i} - 1 \right) + \sum_{i \neq j=1}^N \sum \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) y_i y_j \\ & + \frac{1-c}{c^2} \cdot \left(\frac{1}{M} \sum_{j=1}^M 1_j^2 \right) \cdot \sum_{i=1}^N \frac{1}{\pi_i} . \end{aligned}$$

If a super-population model, say M_1 , is postulated so that using the model - expectation operator E_M , (and V_M , C_M as model variance, co-variance operators) we have

$$E_M(y_i) = \theta_i, \quad V_M(y_i) = \sigma_i^2, \quad C_M(y_i, y_j) = \rho \sigma_i \sigma_j,$$

where θ_i 's are reals, σ_i^2 's are positive numbers and ρ is in $(0,1)$, these being otherwise unknown. Then, one has, [c.f. Cassel/Särndal/Wretman (1976), in 'direct' survey situations].

Theorem 4.3.1. For any p and e_b in C_H we have

$$\begin{aligned} E_M V_{Rp}(e_b) \geq & (1-\rho) \sum_{i=1}^N \sigma_i^2 \left(\frac{1}{c\pi_i} - 1 \right) + \frac{1-c}{c} \sum_{i=1}^N \left(\frac{\theta_i^2 + \rho \sigma_i^2}{\pi_i} \right) \\ & + \frac{1-c}{c^2} \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \sum_{i=1}^N \frac{1}{\pi_i} \end{aligned} \quad (4.3.2)$$

Proof. Under model M_1 , we have

$$\begin{aligned} E_M V_{Rp}(e_b) &= E_M E_R E_p (e_b - Y)^2 \\ &= \sum_{i=1}^N (\sigma_i^2 + \theta_i^2) \left[c \cdot \sum_{s \rightarrow i} b_{si}^2 p(s) - 1 \right] \\ &\quad + (1-c) \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \sum_{i=1}^N \sum_{s \rightarrow i} b_{si}^2 p(s) \\ &\quad + \sum_{i \neq j=1}^N (\theta_i \theta_j + \rho \sigma_i \sigma_j) \left[c^2 \sum_{s \rightarrow i, j} b_{si} b_{sj} p(s) - 1 \right] \\ &= (1-\rho) \sum_{i=1}^N \sigma_i^2 \left[c \cdot \sum_{s \rightarrow i} b_{si}^2 p(s) - 1 \right] \\ &\quad + c(1-c) \sum_{i=1}^N (\theta_i^2 + \rho \sigma_i^2) \sum_{s \rightarrow i} b_{si}^2 p(s) \\ &\quad + \rho E_p \left[c \cdot \sum_{i \neq s} b_{si} \sigma_i \right]^2 - \rho \left(\sum_{i=1}^N \sigma_i \right)^2 \end{aligned}$$

$$\begin{aligned}
 & + E_p \left[c \cdot \sum_{i \in S} b_{si} \theta_i \right]^2 - \left(\sum_{i=1}^N \theta_i \right)^2 \\
 & + (1-c) \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \sum_{i=1}^N \sum_{s \rightarrow i} b_{si}^2 p(s) \\
 & \geq (1-\rho) \sum_{i=1}^N \sigma_i^2 \left(\frac{1}{c\pi_i} - 1 \right) + \frac{1-c}{c} \sum_{i=1}^N \left(\frac{\theta_i^2 + \rho \sigma_i^2}{\pi_i} \right) \\
 & + \frac{1-c}{c^2} \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \sum_{i=1}^N \frac{1}{\pi_i}
 \end{aligned}$$

using the inequality $\sum_{s \rightarrow i} b_{si}^2 p(s) \geq \frac{1}{c^2 \pi_i} \quad \forall i$

and the fact that $E_p \left[c \cdot \sum_{i \in S} b_{si} \theta_i \right]^2 \geq E_p^2 \left[c \cdot \sum_{i \in S} b_{si} \theta_i \right]$

$$= \left(\sum_{i=1}^N \theta_i \right)^2$$

and $E_p \left[c \cdot \sum_{i \in S} b_{si} \sigma_i \right]^2 \geq E_p^2 \left[c \cdot \sum_{i \in S} b_{si} \sigma_i \right]$

$$= \left(\sum_{i=1}^N \sigma_i \right)^2.$$

The lower bound (4.3.2) is attained if the following hold together, namely,

$$\left. \begin{aligned} \text{c. } \sum_{i \in S} b_{si} \theta_i &= \sum_{i=1}^N \theta_i, \quad \text{c. } \sum_{i \in S} b_{si} \sigma_i = \sum_{i=1}^N \sigma_i \\ b_{si} &= \frac{1}{c\pi_i} \quad \forall i \text{ and } \forall s \text{ with } p(s) > 0. \end{aligned} \right\} \quad (4.3.3)$$

Let positive numbers W_i 's (with $W = \sum_{i=1}^N W_i$, $p_i = \frac{W_i}{W}$)

be available as size-measures which may be used in choosing the designs. From now on we consider mostly designs involving a fixed sample size n , say, to be denoted as $p(n)$. We shall denote every design $p(n)$ as $p_W(n)$ when $\pi_i = np_i \quad \forall i$.

Let, in particular, $\theta_i = \rho W_i$ and $\sigma_i^2 = \sigma^2 W_i^2 \quad \forall i$ (4.3.4)

with ρ and σ^2 as unknown positive constants. Now (4.3.3) will hold if (4.3.4) hold and p is any $p_W(n)$. So assuming (4.3.4) and p as a $p_W(n)$ the lower bound (on the right-hand side) of (4.3.2) equals

$$\begin{aligned} & \frac{1-c}{c} \cdot \frac{\rho^2 W^2}{n} + \frac{\sigma^2 W^2}{nc} - \rho \frac{\sigma^2 W^2}{n} - (1-\rho) \sigma^2 \sum_{i=1}^N W_i^2 \\ & + \frac{1-c}{c^2} \cdot \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \frac{W}{n} \sum_{i=1}^N \frac{1}{W_i} = E_M^{V_{Rp}}(e^*), \end{aligned}$$

when p equals $p_W(n)$, giving the

Theorem 4.3.2. For any p as $p_W(n)$ and e_b in C_H , we have under M_1 with $\theta_i = \rho w_i$ and $\sigma_i^2 = \sigma^2 w_i^2 \forall i$,

$$E_M V_{Rp}(e_b) \geq E_M V_{Rp}(e^*).$$

If M_1 is modified into M_1' (say) such that $\rho = 0$, then it is easy to check that for any p and any e_b in C_H , under M_1'

$$E_M V_{Rp}(e_b) \geq \sum_{i=1}^N \sigma_i^2 \left(\frac{1}{c w_i} - 1 \right) + \frac{1-c}{c} \sum_{i=1}^N \frac{\theta_i^2}{w_i} + \frac{1-c}{c^2} \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \sum_{i=1}^N \frac{1}{w_i}$$

To derive this we cannot directly use (4.3.1) but use the relation

$$V_{Rp}(e) = V_p E_R(e) + E_p V_R(e) \quad (4.3.5)$$

interchanging the orders in E_p, E_R , remembering their commutativity.

Also it easily follows that for any p as $p_W(n)$ and e_b in C_H , we have under M_1' with $\theta_i = \rho w_i \forall i$,

$$\begin{aligned} E_M V_{Rp}(e_b) \geq E_M V_{Rp}(e^*) &= \frac{1-c}{c} \cdot \frac{\rho^2 w^2}{n} + \sum_{i=1}^N \sigma_i^2 \left(\frac{w}{n} \cdot \frac{1}{c w_i} - 1 \right) \\ &+ \frac{1-c}{c^2} \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \frac{w}{n} \sum_{i=1}^N \frac{1}{w_i}. \end{aligned}$$

[c.f. Godambe's(1955) result with 'direct' surveys].

If M_1 is modified into M_2 (say) so that y_1 's are 'independently' distributed, we have, following and extending Godambe/Joshi's (1965) work with 'direct' surveys

Theorem 4.3.3. Under M_2 , for any e in C_G and any p ,

$$E_M V_{Rp}(e) \geq \sum_{i=1}^N \sigma_i^2 \left(\frac{1}{c\pi_i} - 1 \right) + \frac{1-c}{c} \sum_{i=1}^N \frac{\theta_i^2}{\pi_i} \\ + \frac{1-c}{c^2} \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \sum_{i=1}^N \frac{1}{\pi_i} .$$

Proof. From (4.3.1) or (4.3.5) we cannot derive it straight away. But we may follow Godambe/Joshi (1965) to write for e in C_G

$$e = e^* + h, \text{ such that}$$

$$0 = E_p(h) = E_p h(s, \underline{z}) = \sum_s h(s, \underline{z}) p(s).$$

$$\text{This implies } \sum_{s \neq i} h(s, \underline{z}) p(s) = - \sum_{s \neq i} h(s, \underline{z}) p(s) \quad (4.3.6)$$

Denoting by E_{MR} , V_{MR} , C_{MR} the operators for expectation, variance and co-variance with respect to model and Bernoullian trial in succession, we have

$$E_p V_{MR}(e) = E_p V_{MR}(e^*) + E_p V_{MR}(h) + 2 E_p C_{MR}(e^*, h).$$

$$\text{But } E_p C_{MR}(e^*, h) = E_p \int C_M(E_R e^*, E_R h) + E_M C_R(e^*, h) \int$$

$$\text{and } E_p C_M(E_R e^*, E_R h) = -E_p E_M \int \sum_{i \in S} \frac{y_i - \theta_i}{\pi_i} \cdot E_R h \int$$

$$= E_M \int \sum_{i=1}^N \left(\frac{y_i - \theta_i}{\pi_i} \right) \sum_{s \neq i} E_R h p(s) \int$$

$$= E_M E_R \int \sum_{i=1}^N \left(\frac{y_i - \theta_i}{\pi_i} \right) \sum_{s \neq i} h p(s) \int$$

$$= E_R E_M \int - \sum_{i=1}^N \left(\frac{y_i - \theta_i}{\pi_i} \right) \sum_{s \neq i} h p(s) \int$$

$$= - E_R \int \sum_{i=1}^N E_M \left(\frac{y_i - \theta_i}{\pi_i} \right) E_M \sum_{s \neq i} h p(s) \int$$

$$= 0.$$

$$\text{Also, } E_p E_M C_R(e^*, h) = E_p E_M E_R \int \sum_{i \in S} \left(\frac{z_i}{c\pi_i} - \frac{y_i}{\pi_i} \right) h \int$$

$$= E_M E_R \int \sum_{i=1}^N \frac{z_i}{c\pi_i} \sum_{s \neq i} h p(s) - \sum_{i=1}^N \frac{y_i}{\pi_i} \sum_{s \neq i} h p(s) \int$$

$$= E_M \int - \sum_{i=1}^N \frac{E_R(z_i)}{c\pi_i} E_R \left\{ \sum_{s \neq i} h p(s) \right\}$$

$$+ \sum_{i=1}^N \frac{y_i}{\pi_i} E_R \left\{ \sum_{s \neq i} h p(s) \right\} \int$$

$$= 0.$$

$$\text{So, } E_p V_{MR}(e) \geq E_p V_{MR}(e^*) \quad (4.3.7)$$

Following Godambe/Thompson (1977), we write

$$E_M V_{Rp}(e) = E_p V_{MR}(e) + E_p \Delta_{MR}^2(e) - V_M(Y)$$

$$\text{where } \Delta_{MR}(e) = E_M E_R(e) - E_M(Y).$$

So, using (4.3.7), we have

$$E_M V_{Rp}(e) \geq E_p V_{MR}(e^*) - V_M(Y)$$

$$= E_p [V_M E_R(e^*) + E_M V_R(e^*)] - V_M(Y)$$

$$= E_p [V_M (\sum_{i \in S} \frac{y_i}{\pi_i}) + E_M \{ c(1-c) \sum_{i \in S} \frac{y_i^2}{c^2 \pi_i^2}$$

$$+ \frac{1-c}{c^2} (\frac{1}{M} \sum_{j=1}^M l_j^2) \sum_{i \in S} \frac{1}{\pi_i^2} \}] - \sum_{i=1}^N \sigma_i^2$$

$$= \sum_{i=1}^N \sigma_i^2 (\frac{1}{c\pi_i} - 1) + \frac{1-c}{c} \sum_{i=1}^N \frac{\sigma_i^2}{\pi_i} + \frac{1-c}{c^2} (\frac{1}{M} \sum_{j=1}^M l_j^2) \sum_{i=1}^N \frac{1}{\pi_i} \quad \text{Q.E.D.}$$

Theorem 4.3.4. For any design p as $p_W(n)$ and for any estimator e in C_G , we have under model M_2 with $\theta_i = \beta w_i \forall i$.

$$E_M V_{Rp}(e) \geq E_M V_{Rp}(e^*) \quad (4.3.8)$$

Proof. $E_M V_{Rp}(e^*) = E_P V_{MR}(e^*) + E_P \Delta_{MR}^2(e^*) - V_M(Y)$

Hence $E_M V_{Rp}(e) - E_M V_{Rp}(e^*) \geq E_P \Delta_{MR}^2(e),$

(since $E_P \Delta_{MR}(e^*) = 0$)

implying the proof.

Next we consider the following model M_3 , say, [taking the cue from Kempthorne (1969), C.R. Rao (1971) and Godambe/Thompson (1973)]. Under this we choose a vector

$\underline{a} = (a_1, \dots, a_1, \dots, a_N)$ of positive numbers a_i 's with $n = \sum_{i=1}^N a_i$, write $r_i = \frac{z_i}{ca_i}$, and assign a prior distribution

F (say, with E_F, V_F as expectation and variance operators) to $\underline{z}$ which gives an equal probability to each of the $N!$ points $(r_{i_1}, \dots, r_{i_N})$ obtainable from

$\underline{r} = (r_1, \dots, r_N)$ on permuting the co-ordinates of $(1, \dots, N)$, a typical permutation being $(i_1, \dots, i_N)$,

keeping $\underline{a}$ fixed. A design $p(n)$ will be denoted as $p_a(n)$

if $\pi_i = a_i \forall i$ and as $p_{\pi_a}(n)$ if π_i 's are subject only

to $\sum_{i=1}^N \pi_i = \sum_{i=1}^N a_i = n$. Writing $e_a^* = e_a^*(s, \underline{z}) = \sum_{i \in s} \frac{z_i}{ca_i}$

and following Godambe/Thompson's (1973) result, we have,

Theorem 4.3.5. (i) For any p as $p_a(n)$,

$$E_R E_F E_p (e_a^* - Y)^2 \text{ has a common value,}$$

$$\text{say, } E_R E_F E_{p(a)} (e_a^* - Y)^2 ;$$

(ii) For any p as $p_{\pi a}(n)$ and any e in C_G ,

$$E_R E_F E_p (e - Y)^2 \geq E_R E_F E_{p(a)} (e_a^* - Y)^2.$$

Proof. First we assume (i) is true and prove (ii) and then prove (i).

Following Godambe/Thompson's (1973) argument it is easy to check that for any p, p' of the type $p_a(n)$,

$$\sum_{\pi} \sum_s \frac{p'(s)}{N!} \left(\sum_{i \in s} r_{\pi(i)} \right)^2 - \frac{1}{N!} \sum_{\pi} \left(\sum_{i=1}^N a_i r_i \right)^2 \leq E_F E_p \left[e - E_p(e) \right]^2 \quad (4.3.9)$$

where $(\pi(1), \dots, \pi(N)) = \pi(1, \dots, N)$ is a permutation π of $(1, \dots, N)$ and $\sum_{\pi}$ denotes summation over $N!$ permutations of $I = (1, \dots, N)$.

Now,

$$\begin{aligned} E_R E_F E_p (e - Y)^2 &= E_R E_F E_p \left[e - E_p(e) \right]^2 + E_R E_F \left[E_p(e) - Y \right]^2 \\ &= E_R E_F E_p \left[e - E_p(e) \right]^2 + E_R (Z/c)^2 - Y^2. \end{aligned}$$

So, operating on both sides of (4.3.9) by E_R and adding $E_R (Z/c)^2 - Y^2$ on both sides we get

$$E_R \left[\frac{n}{N} \sum_{i=1}^N r_i^2 + \frac{n(n-1)}{N(N-1)} \sum_{i \neq j=1}^N r_i r_j \right] - Y^2 \leq E_R E_F E_p (e-Y)^2 \quad (4.3.10)$$

The left-hand side of (4.3.10) works out as

for any p, p' of the type $p_a(n)$

$$\frac{n}{N} \sum_{i=1}^N \frac{y_i^2}{c a_i^2} + \frac{n(n-1)}{N(N-1)} \sum_{i \neq j=1}^N \sum \frac{y_i y_j}{a_i a_j} + \frac{n}{N} \cdot \frac{1-c}{c^2} \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \sum_{i=1}^N \frac{1}{a_i^2}$$

which we may denote as $E_R E_F E_{p(a)} (e_a^* - Y)^2$.

Proof of (ii) is finished.

For any p of the type $p_a(n)$,

$$E_R E_F E_p (e_a^{*2}) = E_R \left[\sum_{\pi} \sum_s \frac{p(s)}{N!} \left(\sum_{i \in s} r_{\pi(i)} \right)^2 \right]$$

$$= E_R \left[\frac{n}{N} \sum_{i=1}^N r_i^2 + \frac{n(n-1)}{N(N-1)} \sum_{i \neq j=1}^N r_i r_j \right]$$

$$= \frac{n}{N} \sum_{i=1}^N \frac{y_i^2}{c a_i^2} + \frac{n}{N} \cdot \frac{1-c}{c^2} \cdot \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \sum_{i=1}^N \frac{1}{a_i^2} + \frac{n(n-1)}{N(N-1)} \sum_{i \neq j=1}^N \sum \frac{y_i y_j}{a_i a_j}.$$

Hence (i) is proved.

Remark. A comprehensive investigation into the rates (needed 100% for our method) of truthful reporting with 'randomized' and comparable 'direct' surveys and corresponding relative performances of relevant sampling strategies is of course important but that is not attempted here. But granting that our 'randomized' method works (in some situation at least, hopefully) for the sampling strategies proposed, it seems interesting to note above a few features comparable with those available in the literature when 'direct' surveys are feasible. The results look similar in form with modifications along expected directions but derivation seems non-trivial — restriction to the class C_G is a novel necessity. Another point to note is that if y_i 's are origin - shifted to $y_i' = y_i - B$ ($i=1, \dots, N$, say), then our method estimates $E y_i'$ directly and adding NB we get the desired estimate for Y . Consequently optimality may be impaired in the absence of 'origin-invariance' which is rare in survey sampling.

4.4 RELATIVE EFFICIENCIES OF A FEW USUAL SAMPLING STRATEGIES MODIFIED UNDER RANDOMIZED TRIALS

We will consider another super-population model, M_4 , say, so as to assume

$$\left. \begin{aligned} E_M(y_i) &= \beta W_i, \quad C_M(y_i, v_j) = \sigma^2 W_i^g \quad \text{if } i = j \\ &= 0 \quad \text{if } i \neq j \end{aligned} \right\}$$

with β any positive number and g in $[0, 2]$. Several strategies will be compared in this section only under M_4 .

Writing $e^* = e_1$ (for comparison with its competitors to follow), for any p as $p_W(n)$, we have,

$$\begin{aligned} E_1 &= E_M V_{Rp} (e_1) = E_M V_{Rp_W} (e_1) \quad (\text{say}) \\ &= \sigma^2 \left(\frac{W}{n} \sum_{i=1}^N W_i^{g-1} - \sum_{i=1}^N W_i^g \right) + \frac{1-c}{c} \left(\frac{\beta^2 W^2}{n} + \frac{\sigma^2 W}{n} \sum_{i=1}^N W_i^{g-1} \right) \\ &\quad + \frac{1-c}{c^2} \cdot \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \cdot \frac{W}{n} \sum_{i=1}^N \frac{1}{W_i}. \end{aligned}$$

For simplicity in comparison, we assume in this section that $\frac{N}{n} = k$ is an integer. Then Rao/Hartley/Cochran (RHC, say, 1962) scheme may be employed taking n groups of size k each using p_i 's as normed size-measures with P_i 's as their group sums formed in sampling. For our 'randomized' survey, an appropriate estimator for this scheme is $e_2 = \frac{1}{c} \cdot \sum_{i \in s} \frac{z_i}{p_i} \cdot P_i$

$$\begin{aligned} \text{with } E_R E_P (e_2) = Y, \quad V_{Rp}(e_2) &= \frac{f_1}{n} \sum_{i=1}^N p_i \left(\frac{y_i}{p_i} - Y \right)^2 \\ &+ \frac{1-c}{c} \left[\frac{f_1}{n} \sum_{i=1}^N \frac{y_i^2}{p_i} + f_2 \sum_{i=1}^N y_i^2 \right] \\ &+ \frac{1-c}{c^2} \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \left[\frac{f_1}{n} \sum_{i=1}^N \frac{1}{p_i} + N f_2 \right] \end{aligned}$$

$$\text{writing } f_1 = \frac{N-n}{N-1}, \quad f_2 = \frac{N(n-1)}{n(N-1)},$$

$$\begin{aligned} E_2 = E_M V_{Rp} (e_2) &= \frac{\sigma^2}{n} \cdot f_1 \left(W \sum_{i=1}^N w_i^{g-1} - \sum_{i=1}^N w_i^g \right) \\ &+ \frac{1-c}{c} \left[f_2 \sigma^2 \sum_{i=1}^N w_i^2 + f_1 \frac{\sigma^2 W^2}{n} + f_2 \sigma^2 \sum_{i=1}^N w_i^g \right. \\ &\quad \left. + f_1 \sigma^2 \frac{W}{n} \sum_{i=1}^N w_i^{g-1} \right] \\ &+ \frac{1-c}{c^2} \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \left[\frac{f_1}{n} \sum_{i=1}^N \frac{1}{p_i} + N f_2 \right]. \end{aligned}$$

Using Lahiri/Midzuno/Sen (LMS, say, 1951, 1952, 1953) sampling scheme with p_i 's as above, for our 'randomized' survey, the modified estimator is $e_3 = \frac{1}{c} \sum_{i \in S} z_i / \sum_{i \in S} p_i$ with $E_R E_P (e_3) = Y$,

$$V_{Rp}(e_3) = \sum_{i=1}^N T_i y_i^2 + \sum_{i \neq j=1}^N T_{ij} y_i y_j - Y^2$$

$$+ \frac{1-c}{c} \sum_{i=1}^N T_i y_i^2 + \frac{1-c}{c^2} \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \sum_{i=1}^N T_i$$

and

$$E_3 = E_M V_{Rp}(e_3) = \sigma^2 \sum_{i=1}^N (T_i - 1) W_i^g + \frac{1-c}{c} \sum_{i=1}^N T_i (\sigma^2 W_i^2 + \sigma^2 W_i^g)$$

$$+ \frac{1-c}{c^2} \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \sum_{i=1}^N T_i$$

Using T.J. Rao's (1967) notations : T_i is average of $\left(\frac{1}{p_i + p_{i_2} + \dots + p_{i_n}} \right)$'s over choice of distinct units $i_2, \dots, i_n$ out of $1, \dots, N$ except i and T_{ij} is average of $\left(\frac{n-1}{N-1} \cdot \frac{1}{p_i + p_j + p_{i_3} + \dots + p_{i_n}} \right)$'s over choice of distinct units $i_3, \dots, i_n$ from $1, \dots, N$ except i and j and

$$\sum_{i=1}^N T_i p_i = \frac{N}{n}, \quad \sum_{i=1}^N T_i p_i^2 + \sum_{i \neq j=1}^N T_{ij} p_i p_j = 1 \quad \int$$

For Hansen-Hurwitz (HH, say, 1943) sampling scheme with same p_i 's the modified estimator is

$$e_4 = \frac{1}{c} \cdot \frac{1}{n} \sum_{i \in s} \frac{z_i}{p_i}, \quad \text{with } E_R E_p(e_4) = Y,$$

$$V_{Rp}(e_4) = \frac{1}{n} \sum_{i=1}^N p_i \left(\frac{y_i}{p_i} - Y \right)^2 + \frac{1-c}{c} \cdot \frac{1}{n} \sum_{i=1}^N \frac{y_i^2}{p_i} \\ + \frac{1-c}{c^2} \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \frac{1}{n} \sum_{i=1}^N \frac{1}{p_i} ,$$

$$\text{and } E_4 = E_M V_{Rp}(e_4) = \frac{\sigma^2}{n} \left(W \sum_{i=1}^N W_i^{g-1} - \sum_{i=1}^N W_i^g \right) \\ + \frac{1-c}{c} \left(\frac{\sigma^2 W^2}{n} + \sigma^2 \frac{W}{n} \sum_{i=1}^N W_i^{g-1} \right) \\ + \frac{1-c}{c^2} \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \frac{1}{n} \sum_{i=1}^N \frac{1}{p_i} .$$

With $n=2$, $s = (i, j)$ and p_i 's as above, Murthy's (1957) estimator is modified as

$$e_5 = \frac{1}{c} \cdot \frac{1}{2-p_1-p_j} \left[\frac{z_i}{p_i} (1-p_j) + \frac{z_j}{p_j} (1-p_i) \right]$$

with $E_R E_p(e_5) = Y$,

$$V_{Rp}(e_5) = \frac{1}{2} \sum_{i \neq j=1}^N \sum_{j=1}^N p_i p_j \frac{(1-p_1-p_j)}{(2-p_1-p_j)} \left(\frac{y_i}{p_i} - \frac{y_j}{p_j} \right)^2 \\ + \frac{1-c}{c} \cdot \sum_{i \neq j=1}^N \sum_{j=1}^N \frac{1}{(2-p_1-p_j)} \cdot \frac{p_j(1-p_j)}{p_i(1-p_i)} y_i^2 \\ + \frac{1-c}{c^2} \cdot \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \sum_{i \neq j=1}^N \sum_{j=1}^N \frac{1}{(2-p_1-p_j)} \cdot \frac{p_j(1-p_j)}{p_i(1-p_i)}$$

$$\begin{aligned} \text{and } E_5 \quad E_M V_{Rp}(e_5) &= \sigma^2 \sum_{i \neq j=1}^N \sum W_i^{g-1} W_j \frac{(1-p_i-p_j)}{(2-p_i-p_j)} \\ &+ \frac{1-c}{c} \cdot \sum_{i \neq j=1}^N \sum \frac{\sigma^2 W_i^g + \beta^2 W_i^2}{(2-p_i-p_j)} \cdot \frac{p_j (1-p_j)}{p_i (1-p_i)} \\ &+ \frac{1-c}{c^2} \cdot \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \cdot \sum_{i \neq j=1}^N \sum \frac{1}{(2-p_i-p_j)} \cdot \frac{p_j (1-p_j)}{p_i (1-p_i)}. \end{aligned}$$

To compare among them [c.f. J.N.K. Rao (1966), Chaudhuri/Arnab (1979) in case of 'direct' surveys]

let us write $N \text{Cov}(x_i, y_i) = \sum_{i=1}^N (x_i - \bar{x}) y_i$ for any

x_i, y_i 's ($i=1, \dots, N$) with $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$,

$$u_i = f_1 \cdot \frac{W}{n} - T_i W_i \div f_2 W_i \quad (\text{which decreases with } W_i),$$

$$\eta_i = \frac{\sigma^2}{c} W_i^{g-1} + \frac{1-c}{c} \beta^2 W_i + \frac{1-c}{c^2} \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \frac{1}{W_i}.$$

We note regarding $\text{Cov}(\eta_i, W_i)$

$$\frac{\partial \eta_i}{\partial W_i} = \frac{\sigma^2}{c} \cdot (g-1) W_i^{g-2} + \frac{1-c}{c} \cdot \beta^2 - \frac{1}{c} \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \frac{1}{W_i^2} \quad \geq 0$$

as $g \geq 1$ and $\beta^2 \geq \frac{1}{c} \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \frac{1}{W_i^2}$ together.

Remark. In case of 'direct' survey, $c=1$ and terms with multiplier $(1-c)$ will be absent and hence the above conditions will be in terms of g only and will not involve β .

Noting $D_1 = E_2 - E_1 = Nf_2 \text{Cov}(\eta_1, W_1)$, we have

$$D_1 \geq 0 \text{ as } g \geq 1 \text{ and } \beta^2 \geq \frac{1}{c} \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \frac{1}{W_1^2} \quad \forall i.$$

$$\begin{aligned} \text{Again, } D_2 &= E_3 - E_1 = \frac{1}{c} \sigma^2 \sum_{i=1}^N (T_i W_i - \frac{W}{n}) W_i^{g-1} \\ &\quad + (1-c) \beta^2 \sum_{i=1}^N (T_i W_i - \frac{W}{n}) W_i \\ &\quad + \frac{1-c}{c} \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \sum_{i=1}^N (T_i W_i - \frac{W}{n}) \frac{1}{W_i} \quad \mathcal{J} \\ &= \sum_{i=1}^N (T_i W_i - \frac{W}{n}) \eta_i = N \text{Cov}(\eta_i, T_i W_i), \text{ say,} \end{aligned}$$

$$\text{noting } \sum_{i=1}^N T_i W_i = \frac{NW}{n}.$$

Then, $D_2 \geq 0$ as $g \geq 1$ together with

$$\beta^2 \geq \frac{1}{c} \cdot \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \frac{1}{W_1^2} \quad \forall i, \text{ noting } T_i W_i \uparrow W_i;$$

$$\begin{aligned}
 D_3 &= E_2 - E_3 = \frac{N}{c} \left[\sigma^2 \text{Cov}(u_i, w_i^{g-1}) \right. \\
 &\quad + (1-c) \rho^2 \text{Cov}(u_i, w_i) \\
 &\quad \left. + \frac{(1-c)}{c} \cdot \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \text{Cov}(u_i, \frac{1}{\bar{w}_i}) \right] \\
 &= N \text{Cov}(u_i, \eta_i) \leq 0 \quad \text{as } g \geq 1
 \end{aligned}$$

along with $\rho^2 \geq \frac{1}{c} \cdot \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \cdot \frac{1}{w_i^2} \quad \forall i.$

Writing $\bar{w} = \frac{1}{N} \sum_{i=1}^N w_i$, $H = \frac{N}{\sum_{i=1}^N \frac{1}{\bar{w}_i}}$, $\sigma_w^2 = \frac{1}{N} \sum_{i=1}^N (w_i - \bar{w})^2$

we note that

$$E_2 = E_1 \text{ if } g=1 \text{ and } \rho^2 \sigma_w^2 + \frac{1}{c} \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) (1 - \frac{\bar{w}}{H}) = 0$$

$$\begin{aligned}
 E_3 &= E_1 \text{ if } g=1 \text{ and } \rho^2 \text{Cov}(T_i w_i, w_i) \\
 &\quad + \frac{1}{c} \cdot \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \text{Cov}(T_i w_i, \frac{1}{\bar{w}_i}) = 0
 \end{aligned}$$

$$\begin{aligned}
 E_3 &= E_2 \text{ if } g=1 \text{ and } \rho^2 \text{Cov}(u_i, w_i) \\
 &\quad + \frac{1}{c} \cdot \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \text{Cov}(u_i, \frac{1}{\bar{w}_i}) = 0.
 \end{aligned}$$

Further, we may check the following :

$$E_4 - E_1 = \sigma^2 \left(1 - \frac{1}{n}\right) \sum_{i=1}^N w_i^g > 0 \quad \forall g;$$

$$\begin{aligned} E_4 - E_2 &= \sigma^2 \left(1 - \frac{1}{n}\right) \sum_{i=1}^N w_i^g - f_2 \left[\frac{\sigma^2}{c} \sum_{i=1}^N (w_i - \bar{w}) w_i^{g-1} \right. \\ &\quad \left. + \frac{1-c}{c} \cdot \rho^2 \sum_{i=1}^N (w_i - \bar{w}) w_i \right. \\ &\quad \left. + \frac{1-c}{c^2} \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \sum_{i=1}^N (w_i - \bar{w}) \frac{1}{w_i} \right] \\ &= \sigma^2 \left(1 - \frac{1}{n}\right) \sum_{i=1}^N w_i^g - N f_2 \text{Cov}(\eta_i, w_i) > 0 \end{aligned}$$

if $g < 1$ simultaneously with $\rho^2 < \frac{1}{c} \cdot \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \frac{1}{w_i^2} \quad \forall i$;

$$\begin{aligned} E_4 - E_3 &= \sigma^2 \left(1 - \frac{1}{n}\right) \sum_{i=1}^N w_i^g - \left[\frac{\sigma^2}{c} \sum_{i=1}^N (T_i w_i - \frac{w}{n}) w_i^{g-1} \right. \\ &\quad \left. + \frac{1-c}{c} \cdot \rho^2 \sum_{i=1}^N (T_i w_i - \frac{w}{n}) w_i \right. \\ &\quad \left. + \frac{1-c}{c^2} \cdot \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \sum_{i=1}^N (T_i w_i - \frac{w}{n}) \frac{1}{w_i} \right] \\ &= \sigma^2 \left(1 - \frac{1}{n}\right) \sum_{i=1}^N w_i^g - N \text{Cov}(T_i w_i, \eta_i) > 0 \end{aligned}$$

if $g < 1$ along with $\rho^2 < \frac{1}{c} \cdot \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \frac{1}{w_i^2} \quad \forall i$.

To compare E_5 with others we assume $n=2$ for each and get (on writing $q_i = 1-p_i$ and

$$\eta_{ij} = \frac{\sigma^2}{c} \cdot \frac{W_i^{g-1} W_j}{q_i} + \frac{1-c}{c} \cdot \frac{\beta^2 W_i W_j}{q_i} + \frac{1-c}{c^2} \cdot \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \frac{W_j}{W_i q_i}$$

$$\begin{aligned} E_5 - E_1 &= \frac{\sigma^2}{c} \cdot \sum_{i \neq j=1}^N \left(\frac{q_j}{q_i + q_j} - \frac{1}{2} \right) \frac{W_i^{g-1} W_j}{q_i} \\ &\quad + \frac{1-c}{c} \cdot \beta^2 \sum_{i \neq j=1}^N \left(\frac{q_j}{q_i + q_j} - \frac{1}{2} \right) \frac{W_i W_j}{q_i} \\ &\quad + \frac{1-c}{c^2} \cdot \left(\frac{1}{M} \sum_{j=1}^M l_j^2 \right) \sum_{i \neq j=1}^N \left(\frac{q_j}{q_i + q_j} - \frac{1}{2} \right) \frac{W_j}{W_i q_i} \\ &= N(N-1) \text{Cov} \left(\frac{q_j}{q_i + q_j}, \eta_{ij} \right), \end{aligned}$$

(on noting $\sum_{i \neq j=1}^N \frac{q_j}{q_i + q_j} = \frac{N(N-1)}{2}$ and extending

notations for covariance involving double summations). But we are unable to conclude about its sign. Similarly, inconclusive are the signs of $E_5 - E_i$ ($i = 2, 3, 4$), expressions for which we omit.

Remark. As opposed to simple criteria for comparison among 'direct' survey strategies (vide J.N.K. Rao (1966) and

Chaudhuri/Arnab (1979) among others), in terms of g alone, with 'randomized' surveys we have more complicated ones involving additional unknown parameters. Thus although we have simple modifications on certain optimality results with 'randomized' as against 'direct' surveys, we have additional difficulties in choosing among the popular strategies modification on 'randomization'.

4.5 VARIANCE ESTIMATION

Writing $\text{Var}(e_1)$ as

$$\begin{aligned} V_{Rp}(e_1) &= E_R V_p(e_1) + V_R(Z/c) \\ &= \frac{1}{c^2} E_R \left[\sum_{i < j=1}^N \sum (\pi_i \pi_j - \pi_{ij}) \left(\frac{z_i}{\pi_i} - \frac{z_j}{\pi_j} \right)^2 \right] + \frac{1-c}{c} \sum_{i=1}^N y_i^2 \\ &\quad + \frac{1-c}{c^2} \cdot \frac{N}{M} \sum_{j=1}^M l_j^2 \end{aligned}$$

we have the following unbiased estimators for $V_{Rp}(e_1)$ (assuming $\pi_{ij} > 0 \forall i \neq j$)

$$\begin{aligned} \hat{V}_1(e_1) &= \frac{1}{c^2} \sum_{i < j \in s} \frac{(\pi_i \pi_j - \pi_{ij})}{\pi_{ij}} \left(\frac{z_i}{\pi_i} - \frac{z_j}{\pi_j} \right)^2 + \frac{1-c}{c^2} \sum_{i \in s} \frac{z_i^2}{\pi_i} \\ &\quad + \frac{1-c}{c} \cdot \frac{N}{M} \sum_{j=1}^M l_j^2 \end{aligned}$$

$$\hat{V}_2(e_1) = \frac{1}{c^2} \sum_{i < j} \sum_s \frac{(\pi_i \pi_j - M_2 p(s))}{M_2 p(s)} \left(\frac{z_i}{\pi_i} - \frac{z_j}{\pi_j} \right)^2$$

$$+ \frac{1-c}{c^2} \sum_{i \in s} \frac{z_i^2}{\pi_i} + \frac{1-c}{c} \cdot \frac{N}{M} \sum_{j=1}^M 1_j^2$$

$$\hat{V}_3(e_1) = \frac{1}{c^2} \sum_{i < j} \sum_s \frac{(\pi_i \pi_j p(s|i,j) - p(s))}{p(s)} \left(\frac{z_i}{\pi_i} - \frac{z_j}{\pi_j} \right)^2$$

$$+ \frac{1-c}{c^2} \sum_{i \in s} \frac{z_i^2}{\pi_i} + \frac{1-c}{c} \cdot \frac{N}{M} \sum_{j=1}^M 1_j^2$$

(applicable for $n > 2$, writing $p(s|i,j)$ for the conditional probability of selecting s given that the units i,j have been selected on the first two draws)

$$\hat{V}_4(e_1) = \frac{1}{c^2} \sum_{i \in s} z_i^2 \left(\frac{1-\pi_i}{\pi_i^2} \right) + \sum_{i \neq j} \sum_s \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij} \pi_i \pi_j} z_i z_j$$

$$+ \frac{1-c}{c^2} \sum_{i \in s} \frac{z_i^2}{\pi_i} + \frac{1-c}{c} \cdot \frac{N}{M} \sum_{j=1}^M 1_j^2$$

$$= \sum_{i \in s} z_i^2 \left(\frac{1-c \pi_i}{c^2 \pi_i^2} \right) + \frac{1}{c^2} \sum_{i \neq j} \sum_s \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij} \pi_i \pi_j} z_i z_j$$

$$+ \frac{1-c}{c} \cdot \frac{N}{M} \sum_{j=1}^M 1_j^2.$$

If $\pi_i = n p_i$ and if the computation of $p(s)$ is simpler than that of the $\frac{n(n-1)}{2}$ distinct π_{ij} 's, $\hat{V}_2(e_1)$ becomes computationally simpler than $\hat{V}_1(e_1)$. We may note that $\hat{V}_1(e_1)$, $\hat{V}_2(e_1)$ and $\hat{V}_3(e_1)$ are unbiased for any fixed sample size (n) design whereas $\hat{V}_4(e_1)$ is unbiased for any design.

Writing $\text{Var}(e_2)$ as

$$\begin{aligned} V_{Rp}(e_2) &= E_R V_p(e_2) + V_R(Z/c) \\ &= \frac{1}{c^2} E_R \left[\frac{N-n}{N-1} \cdot \frac{1}{n} \sum_{i=1}^N p_i \left(\frac{z_i}{p_i} - Z \right)^2 \right] \\ &\quad + \frac{1-c}{c} \sum_{i=1}^N y_i^2 + \frac{1-c}{c^2} \cdot \frac{N}{M} \sum_{j=1}^M l_j^2 \end{aligned}$$

we have an unbiased estimator for $V_{Rp}(e_2)$ given by

$$\begin{aligned} \hat{V}_{Rp}(e_2) &= \frac{1}{c^2} \cdot \frac{N-n}{n-1} \cdot \frac{1}{N} \sum_{i \in S} p_i \left(\frac{z_i}{p_i} - c e_2 \right)^2 \\ &\quad + \frac{1-c}{c^2} \sum_{i \in S} \frac{z_i^2}{p_i} \cdot p_i + \frac{1-c}{c} \cdot \frac{N}{M} \sum_{j=1}^M l_j^2 \\ &= \frac{N-n}{n-1} \cdot \frac{1}{N} \sum_{i \in S} p_i \left(\frac{z_i}{c p_i} - e_2 \right)^2 + \frac{1-c}{c^2} \sum_{i \in S} \frac{z_i^2}{p_i} \cdot p_i + \frac{1-c}{c} \cdot \frac{N}{M} \sum_{j=1}^M l_j^2 \end{aligned}$$

which is uniformly non-negative.

Writing Var (e_3) as

$$\begin{aligned} V_{Rp}(e_3) &= E_R V_p(e_3) + V_R(Z/c) \\ &= \frac{1}{c^2} E_R \left[\frac{1}{\left(\frac{N-1}{n-1} \right)} \sum_s \frac{\left(\sum_{i \in s} z_i \right)^2}{\sum_{i \in s} p_i} - Z^2 \right] \\ &\quad + \frac{1-c}{c} \cdot \frac{N}{\sum_{i=1}^N} y_i^2 + \frac{1-c}{c^2} \cdot \frac{N}{M} \sum_{j=1}^M 1_j^2 \end{aligned}$$

we have following Raj (1954) an unbiased estimator for $V_{Rp}(e_3)$ as follows

$$\begin{aligned} \hat{V}_{Rp}(e_3) &= e_3^2 - \frac{1}{c^2 p(s)} \left[\frac{1}{M_1} \sum_{i \in s} z_i^2 + \frac{1}{M_2} \sum_{i+j \in s} z_i z_j \right] \\ &\quad + \frac{1-c}{c^2} \cdot \frac{\sum_{i \in s} z_i^2}{\sum_{i \in s} p_i} + \frac{1-c}{c} \cdot \frac{N}{M} \sum_{j=1}^M 1_j^2. \end{aligned}$$

Incidentally it may be noted that this $V_{Rp}(e_3)$ should often turn out negative because it is known that Raj's variance estimator (for ratio estimator based on LMS - scheme) is negative for samples with high selection - probabilities.

Following Rao (1979) and noting that $E_R V_p(e_3)$ equals 0, when $z_i = \lambda p_i$, $i=1, \dots, N$ (for some $\lambda \neq 0$), it

can be written as

$$E_R V_p(e_3) = \frac{1}{c^2} E_R \left[- \sum_{i < j=1}^N \sum d_{ij} p_i p_j \left(\frac{z_i}{p_i} - \frac{z_j}{p_j} \right)^2 \right]$$

$$\text{where } d_{ij} = E_p (d_{is} d_{js}) - 1$$

$$= \sum_{s \rightarrow i, j} d_{is} d_{js} p(s) - 1$$

$$= e_{ij} - 1, \text{ say,}$$

$$\text{where } d_{is} = 0 \text{ if } s \nrightarrow i$$

$$= \frac{1}{\sum_{i \in s} p_i} \text{ if } s \rightarrow i.$$

Then, non-negative quadratic unbiased estimator for

$E_R V_p(e_3)$ is necessarily of the form

$$- \frac{1}{c^2} \sum_{i < j \in s} \sum d_{ij}(s) p_i p_j \left(\frac{z_i}{p_i} - \frac{z_j}{p_j} \right)^2$$

$$\text{where } d_{ij}(s) = 0 \text{ if } s \nrightarrow i, j \text{ and } \sum_{s \rightarrow i, j} d_{ij}(s) p(s) = d_{ij}$$

$i < j$. We may consider three classes of estimators for

$E_R V_p(e_3)$ of the above-noted form as follows [c.f. Chaudhuri (1981)] :

$$C_1 : d_{ij}(s) = d_{ij} \left[\frac{\alpha}{\pi_{ij}} + \frac{1-\alpha}{M_2 p(s)} \right]$$

$$C_2 : d_{ij}(s) = e_{ij} \left[\frac{\alpha}{\pi_{ij}} + \frac{1-\alpha}{M_2 p(s)} \right] - \left[\frac{\beta}{\pi_{ij}} + \frac{1-\beta}{M_2 p(s)} \right]$$

$$C_3 : d_{ij}(s) = d_{is} d_{js} - \left[\frac{\lambda^*}{\pi_{ij}} + \frac{1-\lambda^*}{M_2 p(s)} \right]$$

where $0 \leq \alpha, \beta, \lambda^* \leq 1$ and

$$\pi_{ij} = \sum_{s \in I, j} p(s) = \frac{(n-1)(n-2)}{(N-1)(N-2)} + \frac{n-1}{N-1} \cdot \frac{N-n}{N-2} (p_1 + p_j).$$

In any particular situation, values of α, β, λ^* are to be appropriately hit upon (prior to sampling) to render the estimator for $E_R V_p(e_3)$ non-negative, other parameters like e_{ij} 's being fully known. Whether such choice are possible seem difficult to investigate.

If however a non-negative quadratic unbiased estimator for $E_R V_p(e_3)$ is thus derived, we get a non-negative unbiased estimator for $V_{Rp}(e_3)$ as

$$\begin{aligned} \hat{V}_{Rp}(e_3) &= -\frac{1}{c^2} \sum_{i \in S} \sum_{j \in S} d_{ij}(s) p_i p_j \left(\frac{z_i}{p_i} - \frac{z_j}{p_j} \right)^2 \\ &\quad + \frac{1-c}{c^2} \cdot \frac{\sum_{i \in S} z_i^2}{\sum_{i \in S} p_i} + \frac{1-c}{c} \cdot \frac{N}{M} \sum_{j=1}^M l_j^2. \end{aligned}$$

Writing $\text{Var}(e_4)$ as

$$\begin{aligned} V_{Rp}(e_4) &= E_R V_p(e_4) + V_R(Z/c) \\ &= \frac{1}{c^2} E_R \left[\frac{1}{n} \sum_{i=1}^N p_i \left(\frac{z_i}{p_i} - Z \right)^2 \right] + \frac{1-c}{c} \cdot \sum_{i=1}^N y_i^2 \\ &\quad + \frac{1-c}{c^2} \cdot \frac{N}{M} \sum_{j=1}^M l_j^2 \end{aligned}$$

we have an unbiased estimator for $V_{Rp}(e_4)$ given by

$$\begin{aligned} \hat{V}_{Rp}(e_4) &= \frac{1}{c^2} \cdot \frac{1}{n(n-1)} \sum_{i \in S} \left(\frac{z_i}{p_i} - c e_4 \right)^2 \\ &\quad + \frac{1-c}{c^2} \cdot \sum_{i \in S} \frac{z_i^2}{p_i} + \frac{1-c}{c} \cdot \frac{N}{M} \sum_{j=1}^M l_j^2 \\ &= \frac{1}{n(n-1)} \sum_{i \in S} \left(\frac{z_i}{c p_i} - e_4 \right)^2 + \frac{1-c}{c^2} \cdot \sum_{i \in S} \frac{z_i^2}{p_i} + \frac{1-c}{c} \cdot \frac{N}{M} \sum_{j=1}^M l_j^2 \end{aligned}$$

which is uniformly non-negative.

Writing $\text{Var}(e_5)$ as

$$\begin{aligned} V_{Rp}(e_5) &= E_R V_p(e_5) + V_R(Z/c) \\ &= \frac{1}{c^2} E_R \left[\sum_{i < j=1}^N p_i p_j \frac{(1-p_i-p_j)}{(2-p_i-p_j)} \left(\frac{z_i}{p_i} - \frac{z_j}{p_j} \right)^2 \right] \\ &\quad + \frac{1-c}{c} \cdot \sum_{i=1}^N y_i^2 + \frac{1-c}{c^2} \cdot \frac{N}{M} \sum_{j=1}^M 1_j^2 \end{aligned}$$

we have an unbiased estimator for $V_{Rp}(e_5)$ given by

$$\begin{aligned} \hat{V}_{Rp}(e_5) &= \frac{1}{c^2} \cdot \frac{(1-p_i)(1-p_j)(1-p_i-p_j)}{(2-p_i-p_j)^2} \left(\frac{z_i}{p_i} - \frac{z_j}{p_j} \right)^2 \\ &\quad + \frac{1-c}{c^2} \cdot \frac{1}{(2-p_i-p_j)} \left[\frac{z_i^2}{p_i} (1-p_j) + \frac{z_j^2}{p_j} (1-p_i) \right] \\ &\quad + \frac{1-c}{c} \cdot \frac{N}{M} \sum_{j=1}^M 1_j^2 \end{aligned}$$

which is uniformly non-negative.

Appendix-1

The condition (2.3.20) is apparently too stringent although it is satisfied if p_2 and p_1 both correspond to simple random sampling without replacement. The following artificial example illustrates the existence of varying probability sampling designs meeting this requirement.

Let $N=4$; $s_1' = (1,2,3)$, $s_2' = (1,2,4)$, $s_3' = (2,3,4)$

be the first phase samples with $p_1(s_1') = 0.30$, $p_1(s_2') = 0.40$, $p_1(s_3') = 0.30$. Let the second phase samples from $s_i' (i=1,2,3)$ and their selection - probabilities be :

$s_{11}=(s_1|s_1')=(1,2)$, $s_{21}=(s_2|s_1')=(1,3)$, $s_{31}=(s_3|s_1')=(2,3)$,

$p_2(s_{11}) = 0.30$, $p_2(s_{21}) = 0.40$, $p_2(s_{31}) = 0.30$

$s_{12}=(s_1|s_2')=(1,2)$, $s_{22}=(s_2|s_2')=(1,4)$, $s_{32}=(s_3|s_2')=(2,4)$,

$p_2(s_{12}) = 0.30$, $p_2(s_{22}) = 0.40$, $p_2(s_{32}) = 0.30$

$s_{13}=(s_1|s_3')=(2,3)$, $s_{23}=(s_2|s_3')=(2,4)$, $s_{33}=(s_3|s_3')=(3,4)$

$p_2(s_{13}) = 0.30$, $p_2(s_{23}) = 0.30$, $p_2(s_{33}) = 0.40$.

Then, $P_1 = 0.70$, $P_2 = 1.00$, $P_3 = 0.60$, $P_4 = 0.70$;

$$Q_1 = 0.49, \quad Q_2 = 0.60, \quad Q_3 = 0.42, \quad Q_4 = 0.49 \quad \text{giving}$$

$$Q_1/P_1 = 0.70, \quad Q_2/P_2 = 0.60, \quad Q_3/P_3 = 0.70, \quad Q_4/P_4 = 0.70.$$

It is readily checked that the above $p_2(s)$'s, $p_1(s')$'s satisfy (2.3.20).

Appendix - 2

Theorem 2.3.7. In C^n , e^* is 'admissible' for any chosen design p .

Proof. For any e in C^n , we have

$$\begin{aligned} V_p(e) &= V_{p_1} E_{p_2} [e_1(s, \underline{U}) + e_2(s', \underline{X})] + E_{p_1} V_{p_2} [e_1(s, \underline{U})] \\ &= V_p [e_1(s, \underline{U})] + V_{p_1} [e_2(s', \underline{X})] \end{aligned}$$

and similarly

$$V_p(e^*) = V_p \left[\sum_{j \in s} \frac{u_j}{Q_j} \right] + V_{p_1} \left[\sum_{j \in s'} \frac{x_j}{P_j} \right].$$

$$\text{So, } V_p(e) - V_p(e^*)$$

$$\begin{aligned} &= V_p [e_1(s, \underline{U})] - V_p \left[\sum_{j \in s} \frac{u_j}{Q_j} \right] + V_{p_1} [e_2(s', \underline{X})] \\ &\quad - V_{p_1} \left[\sum_{j \in s'} \frac{x_j}{P_j} \right]. \end{aligned}$$

Let for a point $\underline{U}$, $u_j \neq 0$, $u_k = 0 \quad \forall k \neq j$.

Then $\Sigma_1 \Sigma_2 e_1(s, \underline{U}) p_2(s) p_1(s') = u_j$ and at such a point

$$\begin{aligned} V_p [e_1(s, \underline{U})] &= V_p \left[\sum_{j \in S} \frac{u_j}{Q_j} \right] \\ &= \Sigma_1 \Sigma_2 e_1^2(s, \underline{U}) p_2(s) p_1(s') - \frac{u_j^2}{Q_j} \geq 0 \end{aligned}$$

by Cauchy - Schwarz's inequality.

Again from Godambe/Joshi(1965) it follows that

$$V_{p_1} [e_2(s', \underline{X})] > V_{p_1} \left[\sum_{j \in S'} \frac{x_j}{P_j} \right]$$

for at least one $\underline{X}$ in R^N (since $\sum_{j \in S'} \frac{x_j}{P_j}$ is 'admissible' in the class of all unbiased estimators for X for any design).

So, at least for a point $(\underline{X}, \underline{Y})$ such that $y_j \neq 0$, $x_j \neq 0$, $y_j - x_j \neq 0$, $y_k = x_k = 0 \quad \forall k \neq j$ and for e not identically same as e^* for samples with $p(\bar{s}) > 0$, we have

$$V_p(e) - V_p(e^*) > 0.$$

Hence e^* is admissible in C^n . Q.E.D.

Appendix - 3

$$\begin{aligned}
 E_p C_M(e^*, h) &= E_M E_p \left[\sum_{j \in s} \frac{e_j - e'_j}{Q_j} + \sum_{j \in s'} \frac{e'_j}{P_j} \right] h(\bar{s}, y, x') \Big] \\
 &= E_M E_{p_1} \left[\sum_{s \subset s'} \left(\sum_{j \in s} \frac{e_j - e'_j}{Q_j} \right) h(\bar{s}, y, x') p_2(s) \right. \\
 &\quad \left. + \sum_{j \in s'} \frac{e'_j}{P_j} \sum_{s \subset s'} h(\bar{s}, y, x') p_2(s) \right] \\
 &= E_M E_{p_1} \left[\sum_{j \in s'} \frac{e_j - e'_j}{Q_j} \sum_{s \rightarrow j | s \subset s'} h(\bar{s}, y, x') p_2(s) \right] \\
 &\quad \text{using (2.3.21)} \\
 &= E_M E_{p_1} \left[\sum_{j \in s'} \frac{e_j - e'_j}{Q_j} \left\{ - \sum_{s \nrightarrow j | s \subset s'} h_1(s, y, x) p_2(s) - h_2(s', x') \right. \right. \\
 &\quad \left. \left. + h_2(s', x') \sum_{s \rightarrow j | s \subset s'} p_2(s) \right\} \right] \text{ using (2.3.30)} \\
 &= - E_M E_{p_1} \left[\sum_{j \in s'} \frac{e_j - e'_j}{Q_j} \left\{ \sum_{s \nrightarrow j | s \subset s'} h_1(s, y, x) p_2(s) \right\} \right] \\
 &\quad + E_M \left[\sum_{j=1}^N \frac{e_j - e'_j}{Q_j} \left(\frac{Q_j}{P_j} - 1 \right) \left\{ \sum_{s \nrightarrow j} h_2(s', x') p_1(s') \right\} \right] \\
 &\quad \text{using (2.3.20)} \\
 &= E_M \left[\sum_{j=1}^N \frac{e_j - e'_j}{Q_j} \left(1 - \frac{Q_j}{P_j} \right) \left\{ \sum_{s \nrightarrow j} h_2(s', x') p_1(s') \right\} \right]
 \end{aligned}$$

= 0 by independence of (y_j, x_j) and

(y_k, x_k) 's for $j \neq k$ and using

$$\begin{aligned} E_{p_1} [h_2(s', x')] &= 0 \Rightarrow \sum_{s'} h_2(s', x') p_1(s') \\ &= - \sum_{s'} h_2(s', x') p_1(s'). \end{aligned}$$

Appendix - 4

Report on 'A randomized response survey on earnings other than salary and expenses on confidential items for some specified ISI workers in 1982'.

We developed certain theoretical results relating to sampling strategies with randomized response trials as presented in Chapter 4. In order to see how they work in practice we conducted a survey among some specified workers of our Institute (ISI), viz. the workers of its Computer Science Unit and Stat-Math Division.

We identified two items namely (1) income earned in excess over usual salary and (2) total expenditures on

certain items like alcoholic drinks, gambling etc. which are not usually approved in conservative Indian Societies as the ones for which responses are normally unlikely to be available with direct enquiries. So, 'randomized trials' using the device described below were made to get information on them. Stratification was done in order to apply our method for this presented in Chapter 4. In order to form an idea about the accuracy and reliability for this 'randomized' procedure as to be measured in terms of estimated co-efficients of variation we also get direct responses for each sampled person on two 'hopefully non-confidential' items like expenses incurred on conspicuous items (vide items numbered 3 and 4 in the Questionnaire canvassed as attached). Reference period for each was January 1 - June 30, 1982 and estimates were needed for each item total. Our stipulation is that if the estimated co-efficients of variation do not differ enormously then we may have some satisfaction about the performance of the 'randomized trial'.

Our population as stated above was divided into 5 strata on the basis of gross salaries for June, 1982. Their net take-home salaries were also recorded as auxiliary size-measures utilized at the selection and estimation stages.*

* Figures supplied by the Accounts Department as advised by the Director, ISI. This co-operation is thankfully appreciated.

The gross and net take-home salaries (for June, 1982) earned by the specified ISI workers are given in Table I below suppressing their identities for the sake of privacy. We started a complete enumeration anticipating refusals. Fortunately refusals were only a few namely 5 out of 89 - they are not included in Table I. Then in order to try out our proposed (vide Chapter 4) sampling strategy we took a (stratum-wise and hence over-all too) 50% sample. The stratum sizes are 20, 22, 10, 18 and 14 and the respective sample sizes are 10, 11, 5, 9 and 7. From each stratum an IPPS sample is drawn independently following Sengupta's (1981) scheme.

The randomizing device we used unobserved by me (the lone investigator for this survey) was as follows:

Each sampled person was approached with the questionnaire attached and provided with a set of 3 containers - one containing 44 red beads and 36 black beads with a hole in it. and the other two containing some cards (different in numbers) serially numbered on one side with additional numbers with sums zero inscribed on the reverse (used to apply the randomizing device, vide Chapter 4). To answer the first question (in the questionnaire given below) the respondent was asked

to shake the first container and see the colour of the bead through the hole of the container. If a red bead was noticed, he was asked to report his true additional income during last six months minus a constant (given to him as stated in Table I, to ensure privacy), called 'additional adjusted income' and otherwise select from the "second" container one of the cards bearing the "serial number" chosen at random using a table of random numbers supplied and report the value inscribed on the reverse. A similar procedure was adopted to answer the second question using instead of the "second" the "third" container.

Let z_i denote the response from the i th sampled individual on item 1. Then an unbiased estimate for the total of item 1 is

$$e_1 = \frac{1}{c} \sum_{i \in S} \frac{z_i}{\pi_i}$$

where $c = 0.55$. Its variance is unbiasedly estimated by [vide Chapter 4]

$$\begin{aligned} \hat{V}_{Rp}(e_1) &= \frac{1}{c^2} \sum_{i < j \in S} \sum_{s} \frac{(\pi_i \pi_j - \pi_{ij})}{\pi_{ij}} \left(\frac{z_i}{\pi_i} - \frac{z_j}{\pi_j} \right)^2 \\ &\quad + \frac{1-c}{c^2} \sum_{i \in S} \frac{z_i^2}{\pi_i} + \frac{1-c}{c} \cdot \frac{N}{M} \sum_{j=1}^M l_j^2. \end{aligned}$$

If the sample were drawn by RHC scheme using the same p_i 's another unbiased estimate would be $e_2 = \frac{1}{c} \cdot \sum_{i \in s} \frac{z_i}{p_i} \cdot p_i$, (vide Chapter 4), with

$$V_{Rp}(e_2) = \frac{1}{c^2} E_R \left[\frac{N-n}{N-1} \cdot \frac{1}{n} \left\{ \sum_{i=1}^N \frac{z_i^2}{p_i} - z^2 \right\} \right] \\ + \frac{1-c}{c} \cdot \sum_{i=1}^N y_i^2 + \frac{1-c}{c^2} \cdot \frac{N}{M} \sum_{j=1}^M l_j^2.$$

On the basis of the sample drawn by an IPPS method as above it may be unbiasedly estimated by

$$\hat{V}_{Rp}(e_2) = \frac{1}{c^2} \cdot \frac{N-n}{N-1} \cdot \frac{1}{n} \left[\sum_{i \in s} \frac{z_i^2}{p_i \pi_i} - c^2 \left\{ e_1^2 - \right. \right. \\ \left. \left. \frac{1}{c^2} \cdot \sum_{i < j \in s} \frac{(\pi_i \pi_j - \pi_{ij})}{\pi_{ij}} \left(\frac{z_i}{\pi_i} - \frac{z_j}{\pi_j} \right)^2 \right\} \right] \\ + \frac{1-c}{c^2} \cdot \sum_{i \in s} \frac{z_i^2}{\pi_i} + \frac{1-c}{c} \cdot \frac{N}{M} \sum_{j=1}^M l_j^2.$$

If the sample were drawn by Midzuno/Sen scheme using the same p_i 's another estimate would be $e_3 = \frac{1}{c} \cdot \sum_{i \in s} z_i / \sum_{i \in s} p_i$ (vide Chapter 4), with

$$V_{Rp}(e_3) = \frac{1}{c^2} E_R \left[\sum_{i=1}^N T_i z_i^2 + \sum_{i \neq j=1}^N \sum T_{ij} z_i z_j - z^2 \right] \\ + \frac{1-c}{c} \cdot \sum_{i=1}^N y_i^2 + \frac{1-c}{c^2} \cdot \frac{N}{M} \sum_{j=1}^M l_j^2.$$

For an IPPS sample drawn, it may be unbiasedly estimated by

$$\hat{V}_{Rp}(e_3) = \frac{1}{c^2} \left[\sum_{i \in s} \frac{T_i z_i^2}{\pi_i} + \sum_{i \neq j \in s} \frac{T_{ij} z_i z_j}{\pi_{ij}} - c^2 \left\{ e_1^2 - \right. \right. \\ \left. \left. \frac{1}{c^2} \cdot \sum_{i < j \in s} \frac{(\pi_i \pi_j - \pi_{ij})}{\pi_{ij}} \left(\frac{z_i}{\pi_i} - \frac{z_j}{\pi_j} \right)^2 \right\} \right] \\ + \frac{1-c}{c^2} \cdot \sum_{i \in s} \frac{z_i^2}{\pi_i} + \frac{1-c}{c} \cdot \frac{N}{M} \sum_{j=1}^M l_j^2.$$

Similarly we estimate the total of item 2. Since we get direct responses on items 3 and 4 their totals can be estimated using the above formulae replacing z_i by y_i and c by 1. The relevant estimates and the samples on which they are based along with other numerical data used are given in Tables I - IV below.

Table I

Giving gross and net take-home salaries (in Rupees) for June, 1982, for some ISI workers serially listed and the values reported on the Items 1,2,3,4 in the Questionnaire.

Stratum I

(with gross salary in the range 500-900; the constants to be subtracted respectively for Items 1 and 2 are 300 and 150)

Serial No.	Gross salary	Net take-home pay - packet	Values reported on Items			
			1	2	3	4
1	880.04	689.49	- 90	150	340	0
2	748.30	624.75	300	-150	0	0
3	764.75	622.14	-300	-150	15298	16
4	655.20	584.69	-290	-150	600	200
5	620.19	511.86	-110	-150	0	50
6	721.84	444.21	310	140	800	100
7	762.78	438.96	- 30	10	740	25
8	630.20	430.92	-300	- 40	65	5
9	610.13	428.76	- 10	- 20	0	10
10	539.91	426.56	-300	-150	300	0
11	816.22	419.39	-300	-150	2135	10
12	630.20	403.67	-235	90	7000	30
13	685.32	386.39	60	- 30	250	0
14	650.29	379.65	200	30	0	20

contd...

Stratum I (Contd.)

Serial No.	Gross salary	Net take-home pay - packet	Values reported on Items			
			1	2	3	4
15	810.22	352.76	900	50	6	10
16	700.38	347.67	300	-150	70	5
17	650.19	343.60	200	450	290	50
18	670.30	307.25	60	- 40	0	0
19	670.30	282.65	- 10	- 80	0	0
20	621.22	161.81	150	450	0	0

Stratum - II

(with gross salary in the range 900-1600; the constants to be subtracted for Items 1 and 2 are 550 and 350 respectively)

Serial No.	Gross salary	Net take-home pay - packet	Values reported on Items			
			1	2	3	4
1	1563.00	1453.58	-550	-350	550	100
2	1582.70	1379.00	- 60	50	3360	2000
3	1582.70	1379.00	-550	-100	200	0
4	1484.60	1375.42	-550	-350	1900	0
5	1396.20	1144.96	-340	-350	75	50
6	1582.70	1005.82	-280	190	76	5
7	1210.20	992.58	120	-350	30	15

contd...

Stratum II (contd.)

Serial No.	Gross salary	Net take-home pay - packet	Values reported on Items			
			1	2	3	4
8	1406.20	989.80	-550	-350	2000	100
9	1367.00	948.71	- 90	230	0	0
10	1101.60	948.08	- 80	30	130	25
11	1529.95	941.06	-250	-350	800	50
12	1024.15	881.79	-410	-340	1316	290
13	1075.85	881.14	- 50	-290	1250	50
14	1288.60	880.95	-220	-220	300	150
15	998.20	858.33	3510	- 70	2800	5
16	998.20	858.33	-600	-350	1096	0
17	998.20	858.33	-475	-350	508	0
18	1484.60	793.99	270	110	85	3000
19	1249.40	771.02	2150	-350	12000	230
20	1127.45	739.04	500	20	50	36
21	1582.70	560.04	1450	500	270	45
22	1075.85	528.91	320	-350	40280	15

Stratum-III

(with gross salary in the range 1600-2100; the constants to be subtracted for Items 1 and 2 are 300 and 150 respectively)

Serial No.	Gross salary	Net take-home pay - packet	Values reported on Items			
			1	2	3	4
1	1859.50	1629.58	400	-150	126	15
2	1984.70	1617.40	-300	-150	2620	300
3	1717.20	1594.42	700	-150	50	50
4	2076.70	1558.65	2700	1150	580	30
5	1717.20	1545.15	-300	40	1050	25
6	1920.50	1455.07	900	- 20	4004	20
7	1677.70	1427.61	10	-150	280	0
8	2063.25	1361.45	-300	-150	965	40
9	1663.85	1225.32	300	300	668	12
10	1987.50	1162.24	0	-150	670	35

Stratum-IV

(with gross salary in the range 2100-2600; the constants to be subtracted for Items 1 and 2 are 300 and 300 respectively)

Serial No.	Gross salary	Net take-home pay - packet	Values reported on Items			
			1	2	3	4
1	2496.70	2194.54	-300	60	500	50
2	2487.50	2158.34	-300	90	0	0
3	2540.70	2122.04	- 50	300	3813	98
4	2462.00	2091.58	300	-300	6265	50
5	2574.50	2088.24	-300	300	0	0
6	2457.50	1957.35	-300	- 40	301	10
7	2257.50	1887.16	-300	-300	4400	0
8	2446.60	1878.66	300	-300	11500	300
9	2539.00	1864.19	- 20	-300	2950	0
10	2539.90	1840.54	100	-300	0	2000
11	2559.10	1831.52	50	-300	4150	50
12	2507.50	1817.42	300	-300	30	0
13	2375.00	1774.37	1700	1500	2000	0
14	2145.00	1672.86	4700	300	2850	0
15	2295.00	1573.86	- 50	- 40	350	0
16	2559.10	1427.02	3400	500	4870	100
17	2559.10	1383.86	-200	300	0	0
18	2462.00	1316.96	100	300	980	900

Stratum-V

(with gross salary in the range 2600-4100; the constants to be subtracted for Items 1 and 2 are 350 and 300 respectively)

Serial No.	Gross salary	Net take-home pay - packet	Values reported on Items			
			1	2	3	4
1	4037.50	2791.70	-110	100	4150	0
2	3340.50	2736.42	-350	-300	3520	300
3	3113.50	2663.91	90	- 50	570	30
4	3224.70	2536.54	-150	150	0	0
5	3347.50	2457.46	380	-250	4150	300
6	2698.95	2353.53	- 60	-300	0	0
7	3199.70	2326.11	350	-300	300	50
8	2865.50	2287.34	-230	-300	4750	0
9	2651.50	2226.92	-150	50	500	0
10	2816.20	2182.82	3010	870	2450	100
11	2738.60	2057.09	-190	-300	400	20
12	2816.20	2003.24	-170	190	1300	0
13	2631.40	1976.67	-350	50	1100	60
14	2826.70	1763.65	60	-300	1200	100

The two anticipated sets of values (denoted as l_j 's) for answers to Items 1 and 2 in the Questionnaire for the respective strata are given as (respectively inscribed with suitable frequencies on cards in containers 2 and 3):

Strata I and III : (0, ± 10 , ± 20 , ± 30 , ± 40 , ± 50 , ± 60 , ± 70 , ± 80 , ± 90 , ± 100 , ± 110 , ± 120 , ± 130 , ± 140 , ± 150 , ± 160 , ± 170 , ± 180 , ± 190 , ± 200 , ± 300)

(0, ± 10 , ± 20 , ± 30 , ± 40 , ± 50 , ± 60 , ± 70 , ± 80 , ± 90 , ± 100 , ± 150)

Stratum II : (0, ± 10 , ± 20 , ± 30 , ± 40 , ± 50 , ± 60 , ± 70 , ± 80 , ± 90 , ± 100 , ± 110 , ± 120 , ± 130 , ± 140 , ± 150 , ± 160 , ± 170 , ± 180 , ± 190 , ± 200 , ± 210 , ± 220 , ± 230 , ± 240 , ± 250 , ± 260 , ± 270 , ± 280 , ± 290 , ± 300 , ± 310 , ± 320 , ± 330 , ± 340 , ± 350 , ± 360 , ± 370 , ± 380 , ± 390 , ± 400 , ± 410 , ± 420 , ± 430 , ± 440 , ± 450 , ± 550)

(0, ± 10 , ± 20 , ± 30 , ± 40 , ± 50 , ± 60 , ± 70 , ± 80 , ± 90 , ± 100 , ± 110 , ± 120 , ± 130 , ± 140 , ± 150 , ± 160 , ± 170 , ± 180 , ± 190 , ± 200 , ± 210 , ± 220 , ± 230 , ± 240 , ± 250 , ± 260 , ± 270 , ± 280 , ± 290 , ± 300 , ± 350)

Stratum IV : (0, ± 10 , ± 20 , ± 30 , ± 40 , ± 50 , ± 300)

(0, ± 10 , ± 20 , ± 30 , ± 40 , ± 50 , ± 60 , ± 70 , ± 80 , ± 90 ,
 ± 100 , ± 300)

Stratum V : (0, ± 10 , ± 20 , ± 30 , ± 40 , ± 50 , ± 60 , ± 70 ,
 ± 80 , ± 90 , ± 100 , ± 110 , ± 120 , ± 130 , ± 140 , ± 150 , ± 160 ,
 ± 170 , ± 180 , ± 190 , ± 200 , ± 210 , ± 220 , ± 230 , ± 240 , ± 250 ,
 ± 350)

(0, ± 10 , ± 20 , ± 30 , ± 40 , ± 50 , ± 60 , ± 70 , ± 80 , ± 90 ,
 ± 100 , ± 110 , ± 120 , ± 130 , ± 140 , ± 150 , ± 160 , ± 170 , ± 180 ,
 ± 190 , ± 200 , ± 300).

The units sampled from respective strata with serial numbers are as follows:

Stratum	The sampled units are :
I	1, 3, 4, 7, 8, 11, 13, 15, 16, 18
II	1, 2, 5, 7, 8, 9, 10, 11, 12, 14, 19
III	1, 3, 4, 8, 10
IV	1, 2, 3, 4, 6, 7, 8, 14, 15
V	1, 2, 3, 4, 6, 11, 14.

Table - II

Stratum-wise estimates (in Rupees) for Item 1 with measures of error

Stratum	e_1	$\hat{V}_{Rp}(e_1)$	$\hat{V}_{Rp}(e_2)$	$\frac{100x(\hat{V}_{Rp}(e_1))^{\frac{1}{2}}}{e_1}$
I	7832.91	3.8525155×10^7	1.7553295×10^7	79.24
II	13773.87	2.3216769×10^8	8.7228455×10^7	110.62
III	14644.22	7.2995975×10^7	6.4659514×10^7	58.34
IV	21795.40	3.6315027×10^8	2.6234211×10^8	87.43
V	2577.07	1.7076855×10^6	1.8613035×10^6	50.71

Table III

Stratum-wise estimates (in Rupees) for Item 2 with measures of error

Stratum	e_1	$\hat{V}_{Rp}(e_1)$	$\hat{V}_{Rp}(e_2)$	$\frac{100x(\hat{V}_{Rp}(e_1))^{\frac{1}{2}}}{e_1}$
I	1117.73	1.2811375×10^6	1.0036955×10^6	101.26
II	- 706.05	6.3581721×10^6	6.8204612×10^6	-357.13
III	3157.04	1.4895214×10^7	1.4153788×10^7	122.25
IV	4614.24	4.5184466×10^6	4.8170751×10^6	46.07
V	192.27	3.5787634×10^6	3.6366348×10^6	983.90

N.B. The estimate e_1 may turn out (as it has for stratum II and Item 2) negative because of the origin-shifting. An interpretation for the resulting estimate is this that the auxiliary values l_j 's are not chosen quite appropriately (rather they are possibly chosen larger than the true values on an average).

Table - IV

Stratum-wise estimates (in Rupees) for Items 3 and 4 with measures of error (the estimator is denoted as e_1' which is e_1 with z_1 as y_1 and c as 1 and its variance estimator as v).

Stratum	<u>Item 3</u>			<u>Item 4</u>		
	e_1'	v	$\frac{100 \times v^{\frac{1}{2}}}{e_1'}$	e_1'	v	$\frac{100 \times v^{\frac{1}{2}}}{e_1'}$
I	29129.92	2.1338043×10^8	50.15	431.84	4.5128985×10^4	49.19
II	44881.78	1.5514465×10^9	87.76	4919.92	3.2441121×10^6	36.61
III	5149.62	2.7316907×10^6	32.09	347.85	2.1328774×10^3	13.28
IV	56835.73	1.5630332×10^8	21.99	941.78	1.0957876×10^5	35.15
V	17847.02	1.9296164×10^7	24.61	865.70	7.8957212×10^4	32.46

Remark. Items 1 and 2 are estimated, as expected, less efficiently than items 3 and 4 - vide the estimated co-efficients of variation. Comparing $\hat{V}_{Rp}(e_1)$ with $\hat{V}_{Rp}(e_2)$ we note from this study that possibly a more efficient strategy would be to use an extension (with randomization) of RHC strategy rather than the IPPS scheme with extended (with randomization) Horvitz/Thompson (1952) estimation procedure.

N.B. Further details about the computations needed for presentation of the findings of the survey summarised in Tables I-IV are being incorporated in a Technical Report of the Indian Statistical Institute, to be available on a request to the author of this thesis.

We used the following one-page questionnaire for the survey - on one side of it the undernoted introductory points (to be explained to an interviewee) were mentioned and the actual questions were listed on the reverse as given below under the caption "Questionnaire".

INDIAN STATISTICAL INSTITUTE

A randomized response survey on earnings other than salary and expenses on confidential items for some specified ISI workers in 1982.

Dear Colleague,

Many of us occasionally earn money in excess over our usual salaries and also spend on certain items of dubious ethical values the extent and nature of which we do not like to disclose to anyone else. So, instead of enquiring about them directly from you I seek your co-operation through a randomized enquiry (the details to be verbally explained to you). As a part of my research programme I am interested to make a brief study about this with reference to some of you working now in ISI. In the short questionnaire below in Item No.1 we ask you about your 'Additional adjusted Income' meaning your income besides your normal salary minus a constant (to be specified by me) and in Item No.2 (called 'private adjusted expenditure', adjusted by subtracting a constant to be given by me) we ask you about your total expenditures on items like consumption of alcoholic drinks, gambling and some such others usually considered unethical in Indian Society, taken all together. To these two items please answer following the randomized device and not directly so that your privacy is maintained. To ensure this, against these items after a slash the phrase 'a related figure' (to be explained by me) is used. On other items I believe you have no objection to give the true information. Reference period is throughout the period 'since January 1, 1982'. For the sake of confidentiality you are advised not to record anywhere your identity.

Questionnaire

Item No.

1. Your additional adjusted income/a related figure
2. Your private adjusted expenditure/a related figure
3. Your expenses (if any) on each of the following items:
 - (a) purchase of TV set
 - (b) purchase of Fridge
 - (c) purchase of books
 - (d) purchase of cassettes/records etc.
 - (e) purchase of record players
 - (f) purchase of gas cylinder
 - (g) purchase of furniture
 - (h) purchase of crockery
 - (i) outing
 - (j) construction and repair of dwelling places,
purchase of land, buildings etc.
 - (k) any other (luxury item)
4. Total voluntary payments made for promotion of scientific, educational, cultural, religious and artistic activities in private and public life and for general social welfare

N.B. This enquiry is being made to enable me to report 'a case study' in my Ph.D. thesis under preparation illustrating an application of certain theoretical results already developed.

28 June 1982.

Sd/- Arun Kumar Adhikary
Computer Science Unit
Roll No.7902

- Chakrabarti, R.P. (1973). A note on the small sample theory of the ratio estimator in certain specified populations. J. Ind. Soc. Agri. Statist. 25, 49-57.
- Chaudhuri, A. (1978). On estimating the variance of a finite population. Metrika, 25, 65-76.
- Chaudhuri, A. (1979). A comparison among sampling strategies under a model. J. Ind. Soc. Agri. Statist. 31, 16-23.
- Chaudhuri, A. (1981). Non-negative unbiased variance estimators. In Current Topics in Survey Sampling, eds. Krewski, Platek and J.N.K. Rao, 317-328, Academic Press, New York.
- Chaudhuri, A. and Adhikary, A.K. (1982). On the relative efficiencies of several strategies for estimating a finite population mean. Accepted for publication in J. Ind. Soc. Agri. Statist. (to appear in 1982).
- Chaudhuri, A. and Arnab, R. (1979). On the relative efficiencies of sampling strategies under a super-population model. Sankhyā, Ser.C., 41, 40-43.
- Cochran, W.G. (1963). Sampling Techniques. Wiley and Sons. New York.

- Durbin, J. (1959). A note on the application of Quenouille's method of bias reduction to the estimation of ratios. *Biometrika*, 46, 477-480.
- Ericson, W.A. (1974). A note on the non-existence of minimum variance unbiased estimators for a finite population total. *Sankhyā, Ser.C.* 36, 181-185.
- Eriksson, Sven A. (1973). A new model for randomized response. *International Statistical Review*, 41, 101-113.
- Godambe, V.P. (1955). A unified theory of sampling from finite populations. *J. Roy. Statist. Soc. Ser. B.*, 17, 269 - 278.
- Godambe, V.P. (1960). An admissible estimate for any sampling design. *Sankhyā*, 22, 285-288.
- Godambe, V.P. (1969). Admissibility and Bayes estimation in sampling finite populations- V. *Ann. Math. Statist.* 40, 672-676.
- Godambe, V.P. (1980). Estimation in randomised response trials. *International Statistical Review*, 48, 29 - 32.

- Godambe, V.P. and Joshi, V.M. (1965). Admissibility and Bayes estimation in sampling from finite populations - I. Ann. Math. Statist. 36, 1707-1722.
- Godambe, V.P. and Thompson, M.E. (1973). Estimation in sampling theory with exchangeable prior distributions. Ann. Statist. 1, 1212-1221.
- Godambe, V.P. and Thompson, M.E. (1977). Robust near optimal estimation in survey practice. Bull. Int. Statist. Inst. 47, 129-146.
- Goodman, R. and Kish, L. (1950). Controlled selection - a technique in probability sampling. J. Amer. Statist. Assoc. 45, 350-372.
- Hansen, M.H. and Hurwitz, W.N. (1943). On the theory of sampling from finite populations. Ann. Math. Statist. 14, 333-362.
- Hanurav, T.V. (1966). Some aspects of unified sampling theory. Sankhyā, Ser. A., 28, 175-204.
- Hanurav, T.V. (1967). Optimum utilizations of auxiliary informations: π ps sampling of two units from a stratum. J. Roy. Statist. Soc. Ser.B., 29, 379-391.

- Hartley, H.O. and Rao, J.N.K. (1962). Sampling with unequal probabilities and without replacement. *Ann. Math. Statist.* 33, 350-374.
- Hege, V. (1965). Sampling designs which admit uniformly minimum variance unbiased estimate. *Cal. Statist. Assoc. Bull.* 14, 160-162.
- Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from finite universes. *J. Amer. Statist. Assoc.* 47, 663-685.
- Kemphorne, O. (1969). Some remarks on statistical inference in finite sampling. In N.L. Johnson and H. Smith edition, *New Developments in Survey Sampling*. Wiley Inter-Science. New York, 671-695.
- Lahiri, D.B. (1951). A method of sample selection providing unbiased ratio estimates. *Bull. Int. Statist. Inst.* 33, 133-140.
- Lanke, J. (1975). Some contributions to the theory of survey sampling. Ph.D. Thesis, University of Lund, Sweden.
- Midzuno, H. (1952). On the sampling system with probability proportional to sum of sizes. *Ann. Inst. Statist. Math.* 3, 99-107.

- Mohanty, S. and Das, M.N. (1971). Use of transformation in sampling. J. Ind. Soc. Agri. Statist. 23, 83-87.
- Murthy, M.N. (1957). Ordered and unordered estimators in sampling without replacement. Sankhyā, 18, 379-390.
- Pascual, Jose Nieto De (1961). Unbiased ratio estimators in stratified sampling. J. Amer. Statist. Assoc. 56, 70-87.
- Raj, D. (1954). Ratio estimation in sampling with equal and unequal probabilities. J. Ind. Soc. Agri. Statist. 6, 127-138.
- Raj, D. (1956). Some estimators in sampling with varying probabilities without replacement. J. Amer. Statist. Assoc. 51, 269-284.
- Ramakrishnan, M.K. (1975). Choice of an optimum sampling strategy - I. Ann. Statist. 3, 669-679.
- Rao, C.R. (1952). Some theorems on minimum variance unbiased estimation. Sankhyā, Ser. A, 12, 27-42.
- Rao, C.R. (1971). Some aspects of statistical inference in problems of sampling for finite populations. In V.P. Godambe and D.A. Sprott edition, Foundations of Statistical Inference. Holt, Rinehart and Winston, Toronto, 177-202.

- Rao, J.N.K. (1966). On the relative efficiency of some estimators in pps sampling for multiple characteristics. *Sankhyā*, Ser.A, 28, 61-70.
- Rao, J.N.K. (1969a). Ratio and Regression estimators. In *New Developments in Survey Sampling*, eds. N.L. Johnson and H. Smith, 213-234, New York, Wiley.
- Rao, J.N.K. (1979a). On deriving mean square errors and their non-negative unbiased estimators in finite population sampling. *J. Ind. Statist. Assoc.* 17, 125-136.
- Rao, J.N.K. and Bellhouse, D.R. (1978). Optimal estimation of a finite population mean under generalized random permutation models. *J.Statist.Plan. and Inf.* 2, 125-144.
- Rao, J.N.K., Hartley, H.O. and Cochran, W.G. (1962). On a simple procedure of unequal probability sampling without replacement. *J.Roy. Statist. Soc., Ser. B.* 24, 482-491.
- Rao, J.N.K. and Ramachandran, V. (1974). Comparison of the separate and combined ratio estimators. *Sankhyā*, Ser. C., 36, 151-156.
- Rao, J.N.K and Webster, J.T. (1966). On two methods of bias reduction in the estimation of ratios. *Biometrika*, 53, 571-577.

- Rao, P.S.R.S. (1968). On three procedures of sampling from finite populations. *Biometrika*, 55, 438-447.
- Rao, P.S.R.S. (1969b). Comparison of four ratio-type estimates under a model. *J.Amer.Statist.Assoc.* 64, 574-580.
- Rao, P.S.R.S. (1974). Jack-Knifing the ratio estimator. *Sankhyā, Series C*, 36, 84-97.
- Rao, P.S.R.S. (1979b). On applying the Jack-Knife procedure to the ratio estimator. *Sankhyā, Series C*, 41, 115-126.
- Rao, P.S.R.S. and Rao, J.N.K. (1971). Small sample results for ratio estimators. *Biometrika*, 58, 625-630.
- Rao, T.J. (1967). On the choice of a strategy for the ratio method of estimation. *J. Roy. Statist. Soc., Ser.B*, 29, 392-397.
- Roy, J. and Chakravarti, I.M. (1960). Estimating the mean of a finite population. *Ann.Math.Statist.* 31, 392-398.
- Sampford, M.R. (1967). On sampling without replacement with unequal probabilities of selection. *Biometrika*, 54, 499-513.
- Sekkappan, R.M. (1973). Bayes estimation and uniform admissibility for sampling from finite populations. Ph.D. Thesis, University of Waterloo, Ontario.

- Sekkappan, R.M. and Thompson, M.E. (1975). On a class of uniformly admissible estimators for finite populations. Ann. Statist. 3, 492-499.
- Sen, A.R. (1953). On the estimate of the variance in sampling with varying probabilities. J. Ind. Soc. Agri. Statist. 5, 119-127.
- Sengupta, S. (1981). A sampling scheme to realize inclusion probabilities exactly proportional to size. J. Ind. Soc. Agri. Statist. 33, 19-24.
- Singh, P. and Srivastava, A.K. (1980). Sampling schemes providing unbiased regression estimators. Biometrika. 67, 205-209.
- Srivenkataramana, T. (1978). Change of origin and scale in ratio and difference methods of estimation in sampling. Canadian. J. Statist. 6, 79-86.
- Warner, S.L. (1965). Randomized Response : A survey technique for eliminating evasive answer bias. J. Amer. Statist. Assoc. 60, 63-69.
- Yates, F. and Grundy, P.M. (1953). Selection without replacement from within strata with probability proportional to size. J. Roy. Statist. Soc., Ser.B, 15, 253-261.

A SUPPLEMENTARY CHAPTER

Motivation

This chapter accommodates discussions intended to reduce certain limitations of some of our results presented in the earlier chapters as pointed out by one of the examiners of the original version of this revised thesis.

Acknowledgement

We are thankful to the examiner on whose helpful suggestions we made further studies to obtain additional results included in this chapter in order to improve upon our original draft.

A SUPPLEMENTARY CHAPTER

1. Optimality in natural class of estimators

Let from a finite population U of size N labelled $1, \dots, i, \dots, N$ a first phase sample s' be chosen with probability $p_1(s')$ and from a selected s' a second phase sub-sample s be chosen with conditional probability $p_2(s|s')$ such that the over-all two-phase (double) sample $\bar{s}$ has the selection probability $p(\bar{s}) = p_1(s')p_2(s|s')$. Let D denote such a general class of two-phase sampling designs p . Let $\underline{Y} = (y_1, \dots, y_j, \dots, y_N)$, $\underline{X} = (x_1, \dots, x_j, \dots, x_N)$, $\underline{W} = (w_1, \dots, w_j, \dots, w_N)$ be vectors of real numbers with $\underline{Y}$, $\underline{X}$ unknown with totals (means) $Y(\bar{Y})$, $X(\bar{X})$ and $\underline{W}$ known with $w_j > 0 \ \forall j$ and total (mean) $W(\bar{W})$. When s' is chosen the values x_j 's for $j \in s'$ are ascertained giving a vector x' , say, of values x_j for $j \in s'$ and when s is chosen the values y_j 's for $j \in s$ are ascertained giving a vector y of values y_j 's for $j \in s$. Thus typically, $d = (\bar{s}, y, x')$ denotes observable survey data. As an estimator for Y one may consider an estimator $e = e(\bar{s}, y, x')$. By C we mean a class of design-unbiased estimators (UE)'s, e , say, such that $E_p(e) = Y \ \forall \underline{Y}$. By $E_p, V_p(\text{Cov}_p)$ we will denote design-expectation and design-variance (co-variance) operators. We have

Theorem 1. There does not exist a UMV (uniformly minimum variance) estimator in C based on any design in D unless one has a census.

Proof. Let $\underline{A}_K = (A_{K1}, \dots, A_{Ki}, \dots, A_{KN})$, $K = 1, 2$, be two fixed known vectors of real numbers with totals A_K ($K = 1, 2$) and let a_1 denote a sequence of A_{1i} values when $i \in s$ and a_2 denote a sequence of A_{2i} values when $i \in s'$. If an estimator $\bar{e} = \bar{e}(\bar{s}, y, x')$ be the UMV estimator for Y within C based on a design p in D , then the estimator $e^* = e^*(\bar{s}, y, x') = \bar{e}(\bar{s}, y, x') - \bar{e}(\bar{s}, a_1, a_2) + A_1$ is also in C and $V_p(e^*) = 0$ when $\underline{Y}$ equals $\underline{A}_1$ and $\underline{X}$ equals $\underline{A}_2$. Hence the theorem 1 follows.

Following Rao/Bellhouse (1978) let us consider the class C_1 (within C) of non-homogeneous linear UE's for Y of the form

$$e = e_{\bar{s}} = b_{\bar{s}} + \sum_{j \in s} b_{\bar{s}j} y_j + \sum_{j \in s} b'_{\bar{s}j} x_j + \sum_{j \in s' - s} b''_{\bar{s}j} x_j$$

with $b_{\bar{s}}, b_{\bar{s}j}, b'_{\bar{s}j}, b''_{\bar{s}j}$'s as quantities free of $\underline{Y}, \underline{X}$ subject to the conditions

$$E_p(b_{\bar{s}}) = 0, \quad E_p\left(\sum_{j \in s} b_{\bar{s}j} y_j\right) = Y \quad \forall \underline{Y},$$

$$E_p\left(\sum_{j \in s} b'_{\bar{s}j} x_j + \sum_{j \in s' - s} b''_{\bar{s}j} x_j\right) = 0 \quad \forall \underline{X}.$$

Let us write $P_j = \sum_{s' \ni j} p_1(s') = E_1 p_1(s')$, say,

$$Q_j = \sum_{s' \ni j} \sum_{s \ni j} p_2(s|s') p_1(s') = E_1 E_2 p(\bar{s}), \text{ say,}$$

$$P_{jk} = \sum_{s' \ni j, k} p_1(s') = E_3 p_1(s'), \text{ say,}$$

$$Q_{jk} = \sum_{s' \ni j, k} \sum_{s \ni j, k} p(\bar{s}) = E_3 E_4 p(\bar{s}), \text{ say, - (these are}$$

the inclusion-probabilities which will be assumed positive whenever required for their use without repeated mentioning);

$$R_{jk} = \sum_{s' \ni j, k} \sum_{s \ni j} p(\bar{s}) = \Sigma_1 \Sigma_2 p(\bar{s}) \text{ and}$$

$$R_{kj} = \sum_{s' \ni j, k} \sum_{s \ni k} p(\bar{s}) = \Sigma_2 \Sigma_7 p(\bar{s}), \text{ say. Let also } \Sigma_5 \equiv \text{sum over } (s'-s) \ni j \text{ and } \Sigma_6 \equiv \text{sum over } s'-s \ni j, k \text{ as these notations will be used later.}$$

By D_1 we will denote the class of uni-cluster designs within D having the property that every two first phase samples with non-zero selection probabilities are either disjoint or equivalent (i.e. having the same set of distinct units) and also every two second phase samples from any first phase sample are also so. Other designs in D constitute the class D_2 of non-unicluster designs. We have then

Theorem 2. (i) If a UMV estimator for Y in C_1 based on any design in D_1 is to exist then it has to be of the form

$$\bar{c} = \bar{c}_0 = \sum_{j \in s} \frac{y_j}{Q_j} + \sum_{j \in s'} b'_j x_j + \sum_{j \in s'-s} b''_j x_j$$

where b'_j, b''_j 's are constants independent of $\bar{s}$ but subject to the conditions

$$b'_j/b''_j = 1 - \frac{P_j}{Q_j} \quad \forall j.$$

(ii) There does not exist a UMV estimator (for Y) in C_1 based on any design in D_2 .

Proof. (i) Necessary and sufficient conditions (NSC) for $e_{\bar{s}}$ to belong to C_1 are that

$$\sum_1 \sum_2 b_{\bar{s}j} p(\bar{s}) = 1 \quad \forall j \quad \text{and} \quad \sum_1 \sum_2 b'_{\bar{s}j} p(\bar{s}) + \sum_1 \sum_5 b''_{\bar{s}j} p(\bar{s}) = 0 \quad \forall j.$$

Also an $e_{\bar{s}}$ must be free of 'order and multiplicity' with which units occur in $\bar{s}$ in order that it may be a UMV estimator (UMVE, say). If it is based on a design in D_1 it follows that $b_{\bar{s}j}$, $b'_{\bar{s}j}$, $b''_{\bar{s}j}$ must each be free of $\bar{s}$ but may depend only on j . Combining this observation with the unbiasedness conditions yields the required proof.

$$(ii) \quad \text{If } \bar{e} = \bar{b}_{\bar{s}} + \sum_{j \in s} \bar{b}_{\bar{s}j} y_j + \sum_{j \in s} \bar{b}'_{\bar{s}j} x_j + \sum_{j \in s'-s} \bar{b}''_{\bar{s}j} x_j$$

is to be the UMVE for Y in C_1 , then by C.R. Rao's (1952) theorem [vide Ericson (1974) and Rao/Bellhouse (1978)] one must have

$$\text{Cov}_p(\bar{e}_{\bar{s}}, e_{\bar{a}s}) = 0 \quad \forall e_{\bar{s}} \text{ in } C_1.$$

$$\begin{aligned} \text{Here } e_{\bar{a}s} &= \bar{e}_{\bar{s}} - e_{\bar{b}s} = (\bar{b}_{\bar{s}} - b_{\bar{s}}) + \sum_{j \in s} (\bar{b}_{\bar{s}j} - b_{\bar{s}j}) y_j \\ &\quad + \sum_{j \in s} (\bar{b}'_{\bar{s}j} - b'_{\bar{s}j}) x_j + \sum_{j \in s'-s} (\bar{b}''_{\bar{s}j} - b''_{\bar{s}j}) x_j \\ &= a_{\bar{s}} + \sum_{j \in s} a_{\bar{s}j} y_j + \sum_{j \in s} a'_{\bar{s}j} x_j + \sum_{j \in s'-s} a''_{\bar{s}j} x_j, \text{ say, with} \end{aligned}$$

obvious notations. Working out (but suppressing) the algebraic details, it may be checked that the following conditions (1.1) - (1.9) together constitute NSC's for $\text{Cov}_p(\bar{e}_{\bar{s}}, e_{\bar{a}s}) = 0$, namely,

$$\sum_{\bar{s}} a_{\bar{s}} \bar{b}_{\bar{s}} p(\bar{s}) = 0 \quad (1.1)$$

$$\Sigma_1 \Sigma_2 (\bar{b}_{\bar{s}} a_{\bar{s}j} + a_{\bar{s}} \bar{b}_{\bar{s}j}) p(\bar{s}) = 0 \quad \forall j \quad (1.2)$$

$$\Sigma_1 \left[\bar{b}_{\bar{s}} a'_{\bar{s}j} + a_{\bar{s}} \bar{b}'_{\bar{s}j} \right] + \Sigma_5 (\bar{b}_{\bar{s}} a''_{\bar{s}j} + a_{\bar{s}} \bar{b}''_{\bar{s}j}) p(\bar{s}) = 0 \quad \forall j \quad (1.3)$$

$$\Sigma_1 \Sigma_2 \bar{b}_{\bar{s}j} a_{\bar{s}j} p(\bar{s}) = 0 \quad \forall j \quad (1.4)$$

$$\Sigma_3 \Sigma_4 \bar{b}_{\bar{s}j} a_{\bar{s}k} p(\bar{s}) = 0 \quad \forall j \neq k \quad (1.5)$$

$$\Sigma_1 \left[\Sigma_2 \bar{b}'_{\bar{s}j} a'_{\bar{s}j} + \Sigma_5 \bar{b}''_{\bar{s}j} a''_{\bar{s}j} \right] p(\bar{s}) = 0 \quad \forall j \quad (1.6)$$

$$\begin{aligned} \Sigma_3 \left[(\Sigma_4 \bar{b}'_{\bar{s}j} a'_{\bar{s}k} + \Sigma_6 \bar{b}''_{\bar{s}j} a''_{\bar{s}k}) + \frac{1}{2} \Sigma_2 (\bar{b}'_{\bar{s}j} a''_{\bar{s}k} + a'_{\bar{s}j} \bar{b}''_{\bar{s}k}) \right. \\ \left. - \Sigma_4 (\bar{b}'_{\bar{s}j} a''_{\bar{s}k} + a'_{\bar{s}j} \bar{b}''_{\bar{s}k}) + \Sigma_7 (\bar{b}'_{\bar{s}k} a''_{\bar{s}j} + a'_{\bar{s}k} \bar{b}''_{\bar{s}j}) \right. \\ \left. - \Sigma_4 (\bar{b}'_{\bar{s}k} a''_{\bar{s}j} + a'_{\bar{s}k} \bar{b}''_{\bar{s}j}) \right] p(\bar{s}) = 0 \quad \forall j \neq k \quad (1.7) \end{aligned}$$

$$\Sigma_1 \Sigma_2 (a_{\bar{s}j} \bar{b}'_{\bar{s}j} + a'_{\bar{s}j} \bar{b}_{\bar{s}j}) p(\bar{s}) = 0 \quad \forall j \quad (1.8)$$

$$\begin{aligned} \Sigma_3 \left[\Sigma_4 (\bar{b}_{\bar{s}j} a'_{\bar{s}k} + \bar{b}'_{\bar{s}k} a_{\bar{s}j}) - (\bar{b}_{\bar{s}j} a''_{\bar{s}k} + a_{\bar{s}j} \bar{b}''_{\bar{s}k}) \right. \\ \left. + \Sigma_2 (\bar{b}_{\bar{s}j} a''_{\bar{s}k} + a_{\bar{s}j} \bar{b}''_{\bar{s}k}) \right] p(\bar{s}) = 0 \quad \forall j \neq k \quad (1.9) \end{aligned}$$

In addition, we need (because of unbiasedness of $e_{\bar{b}\bar{s}}$ and $\bar{e}_{\bar{b}\bar{s}}$) of course

$$\sum_{\bar{s}} a_{\bar{s}} p(\bar{s}) = 0 \quad (1.10)$$

$$\sum_1 \sum_2 a_{\bar{s}_j} p(\bar{s}) = 0 \quad \forall j \quad (1.11)$$

$$\text{and} \quad \sum_1 (\sum_2 a'_{\bar{s}_j} + \sum_5 a''_{\bar{s}_j}) = 0 \quad \forall j. \quad (1.12)$$

Let $\bar{s}_1$ be two distinct two-phase samples with $p(\bar{s}_1) > 0$, $i = 1, 2$ such that (i) for the supposedly UMVE $\bar{e}_{b\bar{s}}, \bar{b}_{\bar{s}_1} \neq \bar{b}_{\bar{s}_2} \neq 0$ and (ii) for another $e = e_{b\bar{s}}$ in C_1 , $\bar{b}_{\bar{s}} = b_{\bar{s}} \quad \forall \bar{s}$ other than $\bar{s}_1, \bar{s}_2$.

Then it follows that $0 = E_p(\bar{e}_b - e_b) = a_{\bar{s}_1} p(\bar{s}_1) + a_{\bar{s}_2} p(\bar{s}_2)$.

This contradicts (1.1) because $\sum_{\bar{s}} a_{\bar{s}} b_{\bar{s}} p(\bar{s})$ now equals

$(\bar{b}_{\bar{s}_1} - \bar{b}_{\bar{s}_2}) a_{\bar{s}_1} p(\bar{s}_1) \neq 0$. Similarly one may check contradictions in the other relations (1.2) - (1.9), following essentially similar steps as in our original dissertation [cf. pp. 26-32] and thus note the completion of the proof.

Next we will consider properties of the class of estimators $e_{b\bar{s}}^* = \sum_{j \in S} y_j / Q_j + \sum_{j \in S} b'_j x_j + \sum_{j \in S, -S} b''_j x_j$ (subject to $b'_j / b''_j = 1 - \frac{P_j}{Q_j} \quad \forall j$) based on any p in D , with $P_j, Q_j > 0 \quad \forall j$.

By C_2 we will denote the sub-class of C_1 consisting of estimators $e_{b\bar{s}}$ for which $b_{\bar{s}} = 0 \quad \forall \bar{s}$ with $p(\bar{s}) > 0$. For any such estimator $e_{b\bar{s}}^0$, say, we have

$$\begin{aligned}
 V_p(e_{b\bar{s}}^0) = & \sum_{j=1}^N y_j^2 (\sum_1 \sum_2 b_{\bar{s}j}^2 p(s) - 1) + \sum_{j \neq k=1}^N y_j y_k (\sum_3 \sum_4 b_{\bar{s}j} b_{\bar{s}k} p(\bar{s}) - 1) \\
 & + \sum_{j=1}^N x_j^2 \left[\sum_2 b_{\bar{s}j}'^2 + \sum_5 b_{\bar{s}j}''^2 \right] p(\bar{s}) \\
 & + \sum_{j \neq k=1}^N x_j x_k \sum_5 \left[\sum_4 b_{\bar{s}j}' b_{\bar{s}k}' + \sum_6 b_{\bar{s}j}'' b_{\bar{s}k}'' + \sum_2 b_{\bar{s}j}' b_{\bar{s}k}'' \right. \\
 & \left. + \sum_7 b_{\bar{s}k}' b_{\bar{s}j}'' - \sum_4 (b_{\bar{s}j}' b_{\bar{s}k}'' + b_{\bar{s}k}' b_{\bar{s}j}'') \right] p(\bar{s}) \\
 & + 2 \sum_{j=1}^N y_j x_j (\sum_1 \sum_2 b_{\bar{s}j} b_{\bar{s}j}' p(\bar{s})) \\
 & + 2 \sum_{j \neq k=1}^N y_j x_k \sum_3 \left[\sum_4 b_{\bar{s}j} (b_{\bar{s}k}' - b_{\bar{s}k}'') + \sum_2 b_{\bar{s}j} b_{\bar{s}k}'' \right] p(\bar{s}).
 \end{aligned}$$

Using Cauchy-Schwarz's inequality it is easy to check that at a point $(\underline{Y}, \underline{X})$ with $y_j \neq 0$, $y_k = 0 \ \forall \ k \neq j$, $x_j = 0 \ \forall \ j$,

$$V_p(e_{b\bar{s}}^0) - V_p(e_{b\bar{s}}^*) = y_j^2 (\sum_1 \sum_2 b_{\bar{s}j}^2 p(\bar{s}) - 1/q_j) \geq 0$$

and at a point $(\underline{Y}, \underline{X})$ with $x_j \neq 0$, $x_k = 0 \ \forall \ k \neq j$, $y_j = 0 \ \forall \ j$,

$$\begin{aligned}
 V_p(e_{b\bar{s}}^0) - V_p(e_{b\bar{s}}^*) = & x_j^2 \left[\sum_1 (\sum_2 b_{\bar{s}j}'^2 + \sum_5 b_{\bar{s}j}''^2) p(\bar{s}) \right. \\
 & \left. - (b_{\bar{s}j}'^2 q_j + b_{\bar{s}j}''^2 (p_j - q_j)) \right] \geq 0.
 \end{aligned}$$

From this one may complete the proof of

Theorem 3. Among estimators in C_2 based on any design in D , every estimator $e_{b\bar{s}}^*$ as above is 'admissible'.

2. Optimality of double sampling regression estimator under super-population models

Let us consider the super-population model M (say), such that writing E_m, V_m, C_m as operators for model-based expectations, variances and co-variances generically, we have

$$E_m(y_j) = \theta_j, \quad V_m(y_j) = \sigma_j^2, \quad C_m(y_j, y_k) = \rho \sigma_j \sigma_k,$$

$$E_m(x_j) = \theta'_j, \quad V_m(x_j) = \sigma'^2_j, \quad C_m(x_j, x_k) = \rho' \sigma'_j \sigma'_k,$$

$$C_m(y_j, x_j) = \xi \sigma_j^{\prime\prime 2}, \quad C_m(y_j, x_k) = \rho'' \xi \sigma_j^{\prime\prime} \sigma_k^{\prime\prime}$$

such that $-\infty < \theta_j, \theta'_j < \infty, 0 < \sigma_j, \sigma'_j, \sigma_j^{\prime\prime} < \infty, 0 \leq \rho, \rho', \rho'' \leq 1$. Also $\xi = \pm 1$ and we will assume that in any specific situation under investigation ξ is known.

We will consider a sub-class C_3 of estimators in C_1 based on any design p in D such that $\xi b'_{sj} \geq 0, \xi b''_{sj} \leq 0$, and

$$\text{Cov}_p\left(\sum_{j \in s} b_{sj} y_j, \sum_{j \in s} b'_{sj} x_j + \sum_{j \in s'-s} b''_{sj} x_j\right) \geq 0. \quad (2.1)$$

Every design in the sub-class $\bar{D}$ of D for which every s' with $p_1(s') > 0$ has a fixed size n' and every s with $p_2(s|s') > 0$ has a fixed size $n (< n')$ will be denoted as $\bar{p}$. If in particular $P_j = n'W_j/W$ and $Q_j = nW_j/W \quad \forall j$ then $\bar{p}$ will be called an IPPS (inclusion probabilities proportional to sizes) design and denoted as $\bar{p}_w$, the corresponding class being denoted as $\bar{D}_w$. Then we get

Theorem 4. For every $e = e_{b\bar{s}}$ in C_3 ,

$$(i) \quad E_{m\bar{p}} V(e_{b\bar{s}}) \geq (1-\rho) \sum_{j=1}^N \sigma_j^2 \left(\frac{1}{Q_j} - 1 \right) + \lambda^2 (1-\rho') \sum_{j=1}^N \sigma_j'^2 \left(\frac{1}{Q_j} - \frac{1}{P_j} \right) \\ + 2\lambda(1-\rho'') \sum_{j=1}^N \sigma_j''^2 \left(\frac{1}{Q_j} - \frac{1}{P_j} \right), \quad \text{for any}$$

p in D and any $\lambda > 0$ such that $\{b_j''\} = -\lambda/P_j$ with the magnitudes of λ, b_j'' unspecified and (ii) in case $\sigma_j = \sigma W_j$, $\sigma_j' = \sigma' W_j$, $\sigma_j'' = \sigma'' W_j \quad \forall j$, ($0 < \sigma, \sigma', \sigma'' < \infty$),

$$E_{m\bar{p}} V(e_{b\bar{s}}) \geq (1-\rho) \sigma^2 \left(\frac{W^2}{n} - \sum_{j=1}^N W_j^2 \right) + \sigma^2 (1-\rho') \sigma'^2 W^2 \left(\frac{1}{n} - \frac{1}{n'} \right) \\ - 2\sigma(1-\rho'') \sigma''^2 W^2 \left(\frac{1}{n} - \frac{1}{n'} \right)$$

$$= E_{m\bar{p}_w} V(\bar{e}_{b\bar{s}}(\theta)), \quad \text{for every } \bar{p} \text{ in } \bar{D} \text{ and } \bar{p}_w$$

in $\bar{D}_w$, where $\theta = \frac{(1-\rho'')\sigma''^2}{(1-\rho')\sigma'^2}$

$$= \frac{C_m \left(\frac{y_j}{w_j}, \frac{x_j}{w_j} \right) - C_m \left(\frac{y_j}{w_j}, \frac{x_k}{w_k} \right)}{V_m \left(\frac{x_j}{w_j} \right) - C_m \left(\frac{x_j}{w_j}, \frac{x_k}{w_k} \right)} \quad (2.2)$$

$$\text{and } \bar{e}_{b\bar{s}}(\theta) = \sum_{j \in s} \frac{y_j}{Q_j} + \theta \left(\sum_{j \in s'} \frac{x_j}{P_j} - \sum_{j \in s} \frac{x_j}{Q_j} \right) = e(\theta), \quad \text{say.}$$

[cf. Rao/Bellhouse (1978)].

Proof. Follows on applying Cauchy-Schwarz's inequality in appropriate manners and using definitions and unbiasedness conditions. Straightforward details are omitted.

Remark 1. The sign of $C_m(y_j, x_j) \forall j$ and that of $C_m(y_j, x_k) \forall j \neq k$ is supposed to be common and this is to decide the sign of $e(\sum_{j \in s} \frac{x_j}{F_j} - \sum_{j \in s} \frac{x_j}{Q_j})$ in the estimator $e(\theta)$. The strategy $(\bar{p}_w, \bar{e}_{bs}(\theta))$ with any $\bar{p}_w$ in $\bar{D}_w$ is optimal among all $(\bar{p}, \bar{e}_{bs})$'s with $\bar{e}_{bs}$ in C_3 based on any $\bar{p}$ design in $\bar{D}$.

Remark 2. In case $\rho'' = 0$, the condition (2.1) is not needed and the Theorem 4 follows with obvious modifications.

Remark 3. The value of θ depends upon model parameters and it is not obvious what estimate for it based on a sample should in practice be used, [vide (2.2)] but intuitively sample regression coefficient should be used to get an approximation for the regression estimator $\bar{e}_{bs}(\theta)$, assuming large samples.

3. An asymptotic approach

If design-unbiasedness condition on estimators be relaxed then with the following modified model one may think of employing an estimator for Y following Royall's (1970) prediction approach as discussed below in brief.

Let $E_m(y_j|x_j) = \gamma x_j$, $V_m(y_j|x_j) = a^2 f(x_j) > 0$, and y_j 's conditionally (given x_j 's) independent; $E_m(x_j|w_j) = \gamma' w_j$, $V_m(y_j|w_j) = b^2 g(w_j) > 0$ and x_j 's, y_j 's be conditionally (given w_j 's) independent. Then, $E_m(y_j|w_j) = \gamma \gamma' w_j = \psi w_j$ (say). Now given a two-phase sample $\bar{s}$ and survey-data $d = (\bar{s}, y, x')$,

one may write $Y = \sum_{j \in S} y_j + \sum_{j \in S' - S} y_j + \sum_{j \in U - S} y_j$

so that $E_m(Y|d) = \sum_{j \in S} y_j + \gamma \sum_{j \in S' - S} x_j + \psi \sum_{j \in U - S} w_j$.

Then, following Royall (1970) one may consider using a model-unbiased predictor for Y as

$$t = t(d) = \sum_{j \in S} y_j + \hat{\gamma} \sum_{j \in S' - S} x_j + \hat{\psi} \sum_{j \in U - S} w_j$$

such that $\hat{\gamma}$ is a function of (y_j, x_j) 's for $j \in S$ and $\hat{\psi}$ is a function of y, x' and of $\underline{w}$ subject to the conditions $E_m(\hat{\gamma} - \gamma) = 0 \quad \forall s'$ with $p_1(s') > 0$ and $E_m(\hat{\psi} - \psi) = 0 \quad \forall \bar{s}$ with $p(\bar{s}) > 0$.

Considering the mean square error (MSE), namely,

$$\begin{aligned} M = E_m(t - Y)^2 &= E_m \left[\sum_{j \in S' - S} (y_j - \gamma x_j) - \sum_{j \in S' - S} x_j (\hat{\gamma} - \gamma) \right. \\ &\quad \left. + \sum_{j \in U - S} (y_j - \psi w_j) - \sum_{j \in U - S} w_j (\hat{\psi} - \psi) \right]^2 \\ &= a^2 \sum_{j \in S' - S} f(x_j) + \left(\sum_{j \in S' - S} x_j \right)^2 V_m(\hat{\gamma}) + b^2 \sum_{j \in U - S} g(w_j) \\ &\quad + \left(\sum_{j \in U - S} w_j \right)^2 V_m(\hat{\psi}) \end{aligned}$$

when appropriate linear regression models as above hold true BLUE's (best linear unbiased estimators) for γ, ψ may be sought to be derived. But following Särndal (1980), Robinson/Särndal (1983), Isaki/Fuller (1981, 1982) one may instead employ the design-based predictor

$$t_D = t_D(d) = \sum_{j \in S} \frac{y_j}{Q_j} + \hat{\gamma}_D \left(\sum_{j \in S} \frac{x_j}{P_j} - \sum_{j \in S} \frac{x_j}{Q_j} \right) + \hat{\psi}_D \left(W - \sum_{j \in S} \frac{w_j}{P_j} \right)$$

$$\text{taking } \hat{\gamma}_D = \frac{\sum_{j \in S} y_j / Q_j}{\sum_{j \in S} x_j / Q_j} \quad \text{and} \quad \hat{\psi}_D = \frac{\hat{\gamma}_D \sum_{j \in S} \frac{x_j}{P_j} - \sum_{j \in S} \frac{y_j}{Q_j}}{\sum_{j \in S} \frac{w_j}{P_j} - \sum_{j \in S} \frac{w_j}{Q_j}}.$$

Then, it follows that t_D is model-unbiased i.e. $E_{\pi}(t_D - Y) = 0$ $\forall \bar{s}$ with $p(\bar{s}) > 0$.

If based on a $\bar{p}_W$ design, t_D reduces to t_W , say,

$$\text{where } t_W = t_W(d) = w \left[\frac{1}{n} \sum_{j \in S} \frac{y_j}{w_j} + \frac{\sum_{j \in S} y_j / w_j}{\sum_{j \in S} x_j / w_j} \left(\frac{1}{n'} \sum_{j \in S} \frac{x_j}{w_j} - \frac{1}{n} \sum_{j \in S} \frac{x_j}{w_j} \right) \right].$$

But neither t_D nor t_W is design-unbiased.

Formally t_W corresponds to $e(\theta)$ when the latter is based on a $\bar{p}_W$ design provided θ is replaced by an estimator for it as equal to $\sum_{j \in S} \frac{y_j}{w_j} / \sum_{j \in S} \frac{x_j}{w_j}$. In general, for θ one may use an estimator $\hat{\theta} = \hat{\gamma}_D$ which will be called $\hat{\theta}_W$ when $\bar{p}_W$ is used. So, we will write t_W as $t_W(\hat{\theta})$. For any constant μ for θ irrespective of any underlying model the estimator $e(\theta) = e(\mu)$ is valid in design-unbiasedly estimating Y . But its optimality in respect of its efficiency follows only when the postulated model holds and θ is as given in (2.2). But t_W is model-unbiased only under the postulated model and is not design-unbiased. But following the approaches due to Brewer (1979) pursued with by Sarndal (1980) and Robinson/Sarndal (1983) or by Isaki/Fuller (1981, 1982)

one may proceed to establish 'Asymptotic design-unbiasedness' and 'consistency' of t_D and t_w in estimating Y . Following Brewer (1979) if we conceptualize that U is reproduced K times to yield amalgamated population totals KY and KX and on each reproduction it is independently sampled in two phases yielding an estimator t_D (or t_w) and following repeated two-phase sampling the same estimator is based on the amalgamated sample $\bar{s}_k$, say, and the estimator t_D (or t_w) is based on $\bar{s}_k$ to estimate KY then 'Asymptotic unbiasedness' for t_D demands

- (1) $\lim_{k \rightarrow \infty} E_p[(t_D(\bar{s}_k) - KY)/NK] = 0$, the probability being calculated with respect to p . Also 'consistency' of t_D demands
- (1i) $P_m[\lim_{k \rightarrow \infty} \text{Prob}\{|t_D(\bar{s}_k) - KY|/NK > \epsilon\} = 0] = 1$ for any fixed $\epsilon > 0$, no matter how small and P_m denotes the probability distribution (which may also be degenerate) of $\underline{y}$ under a postulated model. Following the works of Brewer (1979), Särndal (1980) and Robinson/Särndal (1983) it follows that t_D (and t_w as well) is both 'asymptotically unbiased' and 'consistent' for Y . We need not elaborate on this further because additional work is hardly needed and also because rigorous investigation of this aspect was not within the scope of the original thesis.

The study of asymptotic efficiency of t_D and derivation of an asymptotically optimal design for its use seem difficult and are not taken up here. But if in the model M we only assume $E_m(y_j|x_j) = Yx_j$ and $E_m(x_j|w_j) = Y'w_j$ and take $\phi = YY'$,

It follows that $e(\hat{\theta})$ is model-unbiased for Y if it is based on any $\bar{p}_w$ design. Since when based on $\bar{p}_w$, $e(\hat{\theta})$ coincides with $\bar{y}_w$, irrespective of any model, it is 'asymptotically design-unbiased' and 'consistent'. Without further work (which we are unable to undertake now) we cannot claim 'asymptotic optimal efficiency' of $e(\hat{\theta})$.

4. Alternative variance estimators

Let the first phase samples s' of a fixed size n' be chosen according to SRSWOR scheme and x_j 's be all positive. Let $q_j = x_j / \sum_{j \in s'} x_j$ and from a selected first phase sample a second phase sample s of a fixed size $n (< n')$ be chosen with a probability $p_2(s|s') = p_2(s)$ according to (i) any suitable IPPS (with conditional inclusion-probabilities $\pi_j = nq_j$ and $\pi_{jk} = E_4 p_2(s)$) scheme or (ii) RHC scheme or (iii) IMS scheme using in each case q_j 's ($j \in s'$) as the size-measures. Let S_j denote the sum of q_j 's over units falling in the j th group (each of a common size $\frac{n}{n'}$, supposed to be an integer) while applying the RHC scheme.

Based on these schemes (i) - (iii), the respective estimators, namely, $e_1 = \frac{N}{n} \cdot \frac{1}{n} \sum_{j \in s} y_j / q_j$, $e_2 = \frac{N}{n} \cdot \sum_{j \in s} \frac{y_j}{q_j} S_j$,

$e_3 = \frac{N}{n} \cdot \frac{\sum_{j \in s} y_j}{\sum_{j \in s} q_j}$ have the variances

$$V_p(e_1) = A + E_{p_1} V_{p_2}(e_1), \quad V_p(e_2) = A + E_{p_1} V_{p_2}(e_2) \quad \text{and}$$

$$V_p(e_3) = A + E_{p_1} V_{p_2}(e_3) \quad \text{where} \quad A = N^2 \left(\frac{1}{n} - \frac{1}{N} \right) \cdot \frac{1}{N-1} \sum_{j=1}^N (y_j - \bar{Y})^2.$$

Following Rao/Vijayan (1977) and using Rao's (1979) general method of unbiasedly estimating mean square errors of linear estimators for Y we obtain respective unbiased estimators for $V_p(e_i)$, $i=1,2,3$ as

$$v_1 = N^2 \left(\frac{1}{n} - \frac{1}{N} \right) \frac{1}{n'(n'-1)} \sum_{j < k \in s} \frac{(y_j - y_k)^2}{\pi_{jk}} \\ + \frac{N^2}{n'^2} \sum_{j < k \in s} \frac{\pi_j \pi_k - \pi_{jk}}{\pi_{jk}} \left(\frac{y_j}{\pi_j} - \frac{y_k}{\pi_k} \right)^2$$

$$v_2 = N^2 \left(\frac{1}{n} - \frac{1}{N} \right) \frac{n}{n'^2(n-1)} \sum_{j < k \in s} \frac{s_j s_k}{q_j q_k} (y_j - y_k)^2 \\ + \frac{N^2}{n'^2} \cdot \frac{n'-n}{n-1} \cdot \frac{1}{n'} \sum_{j < k \in s} s_j s_k \left(\frac{y_j}{q_j} - \frac{y_k}{q_k} \right)^2$$

$$\text{and } v_3 = N^2 \left(\frac{1}{n} - \frac{1}{N} \right) \cdot \frac{1}{n'(n-1)} \sum_{j < k \in s} \frac{(y_j - y_k)^2}{q(s)} \\ + \frac{N^2}{n'^2} \sum_{j < k \in s} \frac{1}{q(s)} \left(\frac{n'-1}{n-1} - \frac{1}{q(s)} \right) q_j q_k \left(\frac{y_j}{q_j} - \frac{y_k}{q_k} \right)^2$$

$$\text{where } q(s) = \sum_{j \in s} q_j.$$

Here v_2 and v_3 are uniformly non-negative but v_1 is so if

$$\pi_j \pi_k > \pi_{jk} \quad \forall j, k (j \neq k), \quad \pi_j q_j q(s) > \frac{n-1}{n'}, \quad \forall s.$$

5. Efficiency comparison

While comparing among the strategies with the estimators t_i ($i = 1, \dots, 6$) on pp.111-125 and t_i ($i = 1, 2, 4, \dots, 9$) on pp.129-149 under linear regression model the appropriate value of α is zero but non-zero α 's are allowed as we also compare with them the regression estimator which is usable for any α . Since we focus on the estimators other than the regression estimator, only small α 's are used while presenting numerical values. When we anticipate large values of α (though its magnitude is not easy to guess) our preference is for estimators with a shift in the origin vide pp.163-193. Of course in large samples the regression estimator is hard to beat. But theoretical comparison of the regression estimator with the other estimators (based on various sampling schemes we consider) when sample size is small is found difficult even under simplistic models we postulate. Incidentally we may consider [Chaudhuri/Adhikary (1984)] a numerical example from Cochran (1977, p.325) with $N=10$, $n=2$ with y_i, x_i values given so that from the observed scatter diagram one may guess an appropriate value of α as 32 for a linear regression model. The actual calculations come out as in the table 5.1 below. Here we directly calculate the exact values of $V'_i = E_p(t'_i - \bar{Y})^2$, $i = 1, \dots, 6$ where t'_i 's are the estimators given on p.167 and show the magnitudes of relative efficiencies $E'_i = 100 V'_2 / V'_i$, $i = 1, \dots, 6$ of the strategies for some values of θ , recalling that the appropriate values of θ should be within $(0, 2\alpha)$.

TABLE 5.1

Relative efficiencies (E'_i) of the sampling strategies for various values of θ within $(0, 2\alpha)$, $\alpha = 32$.

θ	E'_1	E'_2	E'_3	E'_4	E'_5	E'_6	
0	103.95	100	89.57	100.77	101.08	2.42	(116.06)
12	103.64	100	89.63	100.85	101.12	2.26	(108.48)
22	103.11	100	89.61	100.80	100.96	2.18	(104.93)
32	102.37	100	89.49	100.62	100.62	2.16	(103.90)
42	101.47	100	89.33	100.50	100.13	2.19	(105.39)
52	100.47	100	89.10	100.23	99.52	2.28	(101.38)

The figures within the parantheses use the asymptotic variance of the regression estimator, using Cochran's well-known formula.

Here we find that the regression estimator is quite unsatisfactory from efficiency point of view compared to the origin-shifted estimators. Of course, using asymptotic formula [vide Cochran (1977)] the situation is not as bad though reference to this is not quite fair with $N=10$ and $n=2$. From this however one can hardly reach any conclusion about what to do in practice since α is not really easy to guess. This is only an illustration that for very small samples the exact MSE of regression estimator may not come out favourable vis-a-vis its possible competitors we consider.

We find it difficult to compare the origin-shifted estimators with the regression estimator in general under model and also without model for moderate sample sizes ($n > 2$). Comparison of them with Särndal's (1980) regression type estimator seems still more difficult and this is not attempted here. We may note however that this as well as the regression estimator is 'model-unbiased' but 'not design-unbiased' while the former 'is asymptotically design-unbiased' and 'consistent'. The origin-shifted estimators are however 'design-unbiased' but not 'model unbiased'.

While comparing the unbiased regression estimator (and also the regression-type estimator) and the classical regression estimator [vide pp.154-159] the use of the asymptotic variance formula for the latter seems inappropriate so that in our original dissertation Result 1 on p.154, Result 2 on p.155 and Results 4 and 5 on p.158, consideration of III should be ruled out; so we have dropped them now. The result 3 on p.156 is of course relevant concerning the classical regression estimator and this we retain.

6. Optimality in Randomized Response Surveys

Let a suitable randomized response technique (RRT, say) be so devised that each individual i when sampled independently of each other makes a random response z_i when the true variate-value associated with him is y_i .

Let E_R, V_R, C_R denote operators for expectations, variance and co-variance with respect to randomization. We shall assume

throughout this section a model such that for an adopted RRT,

$$E_R(z_i) = a_i y_i + b_i$$

$$E_R(z_i^2) = a_i y_i^2 + d_i$$

where a_i is in $(0,1)$, b_i real, $d_i (\geq b_i^2)$ are known constants for $i = 1, \dots, N$. Various RRT's for which this model seems appropriate are well-known [vide Bellhouse (1980) for a review and also Eriksson (1973) for a particular case].

Writing $z'_i = \frac{z_i - b_i}{a_i}$, $\underline{z}' = (z'_1, \dots, z'_i, \dots, z'_N)$ we will consider estimators for Y based on $\underline{z}'$ of the class C_b of the form

$$e_b = e_b(s, \underline{z}') = b_s + \sum_{i \in S} b_{s_i} z'_i$$

such that $E_p(b_s) = 0$ and $\sum_{s \ni i} b_{s_i} p(s) = 1 \quad \forall i$.

Then $E_R(e_b) = b_s + \sum_{i \in S} b_{s_i} y_i = t_b(s, \underline{y}) = t_b$ (say)

[vide Eriksson (1973) and note that $E_R(z'_i) = y_i$].

Writing $E_{pR} = E_{Rp}$, $V_{Rp} = V_{pR}$ as operators for expectation and variance jointly with respect to both design and randomization it follows that

$$E_{pR}(e_b) = Y.$$

Then, $V_{Rp}(e_b) = \sum_{i=1}^N V_R(z'_i) \sum_{s \ni i} b_{s_i}^2 p(s) + V_p(t_b)$

[c.f. Eriksson (1973)]

Let $c_b^* = \sum_{i \in S} \frac{z_i'}{\pi_i}$. Then $E_R(c_b^*) = \sum_{i \in S} \frac{y_i}{\pi_i} = t_b^*$ (say)

and $V_{Rp}(c_b^*) = \sum_{i=1}^N \frac{V_R(z_i')}{\pi_i} + V_p(t_b^*)$.

So, $V_{Rp}(c_b) - V_{Rp}(c_b^*) > V_p(t_b) - V_p(t_b^*)$ since

$$\sum_{s \geq i} b_{s_i}^2 p(s) > \frac{1}{\pi_i} \quad \forall i.$$

Let us consider the super-population model M on $\underline{Y} = (y_1, \dots, y_i, \dots, y_N)$ such that

$E_M(y_i) = \theta_i$, $V_M(y_i) = \sigma_i^2$ and $C_M(y_i, y_j) = \rho \sigma_i \sigma_j \quad \forall i \neq j$

where θ_i in $(-\infty < \theta_i < \infty)$, σ_i^2 in $(0, \infty)$ and ρ in $(0, 1)$ are unknown constants. Then, under M ,

$$\begin{aligned} E_M V_{Rp}(c_b) &= \sum_{i=1}^N E_M V_R(z_i') \sum_{s \geq i} b_{s_i}^2 p(s) \\ &+ V_p(b_s + \sum_{i \in S} b_{s_i} \theta_i) + \rho V_p(\sum_{i \in S} b_{s_i} \sigma_i) \\ &+ (1-\rho) \sum_{i=1}^N \sigma_i^2 (\sum_{s \geq i} b_{s_i}^2 p(s) - 1) \end{aligned}$$

Also, under M , $E_M V_{Rp}(c_b^*) = \sum_{i=1}^N E_M V_R(z_i') / \pi_i$

$$\begin{aligned} &+ V_p(\sum_{i \in S} \frac{\theta_i}{\pi_i}) + \rho V_p(\sum_{i \in S} \frac{\sigma_i}{\pi_i}) \\ &+ (1-\rho) \sum_{i=1}^N \sigma_i^2 (\frac{1}{\pi_i} - 1). \end{aligned}$$

If we restrict within the class of designs of the form p_w^* , say,

such that the sample-size is a constant n and $\pi_i = nW_i/W$ for given W_i ($> 0 \forall i$), then under the model $M(W)$, say, modified from M so that

$\theta_i = \beta W_i$, $\sigma_i = \sigma W_i \forall i$ with β, σ^2 as unknown positive constants, it follows that

$$\begin{aligned} E_m V_{Rp_w^*}(a_b) &> \frac{W}{n} \sum_{i=1}^N \frac{1}{W_i} E_m V_R(z_i') + (1-\rho)\sigma^2 \left(\frac{W^2}{n} - \sum_{i=1}^N W_i^2 \right) \\ &= E_m V_{Rp_w^*}(e_b^*). \end{aligned}$$

For any other model as well on $\underline{Y}$ including 'Random permutation model' [vide, e.g. Bellhouse (1980)] and 'exchangeable distribution model' [vide Godambe/Thompson (1973)] we may argue that when $E_m V_p(t_b) > E_m V_p(t_b^*) \dots (i)$ holds, it will follow that $E_m V_{Rp}(a_b) > E_m V_{Rp}(e_b^*) \dots (ii)$. In the literature there exist well-known situations when (i) holds establishing (ii).

Let us consider another super-population model M' (say) on $\underline{Y}$ under which y_i 's are independently distributed with finite means θ_i 's. Then we will consider estimators for Y based on $\underline{z}$ of the class C of the form

$$e = e(s, \underline{z}'), \text{ free as usual of } z_i' \text{'s for } i \neq s,$$

$$\text{such that } E_p(e) = \sum_{i=1}^N z_i' = z' \text{ (say).}$$

$$\text{Let } e^* = e^*(s, \underline{z}') = \sum_{i \in S} z_i' / \pi_i$$

$$\text{and } e = e(s, \underline{z}') = e^*(s, \underline{z}') + h(s, \underline{z}') = e^* + h, \text{ (say)}$$

so that $E_p h(s, \underline{z}') = E_p(h) = 0 \quad \forall \underline{z}'$.

Then, we have $V_{pR}(e) = V_{pR}(e^* + h)$

$$\begin{aligned} &= E_p [\bar{V}_R(e^* + h)] + V_p E_R(e) \\ &= E_p V_R(e^*) + E_p V_R(h) + V_p E_R(e). \end{aligned}$$

This is so because $E_p C_R(e^*, h) = E_p E_R \left[\left(\sum_{i \in S} \frac{z'_i - y_i}{\pi_i} \right) h(s, \underline{z}') \right]$

$$= E_R \left[\sum_{i=1}^N \frac{z'_i - y_i}{\pi_i} \sum_{s \ni i} h(s, \underline{z}') p(s) \right]$$

$$= - E_R \left[\sum_{i=1}^N \frac{z'_i - y_i}{\pi_i} \sum_{s \not\ni i} h(s, \underline{z}') p(s) \right]$$

= 0, using $E_p h(s, \underline{z}') = 0$ and independence of

z'_i 's for the RRT.

So, $V_{pR}(e^*) = E_p V_R(e^*) + V_p E_R(e^*)$

and $V_{pR}(e) - V_{pR}(e^*) = E_p V_R(h) + V_p E_R(e^* + h) - V_p(E_R(e^*))$

$$= E_p V_R(h) + V_p E_R(e) - V_p \left(\sum_{i \in S} y_i / \pi_i \right).$$

Now, assuming that E_m, E_p, E_R commute, we get

$$\begin{aligned} E_M V_p(E_R(e)) &= E_M E_p [E_R(e) - E_p E_R(e)]^2 \\ &= E_p E_M [(E_R(e) - E_M E_R(e)) + (E_M E_R(e) - E_M E_p E_R(e)) \\ &\quad - (E_p E_R(e) - E_M E_p E_R(e))]^2 \\ &= E_p [V_M(E_R(e)) + \{E_M(E_R(e) - E_M Y)\}^2 - V_M(Y)] \end{aligned}$$

$$= E_p V_M(E_R(e)) + E_p \Delta_M^2(E_R(e)) - V_M(Y)$$

(writing $\Delta_M(E_R(e)) = E_M(E_R(e) - E_M(Y))$).

$$\begin{aligned} & E_M V_p(E_R(e)) - E_M V_p(E_R(e^*)) \\ &= E_p V_M(E_R(e^* + h)) - E_p V_M(E_R(e^*)) + E_p \Delta_M^2(E_R(e)) - E_p \Delta_M^2(E_R(e^*)) \\ &= E_p V_M(E_R(h)) + E_p \Delta_M^2(E_R(e)) - E_p \Delta_M^2(E_R(e^*)) \end{aligned}$$

because $E_p C_M(E_R(e^*), E_R(h))$

$$\begin{aligned} &= E_p E_M \left[\sum_{i \in S} \frac{y_i - \theta_i}{\pi_i} E_R(h) \right] \\ &= E_M \left[\sum_{i=1}^N \frac{y_i - \theta_i}{\pi_i} E_R \left(\sum_{s \neq i} h(s, \underline{z}') p(s) \right) \right] \\ &= - E_M \left[\sum_{i=1}^N \frac{y_i - \theta_i}{\pi_i} E_R \left(\sum_{s \neq i} h(s, \underline{z}') p(s) \right) \right] \end{aligned}$$

(using $E_p(h(s, \underline{z}')) = 0$)

= 0, using independence of y_i 's noting that

$E_R \left(\sum_{s \neq i} h(s, \underline{z}') p(s) \right)$ involves only y_i 's for $i \neq s$.

So, $E_M[V_{pR}(e) - V_{pR}(e^*)]$

$$\begin{aligned} &= E_M E_p V_R(h) + [E_M V_p(E_R(e)) - E_M V_p(E_R(e^*))] \\ &= E_M E_p V_R(h) + E_p V_M E_R(h) + E_p \Delta_M^2(E_R(e)) - E_p \Delta_M^2(E_R(e^*)). \dots (iii) \end{aligned}$$

Let a particular sub-class of M' be M'' for which

$\theta_i \propto w_i$. Then, for a p_W^* design $\Delta_M E_R(e^*) = 0$ implying

$$E_M[V_{p_{WR}}^*(e) - V_{p_{WR}}^*(e^*)] > 0 \text{ under model } M''.$$

Of course for any model, if $E_M V_p(E_R(e)) > E_M V_p(\sum_{i \in S} y_i / \pi_i)$... (iv)

it will follow that $E_M[V_{pR}(e)] > E_M[V_{pR}(\sum_{i \in S} z_i / \pi_i)]$ (v)

For non-linear estimators e however it is not clear how $E_R(e)$ is to be determined to check (iv) in practice in order to be able to claim the result (v) for other models in general.

REFERENCES

- Bellhouse, D.R. (1980). Linear models for randomized response designs. J. Amer. Statist. Assoc. 1001-1004.
- Brower, K.R.W. (1979). A class of robust sampling designs for large-scale surveys. J. Amer. Statist. Assoc. 74, 911-915.
- Chaudhuri, A. and Adhikary, A.K. (1984). A study of the effect of variate-transformations on strategies of sampling finite population. J. Ind. Soc. Agri. Statist. 36, 51-64.
- Cochran, W.G. (1977). Sampling Techniques. John Wiley and Sons. New York.
- Ericson, W.A. (1974). A note on the non-existence of minimum variance unbiased estimators for a finite population total. Sandhyā, Ser.C, 36, 181-185.
- Eriksson, Sven (1973). Randomized interviews for sensitive questions. Ph.D. Thesis, Dept. of Statistics, Gothenburg University.
- Fuller, W.A. and Isaki, C.T. (1981). Survey design under super-population models. In 'Current topics in Survey Sampling'. Eds. Krewski, D., Platek, R. and Rao, J.N.K., 199-226.
- Godambe, V.P. and Thompson, M.E. (1973). Estimation in sampling theory with exchangeable prior distributions. Ann. Statist. 1, 1212-1221.

- Isaki, C.F. and Fuller, W.A. (1982). Survey designs under the regression super-population model. *J. Amer. Statist. Assoc.*, 77, 89-96.
- Rao, C.R. (1952). Some theorems on minimum variance unbiased estimation. *Sankhyā A*, 12, 27-42.
- Rao, J.N.K. (1979). On deriving mean square errors and their non-negative unbiased estimators in finite population sampling. *J. Ind. Statist. Assoc.*, 17, 125-136.
- Rao, J.N.K and Bellhouse, D.R. (1978). Optimal estimation of a finite population mean under generalized random permutation models. *J. Stat. Plan. Inf.*, 2, 125-141.
- Rao, J.N.K. and Vijayan, K. (1977). On estimating the variance in sampling with probability proportional to aggregate size. *J. Amer. Stat. Assoc.*, 72, 579-584.
- Robinson, P.M. and Sarndal, C.E. (1983). Asymptotic properties of the generalized regression estimator in probability sampling. *Sankhyā, Ser.B.*, 45, 240-248.
- Royall, R.M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika*, 57, 377-387.
- Sarndal, C.E. (1980). On π -inverse weighting versus BLU weighting in probability sampling. *Biometrika*, 67, 639-650.