MODEL AND DESIGN-BASED ANALYSIS OF COMPLEX SURVEYS

(A Revised Version)

JOYDIP MITRA

A thesis submitted to the Indian Statistical Institute in partial fulfilment of the requirements for the degree of Doctor of Philosophy.
Calcutta
1993

1995

Declaration and Acknowledgements

This thesis is being submitted to the Indian Statistical Institute in partial 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.

I acknowledge my deep debt of gratitude to Dr. Arijit Chaudhuri who supervised the entire work, constantly encouraged me to work on the problems elaborated in what follows and permitted me to include in it a portion which has appeared as a paper entitled "A note on two variance estimators for Rao-Hartley-Cochran estimator", jointly by me with him in Communications in Statistics - Theory and Methods, 21(12), 1992, 3435-3443.

I am also grateful to the Dean of Studies, Heads of Computer Science Unit and the Computing and Statistical Services Centre of our Institute who granted me facilities for work including the use of PC's and the VAX mainframe computer.

Finally I must thank two anonymous external examiners whose constructive suggestoins enabled me to prepare this revised version as an improvement upon my original draft.

Joydip Mitra
December, 1993
December 1994.

PREFACE

We consider estimating the total Y of a variable y defined on a survey population. The survey is complex only in the sense that we admit sample selection with arbitrary probabilities. Our 'analysis' consists in examining efficacies of confidence intervals for the total. For this we need point estimators and the corresponding variance (or mean square error (MSE)) estimators, respectively say, e and v. The distribution, resulting from repeated sampling, of the pivotal quantity $d = (e-Y)/\sqrt{v}$ is supposed to approximate that of standard normal deviate τ or of Student's t with (n-1) degrees of freedom, assuming large sample size n. We will consider three general situations, namely when we presume that (i) 'direct responses' (DR) are available from sampled individuals, (ii) no direct but only 'randomized responses' (RR) may be gathered and (iii) there may be positive probability of nonresponse (NR) from at least some individuals sampled. In such cases we consider deriving new choices of (e, v)'s as alternatives to those existing in the current literature.

The thesis consists of eight chapters. Throughout the first seven of them we postulate a super-population model envisaging a linear regression of y on an auxiliary variable x. Our plan is to make use of the model in choosing appropriate v's, though e's may or may not be model-assisted. For a chosen e we consider the design-based MSE or an approximation of it. Every e we consider is either design-unbiased or asymptotically design-unbiased (ADU) in the sense of Brewer (1979) and Särndal (1980). The asymptotic approach of Fuller and Isaki (1981) and Isaki and Fuller (1982), however is nowhere followed in this thesis. Discussion will not be complete unless we refer to the recent text by Wolter (1985) that deals with variance estimation which also forms a principal endeavour on our part in this thesis.

To utilize both the design and the model in the choice of v we draw inspiration from the works of Brewer and Hanif (1983), Kumar, Gupta and Agarwal (1985), Brewer (1990) and Kott (1990a). Their approach is to consider the "model-based expectation" of (i) the design-based expectation of $(e-Y)^2$ and we intend first to extend it by permitting approximation of (i). Denoting this expectation by, say M, their procedures give a v such that the 'model-expectation' of the 'design-expectation' of v equals M. Kott's procedure goes a step further in that the 'model-expectation' of v equals the 'model-expectation' of $(e-Y)^2$. Novelty in our approach is that we find it 'necessary' and 'useful' to replace 'design-expectation' in this context by 'asymptotic design-expectation' in Brewer's sense. This modification leads to a series of alternative choices. This necessitates investigation of their efficacies ralative to their predecessors. In particular we also consider estimators for totals of y for specific domains. Necessary adjustments are made to cover (a) randomized responses and (b) 'non-responses'.

Realtive performances of alternative confidence intervals are dificult to examine theoretically. So we resort to numerical exercises. For this we undertake simulation studies. Through simulation-based studies we demonstrate that most of our newly proposed (e,v)'s yield competitively viable confidence intervals as assessed in terms of several well-known and a few new criteria for comparison, though in case of partial non-response situation we cannot be so assertive.

In the last chapter we evaluate relative efficacies of two well-known model-free but design-based (e,v)'s and utilize models exclusively for simulations in drawing conclusions.

For direct response surveys we cover only the ratio estimator, Horvitz-Thompson (1952) estimator (HTE), Särndal's (1980) generalized regression (Greg) predictor and Rao-Hartley-Cochran (RHC, 1962) estimator. Only sampling schemes employed are simple random sampling without replacement (SRSWOR), Hartley and Rao's (HR, 1962) sampling scheme and Rao-Hartley-Cochran sampling (RHC) scheme.

The table below briefly highlights our coverage of topics in brief.

Table of topics at a glance

Chap.	Estimators	Data set-up	Use of models	Sampling Scheme
1	HTE	DR	For v only	HR
2	Greg	DR	For both e and v	HR
3	Ratio	DR	For v only	SRSWOR
4	Domain	DR	For both e and v	HR
5	Ratio	RR	For v only	SRSWOR
6	Greg	RR	For both e and v	HR
7	(i) HTE	NR	no model for e or v	HR
	(ii) Greg	NR	For both e and v	HR
8	RHC	DR	no model for e or v	RHC

Note : DR ≡ Direct Response
RR ≡ Randomized Response
NR ≡ Partial Non-response
HR ≡ Hartley and Rao
RHC ≡ Rao, Hartley and Cochran

CONTENTS

INTRODUCTION

In this thesis we consider estimating the total Y of a real variable y defined on a survey population of a known number N of identifiable units labelled $i=1,\dots,N$. For this, sampling schemes are considered that are complex in the sense of differing from simple random sampling (SRS) with replacement (WR). For Y, confidence intervals (CI) are constructed involving a point estimator e of Y and a variance or mean square error (MSE) estimator v for e. The size n of a sample is supposed large. The distribution over repeated sampling of the pivotal quantity $d = (e-Y)/\sqrt{v}$ is supposed approximately to be close to that of the standard normal deviate τ or of Student's t-statistic with $(n-1)$ degrees of freedom. From this, $e \pm k\sqrt{v}$ provides a desired CI with k chosen from τ- or t- table for a pre-assigned nominal confidence coefficient. In the first seven chapters of this thesis, containing eight chapters in all, we shall concentrate on choice of v from considerations of both design and a model postulated connecting y and an auxiliary related variable x with known positive x_i- values. For e we shall consider well-known estimators either model-free or model assisted. We draw inspirations from Brewer and Hanif (1983), Kumar, Gupta and Agarwal (1985) and Brewer (1990) to obtain v not just as an estimator of the design mean-square error (MSE) of e but of the model expectation of the design MSE. Consequently v derives model-cum-design based properties. One cannot be sure if a postulated model is correct or wrong. But our intention is to get an improved variance estimator and hence a better CI if a postulated model may in fact happen to be correct and take advantage of that. Kott (1990a,b) also employs variance estimators studded simultaneously with model- as well as design- properties. As a generalization over these approaches we consider it "useful" as well as "necessary" to consider model-cum 'asymptotic-design'-based properties rather than exact 'design-based' properties.
Performances of variance or MSE estimators and CI's are evaluated from considerations of behaviours in hypothetically repeatable sampling. Improvement is of course not assured by invoking a model. By simulation we examine if a model-assisted procedure may fare better than a model-free procedure and if so to what extent.

Through the first seven chapters we adopt Brewer's (1979) 'asymptotic-design-based' approach. This requires mainly the use not of the design expectation operator E_p but of the 'limiting design expectation' operator $\lim E_p$ and application of Slutzky's (cf. Cramer (1966)) limiting theorem for convenience. The details will be explained in chapter one. To briefly indicate how we may derive a series of alternative choices of v for any fixed well-known e, let us note the following, with E_m as the model expectation operator. For e the design MSE is $E_p(e-Y)^2$, which is the variance $V_p(e)$ if e is design unbiased for Y. If e is not design unbiased we will always take it as an 'asymptotic design unbiased' (ADU) estimator in Brewer's sense. Sometimes we shall use only an approximation for the above MSE or variance. The model expectation of any of these design parameters or of $\lim E_p(e-Y)^2$ will be denoted as M. Our principal approach is to employ a v satsfying

$$\lim E_p E_m(v) = M.$$

We shall throughout assume that E_m commutes with E_p and $\lim E_p$. A considerably large series of such v's will be derived for the standard choice of e as the well-known (i) ratio estimator, (ii) Horvitz-Thompson (HT, 1952) estimator and quite a few also for (iii) Särndal's (1980) generalized regression (greg) predictor for Y. Our purpose is then to examine the efficacies of the resulting CI's relative to the traditional ones. Theoretical comparison seems difficult. So, we resort to numerical comparisons. For this we adopt simulation-based studies. In these studies we consider only design-based performances of the CI's. The role of model is only in yielding alternative choices of v. Their efficacies are examined via repeated sampling. For assessing the relative performances of the CI's we employ well-known criteria adding to them a few of our own.
Since we are not aiming at deriving any optimal variance or MSE estimator and the above model-cum-asymptotic design based approach obviously yields infinitely many alternatives, we find it imperative to resort only to numerical exercises through simulations.

In the first four chapters we restrict to simulations where values y_i of y are supposed available as direct responses (DR) from any individual i selected in a sample s taken from the population $U=(1,\ldots,i,\ldots,N)$. In chapters five and six we allow y to relate to sensitive and stigmatizing characteristics and hence instead of DR only 'randomized responses' (RR) are supposed to be available through suitably implemented devices from sampled persons. So, necessary adjustments are employed in choosing the combination (e,v). In chapter seven we allow positive probabilities of non-responses at least for some members of the population. So, further modifications are introduced for our 'analysis'. In chapter four we consider adjustments needed in analysis for estimating not Y but totals of y_i for units within a part called 'domain' of U though sample is drawn from U itself when a postulated model may apply either to the 'specific' domain or to the entire population.

In chapters three and five we consider ratio estimator and its RR-based modification, both based on simple random sampling without replacement (SRSWOR). In numerical illustrations concerning the Horvitz and Thompson's estimator and the generalized regression predictor the only scheme of sampling we use is that due to Hartley and Rao (1962). The size-measures needed for applying the sampling scheme are supposed available as the values of a third variable, say z, well- and positively- associated with y. However we take care "not" to keep the inclusion-probabilities π_i of the units proportional to x_i, $i \in U$. The sampling schemes with π_i proportional to x_i, called schemes with 'inclusion probability proportional to size' (IPPS or πps in brief) yield simplifications in analysis. But we avoid them, treating them as too restrictive because large-scale surveys cover many items or variables and so a particular z yielding π_i's cannot be supposed to 'meet this IPPS requirement' for every y of our interest to which an x is related permitting postulation of a *linear regression* of y on x. Our stress is mainly on data analysis after the sample is drawn, only taking care that a design may not be too bad to lead to poor analysis.

In chapter eight we consider the estimator given by Rao, Hartley and Cochran (1962) based on their own scheme of sampling and two variance estimators for it, one of which is given by themselves and another by Ohlsson (1989). Ohlsson's investigation seems to imply superiority of his estimator. We take up here a design-based comparison of the CI's respectively using these two variance estimators and report a simulation study which indicates a conclusion essentially to the contrary. Here we use a model only for generating the vectors $\underline{Y}=(y_1,\ldots,y_i,\ldots,y_N)$ and $\underline{X}=(x_1,\ldots,x_i,\ldots,x_N)$. Modifications here are not attempted to cover situations permitting RR and partial non-response.

The detailed findings are reported in the following chapters.

We may modestly add that through this thesis we do not intend to propagate any particular dogmatic view of our own about aptness of model-based or model-cum-design-based or classical approaches in sampling or of asymptotic theory in finite population inference. For our ideas about these we simply fall back upon well-known text books on sampling and on review papers one of which is the one by Bellhouse (1988).

CHAPTER ONE

A SIMULATION STUDY OF CONFIDENCE INTERVALS FOR SURVEY THROUGH HORVITZ - THOMPSON STRATEGIES

1.0 SUMMARY.

In order to construct appropriate confidence intervals for a finite population total with the Horvitz - Thompson estimator, (HTE, say) as a point estimator at the base we derive alternative variance estimators postulating the correctness of a linear regression model with a zero intercept. Permitting the use of sampling designs not necessarily with inclusion probabilities proportional to size-measures we find it convenient to aim at estimating the model expectation of the design-variance of HTE. We find a large number of variance estimators with limiting design - expectations of their model - expectations required to match the above aimed-at value. Analytic comparison of the resulting confidence intervals is difficult. So, we resort to a numerical comparison through a simulation study. We find the newly constructed variance estimators to yield confidence intervals promisingly competitive against the traditional Yates - Grundy variance estimator which does not utilize any model at all.

1.1 INTRODUCTION.

We consider a survey population $U=(1,\dots,i,\dots,N)$ on which are defined two real variables x and y with values x_i (>0, known) and y_i, $i=1,\dots,N$, with totals X and Y. The problem is to estimate Y. A super-population model M, say, is postulated permitting one to write

$$y_i = \beta x_i + \varepsilon_i, \quad i \in U. \tag{1.1.1}$$

Here β is an unknown constant; ε_i's are random variables with means $E_m(\varepsilon_i)=0$, variances $V_m(\varepsilon_i)=\sigma_i^2$ and covariances $C_m(\varepsilon_i,\varepsilon_j)=0, i\neq j$. A sample s from U is supposed to be drawn with probability p(s) according to a design p admitting positive inclusion-probabilities π_i for each i in U and π_{ij} for each distinct pair i, j in U. Each unit in s is supposed to be distinct and the size of s a fixed integer n. By Σ, $\Sigma\Sigma$ we shall denote sums over i in U and i, j (i<j) in U; Σ', $\Sigma'\Sigma'$ will mean the corresponding sums for units in s; by Σ_C we mean sum over i in U outside s. A design for which $\pi_i = nx_i/X$ (<1), $i \in U$, is called an IPPS or πps (inclusion probability proportional to size) design. Any other design is a non-IPPS design. By E_p (V_p) we shall mean expectation (variance) over design p. From Godambe and Joshi (1965), Godambe and Thompson (1977), Cassel, Särndal and Wretman (CSW, say, 1977) among others it is well-known that based on an IPPS design the classical Horvitz-Thompson (HT, say, 1952) estimator (HTE, say), namely

$$\bar{t} = \sum{}' \frac{y_i}{\pi_i}$$

is a good point estimator for Y. For $\bar{t}$ the value of

$$M = E_m E_p (\bar{t}-Y)^2$$

is suitably controlled vis-a-vis

$$E_m E_p (e-Y)^2$$

for a rival estimator e for Y satisfying design-unbiasedness condition

$$E_p(e) = Y \text{ for every } \underline{Y} = (y_1, \ldots, y_i, \ldots, y_N).$$

The value of M is further controlled if 'σ_i is proportional to x_i'.

In large scale sample surveys, however, in practice one can hardly employ an IPPS design. This is because (a) they involve many variables for each of which the survey population total or mean is required to be estimated, (b) a single design is adopted for the entire survey and as such (c) even though for every variable y of interest one may find an auxiliary variable x for which (1.1.1) is plausible, the IPPS requirement cannot be met for each such pair (y,x). Yet, HTE is traditionally 'an oft employed estimator' and is believed to perform well whether (1.1.1) is tenable or not, in the sense that

$$V=E_p(\bar{t}-Y)^2$$

is suitably under control if p is so designed that y_i's and π_i's may be positively well-correlated.

In this chapter we (i) rule out IPPS designs, (ii) consider HTE alone as a point estimator for Y, (iii) believe the model $\underline{M}$ of (1.1.1) as appropriate, propose to (iv) derive estimators say, v, for a suitably defined measure of error of $\bar{t}$, as done below and (v) examine the performances of confidence intervals (CI,say) for Y based on $(\bar{t},v)$ as competitors against a standard one, namely, $(\bar{t},v_{YG})$. Here

$$v_{YG} = \sum\sum' \Delta_{ij} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2$$

with $\Delta_{ij} = (\pi_i\pi_j - \pi_{ij}) / \pi_{ij}$, is the well-known estimator of V given by Yates and Grundy (YG, say, 1953).
Our intention is to demonstrate, if possible, that a variance estimator that unlike v_{YG} uses $\underline{M}$ is preferable to v_{YG} when $\underline{M}$ is plausible.

With any point estimator e for Y, linear in y_i, i in s, admitting a positive-valued variance estimator v, for large samples, it is usual to regard

$$d = \frac{(e-Y)}{\sqrt{v}}$$

as a variable t_{n-1} following Student's t-distribution with (n-1) degrees of freedom (d.f.) or as a standardized normal deviate τ with N(0,1) distribution. From this one justifiably sets up $100(1-\alpha)$ percent confidence interval (CI) as

$$\text{`}e \pm C_{\alpha/2}\sqrt{v},\ \alpha \text{ in } (0,1), \text{ for } Y\text{'},$$

where $C_{\alpha/2}$ is the upper $100\alpha/2$ percent point of the distribution of t or τ. Since v_{YG} does not use (1.1.1), we consider it of interest to try alternative variance estimators v for $\bar{t}$ in constructing CI's, namely, $\bar{t} \pm C_{\alpha/2}\sqrt{v}$, which utilize (1.1.1). For $\bar{t}$ 'based on IPPS designs' model-based variance estimators exist in the literature.

Brewer and Hanif (1983), Kumar, Gupta and Agarwal (1985) and Brewer (1990) approve of such a variance estimator, namely,

$$v_{KGA} = K_0 \sum\sum' \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2.$$

This is proposed by them to rectify the alleged deficiency of v_{YG} in the latter's possibility of yielding negative values. To fix the constant K_0 in it Kumar et al (1985)

(i) assume, following Smith (1938) and Brewer, Foreman, Mellor and Trewin (1979) among others, that

$$\sigma_i^2 = \sigma^2 x_i^g = \sigma^2 f_i, \text{ say, } \sigma\ (>0) \text{ unknown} \qquad (1.1.2)$$

but g is a known constant within [0,2]; in this case the model occassionally will be denoted by $\underline{M}(f)$;

(ii) note that for an IPPS design M equals

$$\sum \frac{\sigma_i^2}{\pi_i}(1-\pi_i) = M_0, \text{ say, and}$$

(iii) equate $E_m E_p(v_{KGA})$ to M_0, with σ_i^2 subject to (1.1.2).

Needless to mention, since $E_p(v_{YG}) = E_p(\bar{t} - Y)^2$, $E_m E_p(v_{YG})$ equals M_0 too.

Of course $E_p(v_{KGA}) \neq V$ i.e. v_{KGA} is not 'design-unbiased' and $E_m(v_{KGA}) \neq E_m(V)$ i.e. v_{KGA} is not 'model-unbiased' for V. For definition of model- unbiasedness we follow Royall (1970). Encouraged by this we seek 'model-based' variance estimators for $\bar{t}$ without insisting on requirements of (a) design-unbiasedness or (b) model-unbiasedness. But taking our main object as construction of CI for Y valid under large samples we seek 'Asymptotically design-cum-model-unbiased' variance estimators for $\bar{t}$. Explicitly, permitting the use "exclusively of non-IPPS designs" we seek variance estimators, say, m, satisfying

$$\lim E_p E_m(m) = E_m E_p(\bar{t}-Y)^2 = M = M_0 + \beta^2 V(x) \qquad (1.1.3)$$

say, writing

$$V(x) = V_p\left(\sum' \frac{x_i}{\pi_i}\right) = \sum\sum \Delta_{ij}\,\pi_{ij}\left(\frac{x_i}{\pi_i} - \frac{x_j}{\pi_j}\right)^2,$$

Noting $E_m(\bar{t}-Y)^2 = \Sigma\sigma_i^2(\frac{1}{\pi_i}-1)^2 + \Sigma_C\sigma_i^2 + \beta^2(\Sigma'\frac{x_i}{\pi_i}-X)^2$ it follows also that $E_pE_m(\bar{t}-Y)^2 = M_0 + \beta^2 V(x) = E_mE_p(\bar{t}-Y)^2$.

We assume throughout that E_p and E_m commute as operators and by $\lim E_p$ we mean the following, adopting Brewer's (1979) approach. According to this approach, we suppose that $U=(1,\ldots,i,\ldots,N)$, $\underline{Y}=(y_1,\ldots,y_i,\ldots,y_N)$, $\underline{X}=(x_1,\ldots,x_i,\ldots,x_N)$ and similarly other related vectors $\underline{W}=(w_1,\ldots,w_i,\ldots,w_N)$, w_i being values of a real variable w, reproduce themselves $T(>1)$ times in a way to yield the following entities:

$$U(j)=(\ (j-1)N+1,\ldots,(j-1)N+i,\ldots,(j-1)N+N\ ),$$
$$\underline{Y}(j)=(y_{(j-1)N+1},\ldots,y_{(j-1)N+i},\ldots,y_{(j-1)N+N}),\ j=1,\ldots,T,$$
$$U_T=(\ U(1),\ldots,U(j),\ldots,U(T)\ ),\ \underline{Y}_T=(\underline{Y}(1),\ldots,\underline{Y}(j),\ldots,\underline{Y}(T)),$$

such that for each fixed i $(=1,\ldots,N)$, $(j-1)N+i$ represents the same i for each $j=1,\ldots,T$. From each $U(j)$, samples $s(j)$ of the form as s are selected, 'independently' across $j=1,\ldots T$, according to the same design as p and these T samples are pooled into an amalgamated sample $s_T=(s(1),\ldots,s(T))$. The selection probability of s_T is then $p_T(s_T)=p(s(1))\cdots p(s(T))$, the resulting sampling design being p_T. If corresponding to an estimator $e=e(s)$ for Y one considers the estimator $e(s_T)$ for TY, then

$$\lim_{T\to\infty} E_{p_T}\left(\frac{1}{T}\,e(s_T)\right)$$

is abbreviated as $\lim E_p(e)$. If this equals Y, then e is 'Asymptotically design unbiased' (ADU) for Y. Employing Slutzky's theorem (ref. Cramér(1966)) applicable for continuous, especially rational functions, several simple and convenient 'asymptotic' results are available with this approach as will be illustrated in later sections.

Under the model $\underline{M}$, an ADU estimator, namely the well-known generalized regression (Greg, say) predictor

$$t_G = \sum{}' \frac{y_i}{\pi_i}\, g_{si},\ \ g_{si}=1+\left(X-\sum{}'\frac{x_k}{\pi_k}\right)\frac{Q_i\pi_i x_i}{\Sigma' Q_k x_k^2}, \qquad \mathbf{(1.1.4)}$$

for Y, with Q_i as any assignable positive constants, is available [ref. Särndal (1980) and Särndal, Swensson and Wretman (1989)] with its properties elaborately described in the recent book by Särndal, Swensson and Wretman (SSW, in brief, 1992). For this t_G, Särndal (1982) gave an approximate variance formula

$$\sum\sum \Delta_{ij}\, \pi_{ij} \left(\frac{E_i}{\pi_i} - \frac{E_j}{\pi_j}\right)^2$$

where $E_i = y_i - x_i B_Q$, $B_Q = \sum y_i x_i Q_i \pi_i \Big/ \sum x_i^2 Q_i \pi_i$, along with two estimators, to be briefly called Tay and Tay-2 respectively given by

$$v_{G1} = \sum\!\!\!\!\!\!\checkmark\sum\!\!\!\!\!\!\checkmark\ \Delta_{ij} \left(\frac{e_i}{\pi_i} - \frac{e_j}{\pi_j}\right)^2$$

and,
$$v_{G2} = \sum\!\!\!\!\!\!\checkmark\sum\!\!\!\!\!\!\checkmark\ \Delta_{ij} \left(\frac{g_{si} e_i}{\pi_i} - \frac{g_{sj} e_j}{\pi_j}\right)^2,$$

where, $e_i = y_i - x_i \hat{\beta}_Q$, $\hat{\beta}_Q = \sum\!\!\!\!\!\!\checkmark\ y_i x_i Q_i \Big/ \sum\!\!\!\!\!\!\checkmark\ x_i^2 Q_i$.

Checking that,

$$\lim E_p E_m (t_G - Y)^2 = \sum \frac{\sigma_i^2}{\pi_i}\left(1 - \pi_i\right) = M_O,$$

in estimating M in (1.1.3) for the component M_O in M we propose v_{Gj}, j=1,2 as two possible estimators — details discussed in section 1.2. Following Kott (1990a, b), a model-based estimator for 'a measure of error of $\bar{t}$' one may take

$$v_K = \frac{v}{E_m(v)} E_m(\bar{t} - Y)^2 \qquad (1.1.5)$$

with v as any variance estimator for $\bar{t}$ to start with, taken, say, as v_{YG} or v_{Gj}, j=1,2, provided v_K is free of unknown 'model parameters'. Unfortunately, in each of these three cases v_K is 'not model-free' and hence unavailable. In the next section we propose various choices of m subject to (1.1.3) and other alternative variance estimators for $\bar{t}$. Since analytic study of their properties is not easy, we consider various performance characteristics of CI's based on various choices of $(\bar{t}, v)$ through a numerical exercise carried out by simulations.

1.2 MODEL-BASED VARIANCE ESTIMATORS.

We throughout assume that $\underline{M}$ in (1.1.1) is tenable. To estimate M in (1.1.3) we need to estimate β^2 and M_0 but V(x) may be used itself or may be estimated by

$$\hat{V}(x) = \sum\nolimits'\sum\nolimits' \frac{\Delta_{ij}}{\pi_{ij}} \left(\frac{x_i}{\pi_i} - \frac{x_j}{\pi_j} \right)^2 \quad \text{or by } V^*(x) = (\Sigma' \tfrac{x_i}{\pi_i} - X)^2.$$

For β^2 the following three estimators are proposed, namely,

$$\hat{\beta}_1^2 = \sum\nolimits'\sum\nolimits' \frac{y_i \, y_j}{\pi_i \, \pi_j} \Big/ \sum\nolimits'\sum\nolimits' \frac{x_i \, x_j}{\pi_i \, \pi_j},$$

$$\hat{\beta}_2^2 = \sum\nolimits'\sum\nolimits' \frac{y_i \, y_j}{\pi_{ij}} \Big/ \sum\nolimits'\sum\nolimits' \frac{x_i \, x_j}{\pi_{ij}}$$

and,
$$\hat{\beta}_3^2 = \frac{2}{n(n-1)} \sum\nolimits'\sum\nolimits' \frac{y_i \, y_j}{x_i \, x_j}. \quad \text{Of course } E_m(\hat{\beta}_j^2) = \beta^2;\ j = 1, 2, 3.$$

Except simplicity these have no other known properties. Many other choices are possible. For illustration we restrict only to these three.

To estimate M_0 we proceed as follows. Let α_i's be weights to be appropriately chosen and $t(\alpha) = \Sigma'\alpha_i(r_i-\bar{r})^2$, where $r_i=y_i/x_i$, $\bar{r} = \Sigma' r_i/n$. Two sets of α_i's are suggested, namely, $\alpha_i(1)$ and $\alpha_i(2)$ respectively obtained on (i) equating $\lim E_p E_m[t(\alpha)]$ to M_0 and (ii) on equating $E_m[t(\alpha)]$ to $\Sigma'\sigma_i^2(1-\pi_i)/\pi_i^2$ as

$$\alpha_i(1) = \frac{n}{n-2} \left[\frac{x_i^2}{\pi_i^2} - \frac{1}{n(n-1)} \sum \frac{x_i^2}{\pi_i} (1-\pi_i) \right]$$

$$\alpha_i(2) = \frac{n}{n-2} \left[\frac{x_i^2}{\pi_i^2} - \frac{1}{n(n-1)} \sum\nolimits' \frac{x_i^2}{\pi_i^2} (1-\pi_i) \right].$$

From these we suggest estimating M_0 by any of the following:

$$\hat{M}_0(1) = \Sigma'\alpha_i(1)(r_i-\bar{r})^2, \quad \hat{M}_0(2) = \Sigma'\alpha_i(2)(r_i-\bar{r})^2,$$

$$\hat{M}_0(3) = \frac{E_p[\Sigma'\alpha_i(1)]}{\Sigma'_p\alpha_i(1)} \hat{M}_0(1)$$

$$= \frac{\hat{M}_0(1) \frac{(n-2)}{(n-1)} \sum \frac{x_i^2}{\pi_i}(1-\pi_i)}{\sum' \frac{x_i^2}{\pi_i^2}(1-\pi_i) - \frac{1}{n-1}\sum \frac{x_i^2}{\pi_i}(1-\pi_i)}$$

$$\hat{M}_0(4) = \frac{E_p[\Sigma'\alpha_i(2)]}{\Sigma'_p\alpha_i(2)} \hat{M}_0(2) = \frac{\sum \frac{x_i^2}{\pi_i}(1-\pi_i)}{\sum' \frac{x_i^2}{\pi_i^2}(1-\pi_i)} \hat{M}_0(2).$$

Alternatively, writing $z_i = y_i/\pi_i$, $\bar{z} = \Sigma' z_i/n$, weights α_i may be determined as $\alpha_i(3)$ and $\alpha_i(4)$ so as to estimate M_0 by $z(\alpha) = \Sigma'\alpha_i(z_i - \bar{z})^2$ such that $\lim E_p E_m[z(\alpha)]$ equals $M_0 + \beta^2 C$, with C as a known constant. This approach yields

$$\alpha_i(3) = \frac{n}{n-2}\left[(1-\pi_i) - \frac{1}{n(n-1)}(n - \Sigma\pi_i^2)\right]$$

and, $$\alpha_i(4) = \frac{n}{n-2}\left[(1-\pi_i) - \frac{1}{n(n-1)}(n - \Sigma'\pi_i)\right]$$

for which C respectively equals

$$C_1 = \sum \alpha_i(3)\pi_i\left(\frac{x_i}{\pi_i} - \frac{X}{n}\right)^2$$

and, $$C_2 = \sum \alpha_i(4)\pi_i\left(\frac{x_i}{\pi_i} - \frac{1}{n}\sum' \frac{x_k}{\pi_k}\right)^2.$$

Hence we propose two more estimators of M_0 as

$$\hat{M}_0(5) = \Sigma\alpha_i(3)(z_i - \bar{z})^2 - \hat{\beta}^2 C_1$$

and, $$\hat{M}_0(6) = \Sigma\alpha_i(4)(z_i - \bar{z})^2 - \hat{\beta}^2 C_2$$

with $\hat{\beta}^2$ as one of $\hat{\beta}_j^2$, j=1,2,3. In using TAY as an estimator for M_0 we consider only 4 alternative choices of Q_i as $\frac{1}{\pi_i x_i}$, $\frac{1-\pi_i}{\pi_i x_i}$, $1/x_i^2$ and $1/x_i$ for which we write v_{G1} respectively as v_H, v_B, v_S, $v_{S'}$ to associate the names of Hájek (1971) and Brewer (1979) with the first two as they adopted these choices and with the last two we associate the name of Särndal (1980) who first proposed the Greg predictor. For TAY-2 we use only one choice of Q_i for simplicity as $\frac{1}{\pi_i x_i}$ for which one may check that $v_{G2} = \left(\frac{X}{\Sigma' x_i/\pi_i}\right) v_H = v_T$ (say). Writing ϕ for B,H,S,S',T we propose then the following 66 choices of m namely,

$$m_j(\phi) = v_\phi + \hat{\beta}_j^2 V(x), \quad m_j(i) = \hat{M}_0(i) + \hat{\beta}_j^2 V(x),$$

and $\hat{m}_j(\phi)$, $\hat{m}_j(i)$ with $\hat{V}(x)$ in place of $V(x)$ in $m_j(\phi)$, $m_j(i)$, for $j=1,2,3$ and $i=1,..,6$ respectively. To these we add a few more, constraining (1.1.1) by (1.1.2). Writing $A(f)=\Sigma f_i(1-\pi_i)/\pi_i$, then M_0 equals $\sigma^2 A(f)$. For any choice of (Q_i,α_i), writing

$$t(Q,\alpha) = \sum{}' \alpha_i (y_i - x_i \hat{\beta}_Q)^2 \quad \text{we work out}$$

$$\text{(i) } E_m[t(Q,\alpha)] = \sigma^2 a(Q,\alpha), \quad \text{(ii) } \lim E_p E_m[t(Q,\alpha)] = \sigma^2 A(Q,\alpha),$$

where,

$$a(Q,\alpha) = \sum{}' \alpha_i \left[f_i \left(1 - \frac{2Q_i x_i^2}{\Sigma' Q_k x_k^2} \right) + x_i^2 \frac{\Sigma' Q_k^2 x_k^2 f_k}{(\Sigma' Q_k x_k^2)^2} \right] \quad \text{and,}$$

$$\begin{aligned} A(Q,\alpha) &= \lim E_p[a(Q,\alpha)] \\ &= \Sigma \alpha_i f_i \pi_i - 2(\Sigma \alpha_i Q_i x_i^2 f_i \pi_i)/(\Sigma Q_k x_k^2 \pi_k) \\ &\quad + (\Sigma \alpha_i x_i^2 \pi_i)(\Sigma Q_i^2 x_i^2 f_i \pi_i)/(\Sigma Q_k x_k^2 \pi_k)^2. \end{aligned}$$

Then we propose

$$\hat{M}_1(f) = \frac{t(Q,\alpha)}{a(Q,\alpha)} A(f) + \hat{\beta}^2 B$$

and,

$$\hat{M}_2(f) = \frac{t(Q,\alpha)}{A(Q,\alpha)} A(f) + \hat{\beta}^2 B$$

as alternative choices of m subject to (1.1.3) with B as either $V(x)$ or $\hat{V}(x)$ and $\hat{\beta}^2$ as $\hat{\beta}_j^2$, $j=1,2,3$.

Moreover, starting with $v_z = \Sigma'(z_i - \bar{z})^2$ and noting that

(i) $E_m(v_z) = \sigma^2 a + \beta^2 b$, say, with

$$a = \frac{n-1}{n}\sum{}' \frac{f_i}{\pi_i^2} \text{ and, } b = \sum{}' \left(\frac{x_i}{\pi_i} - \frac{1}{n}\sum{}' \frac{x_k}{\pi_k}\right)^2$$

and, (ii) $\lim E_p E_m(v_z) = \sigma^2 a' + \beta^2 b'$, where,

$$a' = \frac{n-1}{n}\sum \frac{f_i}{\pi_i} \text{ and, } b' = \sum \pi_i \left(\frac{x_i}{\pi_i} - \frac{X}{n}\right)^2,$$

we propose further alternative forms of m as

$$\hat{M}_1(z) = A(f)\left(v_z - \hat{\beta}^2 b\right)/a + \hat{\beta}^2 B$$

and, $$\hat{M}_2(z) = A(f)\left(v_z - \hat{\beta}^2 b'\right)/a' + \hat{\beta}^2 B.$$

In our simulation studies reported in Appendix-A, at the end of this chapter, we do not cover $\hat{M}_2(f)$ and $\hat{M}_2(z)$ but treat $\hat{M}_1(f)$ with 8 alternative choices of (Q_i, α_i) as $(\frac{1}{f_i}, 1)$, $(\frac{1}{f_i}, \frac{1}{f_i})$, $(\frac{1}{f_i}, \frac{1}{\pi_i})$, $(\frac{1}{f_i}, \frac{1}{\pi_i^2})$, $(\frac{1}{f_i}, \frac{1-\pi_i}{\pi_i^2})$, $(\frac{1}{\pi_i x_i}, \frac{1}{\pi_i})$, $(\frac{1}{\pi_i x_i}, \frac{1}{\pi_i^2})$ and $(\frac{1}{\pi_i x_i}, \frac{1-\pi_i}{\pi_i^2})$. These 8 forms of $\hat{M}_1(f)$ will be denoted by m_{ij} with B as V(x) and $\hat{m}_{ij}$ with B as $\hat{V}(x)$, i=1,..,8, and j=1,2,3 for $\hat{\beta}_j^2$ as $\hat{\beta}^2$. The 3 choices of $\hat{M}_1(z)$ with $\hat{\beta}_j^2$ as $\hat{\beta}^2$ will be denoted by m_{9j} with B as V(x) and by $\hat{m}_{9j}$ with B as $\hat{V}(x)$.

Infinitely many more choices of m subject to (1.1.8) are obviously possible. We consider the above choices only to try for alternatives to v_{YG} which may fare better than it when $\underline{M}$ is appropriate. Any general quadratic form in the sampled y_i's or $\Sigma\alpha_i(y_i - \widehat{\beta_Q}x_i)^2$ instead of $t(\alpha)$ or $s(\alpha)$ might be tried applying the constraint (1.1.8) and $V^*(x)$ in lieu of $V(x)$ and $\widehat{V}(x)$. Since our exercise is numerical we illustrate only a few.

1.3 SIMULATION AND CRITERIA MEASURE FOR COMPARISON OF COMPETITIVE CONFIDENCE INTERVALS.

We take N=150, σ=1, β=1, a few choices of g in [0,2], ε_i's as independent $N(0, x_i^g)$ variables and draw x_i's independently from the density

$$f(x) = \frac{1}{8.5} e^{-x/8.5}, \quad x>0$$

to generate $\underline{Y}=(y_1,\ldots,y_i,\ldots,y_N)'$, $\underline{X}=(x_1,\ldots,x_i,\ldots,x_N)'$ subject to (1.1.1). Also we take a few choices of h in [0,2] to obtain size-measures $w_i=x_i^h$ to be used in drawing samples of size n=32 *from* U=(1,...,150) adopting the well-known sampling scheme due to Hartley and Rao (HR, in brief, 1962). For this, π_i taken proportional to w_i will correlate well and positively with y_i, $i \in U$. To calculate $d=(e-Y)/\sqrt{v}$ for various choices of v and e taken as $\bar{t}$, we replicate sampling R=1000 times. By Σ_r we denote sum over these R replicates. To discriminate among the CI's given by $e \pm C_{\alpha/2}\sqrt{v}$ we consider the following criteria heeding Rao and Wu's (1983) works. In our numerical illustrations we show only $\alpha=0.05$ and $C_{\alpha/2}$ as $\tau_{\alpha/2}$ and write

$$A=\frac{1}{R}\sum_r (e-Y)^2 \quad \text{and,} \quad P=\frac{1}{R}\sum_r v.$$

(1) ACP (Actual coverage percentage) : The percentage of the R replicates for which CI covers Y. The closer this is to $100(1-\alpha)$ the better.

(2) ACV (Average coefficient of variation) : The average over R replicates, *of* $\sqrt{v}/e$ *reflecting* the length of CI relative to e. The smaller it is the better.

(3) PB (Pseudo relative bias of v) : $B(v)=\frac{1}{A}\left(\frac{1}{R}\sum_r v - A\right)$.

(4) PS (Pseudo relative stability of v) : $S(v)=\frac{1}{A}\left[\frac{1}{R}\sum_r (v-A)^2\right]^{1/2}$.

(5) PL (Pseudo standardized length) : $L(v)=\frac{1}{R}\sum_r \sqrt{v}\,/\sqrt{A}$.

(6) $B(\cdot)$ (Bias of d) : $B(d)=\frac{1}{R}\sum_r d$.

(7) $M(\cdot)$ (Mean square error (MSE) of d) : $M(d)=\frac{1}{R}\sum_r (d-B(d))^2$.

(8) $\sqrt{\beta_1(\cdot)}$ (Root beta *one*) : $\sqrt{\beta_1(d)}=\frac{1}{R}\sum_r\left(\frac{d-B(d)}{\sqrt{M(d)}}\right)^3$.

(9) $E(\cdot)$ (Excess measure) : $E(d)=\beta_2(d)-3=\frac{1}{R}\sum_r\left(\frac{d-B(d)}{\sqrt{M(d)}}\right)^4-3$.

(10) PCV (pseudo coefficient of variation) : $\frac{1}{P}\left[\frac{1}{R}\sum_r (v-P)^2\right]^{1/2}$.

The smaller these (3)-(10), the better.

To use (e,v) with (1.1.2) assumed to hold good we take 4 choices of g as 0.4, 0.8, 1.2 and 1.6 and compare the CI's in terms of (1)-(10) above. If the discrepancies over these 4 choices of g are small then we claim 'robustness' of the procedures. To make the procedures still more robust we add an intercept θ in the model $\underline{M}$ of (1.1.1). We numerically illustrate only one choice of θ as 10.0. We write g_0 for this g to distinguish it from g in $N(0, x_i^g)$ used in generating $\underline{Y}$.

Besides the global empirical studies as above where the criterion measures relate to all the R replicates, following Royall and Cumberland (1985) and Wu and Deng (1983) among others, a conditional empirical study is also made. Following Godambe (1989) we take $\sum' \frac{x_i}{\pi_i}$ as an 'ancillary' statistic and split the R samples into 10 equal groups. The r-th group is so formed that the set of 100 replicates for which the values of $\sum' \frac{x_i}{\pi_i}$ are the least, constitute the first group, the second group consisting of the next 100 consecutive higher values of $\sum' \frac{x_i}{\pi_i}$ and so on. Then, G=10 sets of each of the measures formed from the respective groups are calculated and compared group-wise. Writing v_r, A_r and P_r for v, A and P respectively as calculated from the r-th group of 100 replicates we write

$$d(v) = \left[\frac{1}{G} \sum_{r=1}^{G} \left(\sqrt{P_r} - \sqrt{A_r} \right)^2 \right]^{1/2}$$

to denote an over-all measure of performance of CI based on $(\bar{t}, v)$. The smaller d(v) the better is v. Further, we take $\sum' \frac{1}{\pi_i}$ and $\bar{X} \left(\sum' \frac{1}{\pi_i} \Big/ \sum' \frac{x_i}{\pi_i} \right)$ as additional 'ancillary' statistics and repeat the same study. Numerical illustrations are not reported for these. For discussions on such criteria one may consult among others, Rao and Wu (1983), Chaudhuri and Stenger (1992).

The findings are illustrated in Tables A.1-A.6 in Appendix-A at the end of this chapter and some remarks are made in section 1.4.

1.4 REMARKS ON ILLUSTRATED FINDINGS.

Numerical values are illustrated selectively to highlight better performances of the newly proposed variance estimators. Inferior performances are generally omitted but *even the inferior* variance estimators whose performances we do not show are never worse than v_{YG} except occasionally in respect of d(v). Even the otherwise good ones turn out worse than v_{YG} and $v(T)$ in terms of d(v). *To* emphasize better performances of some of our proposed procedures some favourable values are 'underscored' while the unfavourable ones are 'starred'. The variance estimators $m_j(5)$, $m_j(6)$, $\hat{m}_j(5)$, $\hat{m}_j(6)$, j=1,2,3 turn out less impressive as improvements over v_{YG}. The choice of g in [0,2] is not very crucial in yielding variance estimators $\hat{m}_{ij}$, i=1,..,9; j=1,2,3, but smaller values of g seem to be better choices.

More comments follow at the bottom of each table below. A message that seems to emerge from our numerical findings is that if the model $\underline{M}$ is correctly postulated, then some of our newly proposed model-based variance estimators may be profitably employed as better alternatives to the traditional Yates-Grundy variance estimator which uses no model. Our findings *displayed in the tables below may* assist in making a judicious choice in a given situation depending on the importance one may attach to the various performance criteria we mention. Of course our findings have limitations because real life situations may not match the simplifying postulations we have made. Since we believe that if a model is correctly postulated then it should be used in analysis expecting better results than without using it but we have no theory to prove that this must be so, we present our numerical findings to provide evidence which seems to support what we anticipate though not in a very pronounced or obvious manner. We believe this exercise is worth reporting.

Appendix A

The tables below use the following abbreviations described on p-15: ACP = Actual coverage percentage; ACV = Average coefficient of variation (CV); PB = Pseudo relative bias of v; PS = Pseudo relative stability of v; PL = Pseudo standardized length; B(.) = bias; M(.) = MSE; $\sqrt{\beta_1}$ = Root beta one; E(.) = Excess measure and PCV = Pseudo CV. Further $\bar{t}$ = Horvitz-Thompson estimator and v = estimator of $E_m E_p(\bar{t} - Y)^2$.

Table A.1

Performances of $(\bar{t}, v)$ under $\underline{M}$. g=1.2, β=1.0, h=1.6.

Especially good (bad) values are underscored (starred).

v	10^4PCV	ACP	10^5ACV	PB	PS	PL	10^3B(d)	M(d)	$\sqrt{\beta_1}$(d)	E(d)
v_{YG}	1768*	94.1	4930*	$\underline{.0012}$	.1772*	.9808*	17.73*	1.097	.13*	.21*
$\hat{m}_1$(B)	1493	94.1	4906	.0128	.1457	.9768	.57	$\underline{1.083}$	.06	.09
$\hat{m}_1$(H)	1493	94.1	4905	.0131	.1456	.9767	$\underline{.54}$	$\underline{1.083}$	.06	.09
$\hat{m}_1$(S)	1493	94.1	4906	.0128	.1457	.9768	.57	$\underline{1.083}$	.06	.09
$\hat{m}_1$(S′)	1491	94.1	4903	.0142	.1451	.9762	$\underline{.44}$	1.085	.06	.09
$\hat{m}_2$(1)	$\underline{1467}$	94.3	$\underline{4899}$	.0158	$\underline{.1423}$	$\underline{.9756}$	1.04	1.085	$\underline{.05}$	.09
$\hat{m}_2$(2)	$\underline{1467}$	94.3	$\underline{4899}$	.0158	$\underline{.1424}$	$\underline{.9756}$	1.01	1.085	$\underline{.05}$	.09
$\hat{m}_3$(3)	1483	94.2	4904	.0138	.1444	.9764	$\underline{.03}$	1.058	.06	.08
$\hat{m}_3$(4)	1483	94.2	4904	.0138	.1444	.9764	$\underline{.03}$	1.084	.06	.08
$\hat{m}_1$(5)	1756*	94.2	4930*	$\underline{.0008}$	.1759*	.9808*	16.71*	1.098	.13*	.21*
$\hat{m}_2$(6)	1478	94.2	4932*	.0016	.1754*	.9812*	16.28*	1.095	.12*	.21*
$\hat{m}_3$(T)	1550	94.3	4908	.0112	.1517	.9771	3.14	1.085	.08	.09

Comments: Possibly because of postulated normality of errors each ACP is so good; each model-based v except $\widehat{m_1}(5)$ has a vastly superior PCV to that of model-free v_{YG}; except $\widehat{m_1}(5), \widehat{m_1}(6)$ each has much better ACV than v_{YG} giving CI's with shorter lengths. In other respects also v_{YG} is outperformed by others except in terms of pseudo relative bias. Since v_{YG} is design unbiased while others are not, PB for v_{YG} should naturally be small as it turns out to be.

Table A.2

Conditional performances of $(\bar{t}, v)$ under $\underline{M}$. g=1.2, β=2.0, h=1.6.

Ancillary=$\Sigma' x_i/\pi_i$. Parentheses give values of (10^4PCV, ACP, 100ACV, B(d)).

Group-wise values given consecutively for 10 groups.

v	(10^4PCV, ACP, 100ACV, B(d))		
v_{YG}	(1303, 92, 2.63, .07)	(1708, 92, 2.65, .10)	(1665, 95, 2.66, .18)
	(1490, 93, 2.65, .02)	(1974, 95, 2.59, .06)	(1289, 92, 2.57, .14)
	(1307, 96, 2.64, .02)	(1823, 98, 2.69, .14)	(1778, 92, 2.63, .10)
	(1725, 97, 2.75, .07)		

Table A.2 (continued)

v	(10^4PCV, ACP, 100ACV, B(d))		
$\hat{m}_1(B)$	(965, 91, 2.60, .05)	(1246, 94, 2.61, .08)	(1123, 96, 2.66, .14)
	(1029, 94, 2.62, .01)	(1425, 94, 2.55, .08)	(849, 92, 2.56, .11)
	(989, 95, 2.60, .07)	(1262, 97, 2.67, .16)	(1177, 94, 2.60, .07)
	(1098, 98, 2.73, .10)		
$\hat{m}_1(1)$	(951, 91, 2.60, .05)	(1228, 94, 2.61, .08)	(1090, 96, 2.65, .14)
	(985, 94, 2.62, .02)	(1394, 94, 2.55, .08)	(826, 92, 2.55, .11)
	(964, 95, 2.60, .01)	(1235, 97, 2.66, -.17)	(1146, 94, 2.61, .06)
	(1094, 98, 2.73, -.10)		
$\hat{m}_3(2)$	(950, 91, 2.60, .05)	(1227, 94, 2.61, .08)	(1089, 96, 2.65, .14)
	(984, 94, 2.62. .02)	(1393, 94, 2.55, .08)	(825, 92, 2.55, .11)
	(964, 95, 2.60, .01)	(1233, 97, 2.66, -.17)	(1144, 94, 2.61, .06)
	(1091, 98, 2.73, -.10)		
$\hat{m}_3(T)$	(993, 90, 2.60, .06)	(1284, 94, 2.61, .08)	(1185, 95, 2.65, .15)
	(1113, 94, 2.62, .01)	(1481, 95, 2.56, .07)	(892, 92, 2.56, .12)
	(1043, 95, 2.60, .00)	(1313, 97, 2.67, -.16)	(1230, 94, 2.60, .07)
	(1099, 98, 2.74, -.10)		

Comments: Possibly because of reduced group-wise numbers of replicates we notice fluctuations in ACP-values in the range 90-98 per cent. Though the smallest and the largest ACP values correspond more or less respectively to the least and the highest values of the ancillary for every v, no definite linear trend is discernible. Similar is for PCV and ACV but B(d) behaves quite irregularly. But in vindication of our approach, v_{YG} is outperformed by the four alternatives illustrated in this table.

Table A.3

Conditional performances of $(\bar{t},v)$ under $\underline{M}$. g=1.2, β=1.0, h=1.6. Ancillary=$\Sigma' x_i/\pi_i$. Parentheses give values of (10^4PCV, ACP, 100ACV, B(d)). Group-wise values given consecutively for 10 groups.

v	(10^4PCV, ACP, 100ACV, B(d))		
v_{YG}	(1467, 93, 4.88, .06)	(1870, 93, 4.92, .06)	(1761, 95, 4.97, .15)
	(1571, 96, 4.95, -.01)	(2219, 93, 4.79, .07)	(1395, 94, 4.77, .15)
	(1466, 93, 4.90, -.01)	(1972, 98, 5.05, -.13)	(1923, 90, 4.90, .05)
	(1791, 96, 5.17, .07)		
$m_1(S')$	(1287, 94, 4.85, .05)	(1604, 94, 4.88, .04)	(1445, 95, 4.95, .13)

Table A.3 (continued)

v	(10^4PCV, ACP, 100ACV, B(d))		
	(1291, 95, 4.92, -.03)	(1905, 93, 4.77, .08)	(1146, 94, 4.75, .14)
	(1296, 92, 4.85, .03)	(1655, 98, 5.02, -.15)	(1582, 90, 4.87, .03)
	(1438, 97, 5.16, .09)		
$m_3(1)$	(1271, 94, 4.85, .05)	(1584, 96, 4.89, .04)	(1436, 95, 4.96, .13)
	(1237, 95, 4.92, .03)	(1867, 93, 4.76, .08)	(1122, 94, 4.75, .13)
	(1266, 92, 4.85, .03)	(1628, 98, 5.02, -.15)	(1551, 90, 4.87, .03)
	(1436, 97, 5.15, -.09)		
$m_3(2)$	(1271, 94, 4.85, .05)	(1585, 96, 4.89, .04)	(1440, 95, 4.96, .13)
	(1239, 95, 4.92, -.03)	(1868, 93, 4.76, .08)	(1123, 94, 4.75, .13)
	(1267, 92, 4.85, -.03)	(1628, 98, 5.02, -.03)	(1552, 90, 4.87, .03)
	(1436, 97, 5.15, -.09)		
$m_3(T)$	(1316, 94, 4.86, .05)	(1651, 94, 4.88, .04)	(1521, 95, 4.96, .13)
	(1400, 96, 4.93, -.03)	(1974, 93, 4.78, -.08)	(1200, 94, 4.76, .14)
	(1363, 92, 4.85, -.26)	(1723, 98, 5.04, -.15)	(1653, 90, 4.87, .03)
	(1440, 97, 5.16, -.09)		

Comments: Here also ACP's range from 90 to 98 per cent possibly again because of small replication sizes but their pattern is quite irregular. Similar is with the other criteria. But to our satisfaction v_{YG} turns out the poorest performer among the five displayed.

Table A.4

Robustness of $(\bar{t}, v)$. g=1.2, β=1.0, h=1.6, θ=10.0. Parentheses give values for four choices of g_0 as .4, .8, 1.2, 1.6. The first rows for a particular v give values of (10^4PCV, ACP, PB) in this order and second rows give values for 100ACV and -10B(d) respectively.

	g_0		
v	(.4 .8 1.2 1.6)	(.4 .8 1.2 1.6)	(.4 .8 1.2 1.6)
$\hat{m}_{11}$	(388, 402, 421, 444)	(97.8, 97.8, 97.3, 97.0)	(.45, .39, .35, .31)
		(3.70, 3.63, 3.57, 3.52)	(.69, .73, .77, .82)
$\hat{m}_{22}$	(390, 406, 426, 451)	(97.8, 97.8, 97.8, 97.6)	(.47, .44, .43, .43)
		(3.73, 3.69, 3.68, 3.68)	(.68, .72, .75, .77)
$\hat{m}_{31}$	(407, 420, 434, 450)	(97.8, 97.8, 97.8, 97.6)	(.56, .52, .47, .43)

Table A.4 (continued)

v	g_0 (.4	.8	1.2	1.6)	(.4	.8	1.2	1.6)	(.4	.8	1.2	1.6)
					(3.85,	3.79,	3.73,	3.68)	(.67,	.70,	.74,	.77)
$\hat{m}_{43}$	(460,	478,	495,	512)	(98.6,	98.3,	98.2,	98.1)	(.74,	.71,	.67,	.64)
					(4.05,	4.01,	3.97,	3.84)	(.63,	.66,	.69,	.71)
$\hat{m}_{52}$	(475,	494,	513,	530)	(98.6,	98.5,	98.2,	98.2)	(.78,	.75,	.72,	.69)
					(4.10,	4.06,	4.03,	3.99)	(.62,	.65,	.67,	.70)
$\hat{m}_{63}$	(453,	473,	493,	513)	(98.4,	98.2,	98.1,	98.1)	(.65,	.61,	.57,	.53)
					(3.94,	3.90,	3.85,	3.80)	(.69,	.73,	.76,	.80)
$\hat{m}_{72}$	(417,	440,	464,	489)	(98.2,	98.1,	98.1,	98.1)	(.62,	.61,	.60,	.59)
					(3.91,	3.91,	3.89,	3.88)	(.60,	.63,	.67,	.70)
$\hat{m}_{81}$	(421,	443,	468,	494)	(98.3,	98.2,	98.0,	98.1)	(.62,	.62,	.62,	.62)
					(3.91,	3.91,	3.91,	3.91)	(.57,	.61,	.64,	.68)
$\hat{m}_{93}$	(552,	564,	578,	594)	(93.0,	92.9,	92.6,	92.6)	(.46,	.51,	.56,	.60)
					(3.00,	2.99,	2.99,	2.98)	(.58,	.62,	.67,	.71)

Comments: Except for $\hat{m}_{93}$, the ACP exceeds 95 per cent but in every case there is little variation with changing g_0. In respect of ACV also there is undoubted robustness. In terms of PB the robust procedures are $\hat{m}_{22}$, $\hat{m}_{72}$ and $\hat{m}_{81}$. None seems robust in terms of PCV. In terms of PB the variace estimators $\hat{m}_{43}$ and $\hat{m}_{52}$ seem to be robust.

Table A.5

d-values for v with ancillary $\Sigma' x_i/\pi_i$ and various (g, β, h)

I. $g=1.2$, $\beta=1.0$, $h=1.6$

v :	v_{YG}	$\hat{m}_1(B)$	$\hat{m}_1(H)$	$\hat{m}_1(S)$	$\hat{m}_1(S')$	$\hat{m}_1(1)$	$\hat{m}_1(2)$	$\hat{m}_1(3,4,5,6)$	$\hat{m}_1(T)$
d :	5.43	5.45	5.45	5.45	5.45	5.46	5.46	5.45	<u>5.42</u>

II. $g=1.6$, $\beta=1.0$, $h=1.6$

v :	v_{YG}	$\hat{m}_1(B)$	$\hat{m}_1(S)$	$\hat{m}_1(S')$	$\hat{m}_1(1)$	$\hat{m}_1(2)$	$\hat{m}_1(3)$	$\hat{m}_1(4)$	$\hat{m}_1(5,6)$	$\hat{m}_1(T)$
d :	8.07	8.06	8.06	8.06	8.07	8.07	8.06	8.06	8.08*	<u>8.03</u>

III. $g=1.2$, $\beta=2.0$, $h=1.6$

v :	v_{YG}	$m_1(B)$	$m_1(H)$	$m_1(S)$	$m_1(S')$	$m_1(1,2,3,4)$	$m_1(5)$	$m_1(6)$	$m_2(6)$	$m_1(T)$
d :	6.17	6.18	6.18	6.19	6.19	6.20*	6.18	<u>6.16</u>	<u>6.16</u>	<u>6.15</u>

Comment: Each variance estimator is nearly at par.

Table A.6

Performances of $(\bar{t}, v)$ under $\underline{M}(f)$. θ=0.0, g=1.2, β=1.0, h=1.6. Parentheses give values for four choices of g_0 as .4, .8, 1.2, 1.6. For v_{YG}, however, only one entry is relevant with g_0 equal to 1.2. The first rows for a particular v (only one entry of course for v_{YG}) give values of (10^4PCV, ACP, 10^5ACV) in this order and the second rows give values of PB and -10B(d) in succession.

v	g_0 (.4 .8 1.2 1.6)	(.4 .8 1.2 1.6)	(.4 .8 1.2 1.6)
v_{YG}	(—, —, 1768, —)	(—, —, 94.1, —)	(—, —, 4930, —)
		(—, —, .00, —)	(—, —, .02, —)
m_{11}	(6333, 6292, 6275, 6280)	(95.9, 95.2, 94.8, 94.2)	(5104, 4989, 4878, 4773)
		(.05, .01, .04, .08)	(-.04, .04, .03, .03)
m_{22}	(6756, 7218, 7812, 8549)	(95.9, 95.1, 94.9, 94.7)	(5084, 4947, 4892, 4833)
		(.05, .00, .03, .05)	(.03, .03, .03, .02)
m_{31}	(8568, 8519, 8507, 8526)	(95.2, 95.1, 94.9, 94.8)	(5030, 4963, 4899, 4837)
		(.03, .00, .03, .05)	(.03, .02, .02, .02)
m_{43}	(12731, 12647, 12615, 12630)	(94.6, 94.6, 94.6, 94.7)	(4993, 4918, 4903, 4889)
		(.00, .01, .02, .02)	(.01, .01, .01, .01)
m_{51}	(14126, 14032, 13995, 14008)	(94.5, 94.5, 94.6, 94.6)	(4907, 4905, 4903, 4902)
		(.01, .01, .01, .02)	(.01, .01, .01, .01)
m_{63}	(8618, 8559, 8530, 8563)	(95.3, 95.1, 94.9, 94.7)	(5028, 4963, 4898, 4886)
		(.03, .00, .03, .05)	(.03, .02, .02, .02)
m_{73}	(12705, 12617, 12571, 12616)	(94.6, 94.6, 94.6, 94.7)	(4940, 4925, 4909, 4893)
		(.00, .01, .02, .02)	(.01, .01, .01, .01)
m_{83}	(14065, 13969, 13918, 13965)	(94.4, 94.5, 94.5, 94.6)	(4917, 4914, 4910, 4808)
		(.00, .01, .01, .01)	(.01, .01, .01, .01)
m_{93}	(15223, 15144, 15119, 15143)	(94.4, 94.3, 94.3, 94.0)	(4969, 4952, 4936, 4921)
		(.01, .01, .00, .01)	(.00, .00, .01, .01)

Comments: All the m_{ij}'s are poorer than v_{YG} in terms of PCV but have better ACV for every g and are better in terms of ACV at least when g_0 is not less than g. In terms of B(d) also they outperform for most choices of g_0.

CHAPTER TWO

CONFIDENCE INTERVAL ESTIMATION USING GENERALIZED REGRESSION PREDICTOR AND ITS MODEL-CUM-DESIGN-BASED VARIANCE ESTIMATORS

2.0 SUMMARY.

In this follow-up of Chapter One we pursue with the same model $\underline{M}$, but consider instead of the Horvitz-Thompson estimator the generalized regression (greg) predictor as the basic point estimator for the survey population total Y. To construct confidence intervals for Y we derive variance estimators for it adopting Brewer's asymptotic approach as in Chapter One. Working out the asymptotic limiting design expectation of the model expectation M of the squared difference between the greg predictor and Y, estimators m for it called variance estimators are derived. Such an m is required to be asymptotically design-cum-model-unbiased for M. Theoretical comparison among the new and traditional variance estimators of the greg predictor is again found dificult. So, design-based efficacies of the confidence intervals based on them are numerically compared through simulations according to various criteria. Granting correctness of a postulated regression model some of the newly derived variance estimators are demonstrated to perform as good competitors against the traditional ones in yielding serviceable confidence intervals.

2.1 INTRODUCTION.

Pursuing with the same model and notations as in Chapter One we consider here for Y the generalized regression (greg) predictor as the basic estimator, given by

$$t_g = \sum{}' \frac{y_i}{\pi_i} g_{si} ,$$

$$g_{si} = 1 + \left(X - \sum{}' \frac{x_i}{\pi_i} \right) \frac{x_i Q_i \pi_i}{\Sigma' x_k^2 Q_k} . \qquad (2.1.1)$$

It follows that,

$$\lim E_p E_m (t_g - Y)^2 = \sum \sigma_i^2 \left(\frac{1-\pi_i}{\pi_i} \right) = M, \text{ say.} \qquad (2.1.2)$$

For t_g we seek a variance estimator m satisfying

$$\lim E_p E_m (m) = M. \qquad (2.1.3)$$

Treating the distribution of

$$d = (t_g - Y) / \sqrt{m}$$

for large n as close to that of τ or of Student's t with (n-1) degrees of freedom, we follow up the work in Chapter One to construct confidence intervals (CI) for Y in terms of (t_g, m). Failing to analytically discriminate among the CI's based on various m's and various t_g's changing with Q_i's, we resort to simulation studies to attempt at numerical comparisons among them. The findings are tabularly displayed later in the chapter. Encouraging competitiveness of some of the newly proposed ones against the traditional ones is well demonstrated there.

2.2 MODEL-CUM-DESIGN-BASED VARIANCE ESTIMATORS, CONFIDENCE INTERVALS AND THEIR ASSESSMENT.

To derive m satisfying (2.1.3) we consider the statistic

$$t(\alpha) = \sum{}' \alpha_i \left(\frac{y_i}{x_i} - \frac{1}{n} \sum{}' \frac{y_k}{x_k} \right)^2 ,$$

with α_i's as constants to be so chosen that $\lim E_p E_m [t(\alpha)]$ equals M.

Of course, as noted in section 1.2 we might use many others.

Noting $E_m[t(\alpha)] = \sum{}' \alpha_i \left[\frac{n-2}{n} \frac{\sigma_i^2}{x_i} + \frac{1}{n^2} \sum{}' \frac{\sigma_k^2}{x_k^2} \right]$ it is possible to choose the following sets of such α_i's, namely,

$$\alpha_i(1) = \frac{n}{n-2}\left[\frac{x_i^2}{\pi_i^2}(1-\pi_i) - \frac{1}{n(n-1)}\sum \frac{x_k^2}{\pi_k}(1-\pi_k)\right] \text{ and}$$

$$\alpha_i(2) = \frac{n}{n-2}\left[\frac{x_i^2}{\pi_i^2}(1-\pi_i) - \frac{1}{n(n-1)}\sum{}' \frac{x_k^2}{\pi_k^2}(1-\pi_k)\right]$$

yielding two alternative forms of m as

$$m_1 = \sum{}' \alpha_i(1)\left(\frac{y_i}{x_i} - \frac{1}{n}\sum{}' \frac{y_k}{x_k}\right)^2 \text{ and}$$

$$m_2 = \sum{}' \alpha_i(2)\left(\frac{y_i}{x_i} - \frac{1}{n}\sum{}' \frac{y_k}{x_k}\right)^2.$$ Two more alternative choices of m subject to (2.1.3) also follow as

$$m_3 = \frac{E_p\Sigma'\alpha_i(1)}{\Sigma'\alpha_i(1)}\, m_1 = \frac{\frac{n-2}{n-1}\sum x_k^2 \frac{1-\pi_k}{\pi_k}}{\sum{}' x_i^2 \frac{1-\pi_i}{\pi_i^2} - \frac{1}{n-1}\sum x_k^2 \frac{1-\pi_k}{\pi_k}}\, m_1$$

and,
$$m_4 = \frac{E_p\Sigma'\alpha_i(2)}{\Sigma'\alpha_i(2)}\, m_2 = \frac{\sum x_k^2 \frac{1-\pi_k}{\pi_k}}{\sum{}' x_i^2 \frac{1-\pi_i}{\pi_i^2}}\, m_2.$$

For analytical simplicity next we restrict to the situations where

$$\sigma_i^2 = \sigma^2 f_i \text{ where } f_i = x_i^g \text{ with g in } [0,2], \qquad (2.2.1)$$

but otherwise unknown and $\sigma(>0)$ unknown. If f_i is arbitrarily assigned, then, one may note that

(i) $M = \sigma^2 \sum f_i \frac{1-\pi_i}{\pi_i}$, (ii) $t(1) = \sum{}' \left(\frac{y_i}{x_i} - \frac{1}{n}\sum{}' \frac{y_k}{x_k}\right)^2$,

(iii) $E_m[t(1)] = \sigma^2 \frac{n-1}{n}\sum{}' \frac{f_i}{x_i^2}$ and

(iv) $\lim E_p E_m[t(1)] = \sigma^2 \frac{n-1}{n}\sum \frac{f_i\pi_i}{x_i^2}$.

Hence result the following two more choices of m as

$$m_5 = \frac{\sum f_i \frac{1-\pi_i}{\pi_i}}{\frac{n-1}{n}\sum \frac{f_i \pi_i}{x_i^2}} \sum' \left(\frac{y_i}{x_i} - \frac{1}{n}\sum' \frac{y_k}{x_k} \right)^2 \text{ and}$$

$$m_6 = \frac{\sum f_i \frac{1-\pi_i}{\pi_i}}{\frac{n-1}{n}\sum' \frac{f_i}{x_i^2}} \sum' \left(\frac{y_i}{x_i} - \frac{1}{n}\sum' \frac{y_k}{x_k} \right)^2 .$$

Kott's (1990a,b) variance estimators v_K corresponding to v as v_j, to be denoted as K_j, j=1,2, are available, though not generally, but only under (2.2.1) with f_i pre-assigned and may be seen to equal

$$K_1 = \frac{F\ v_1}{V_1} \quad \text{and,} \quad K_2 = \frac{F\ v_2}{V_2} \quad \text{where,}$$

$$F = \frac{1}{\sigma^2} E_m(t_g - Y)^2 = \sum' \left(\frac{g_{si}}{\pi_i} - 1 \right)^2 f_i + \Sigma_c f_i \ .$$

Here $\Sigma_c \equiv$ sum over i not in s,

$$V_1 = \frac{1}{\sigma^2} E_m(v_1) = \sum'\sum' \Delta_{ij} \left(\frac{f_i}{\pi_i^2} + \frac{f_j}{\pi_j^2} \right)$$

$$- \frac{2}{\sum' x_k^2 Q_k} \sum'\sum' \Delta_{ij} \left(\frac{x_i}{\pi_i} - \frac{x_j}{\pi_j} \right) \left(\frac{x_i f_i Q_i}{\pi_i} - \frac{x_j f_j Q_j}{\pi_j} \right)$$

$$+ \frac{\sum' x_i^2 f_i Q_i^2}{\left(\sum' x_k^2 Q_k \right)^2} \sum'\sum' \Delta_{ij} \left(\frac{x_i}{\pi_i} - \frac{x_j}{\pi_j} \right)^2$$

$$V_2 = \frac{1}{\sigma^2} E_m(v_2) = \sum'\sum' \Delta_{ij} \left(\frac{g_{si}^2 f_i}{\pi_i^2} + \frac{g_{sj}^2 f_j}{\pi_j^2} \right)$$

$$- \frac{2}{\sum' x_k^2 Q_k} \sum'\sum' \Delta_{ij} \left[\left(\frac{g_{si} x_i}{\pi_i} - \frac{g_{sj} x_j}{\pi_j} \right) \cdot \left(\frac{g_{si} x_i f_i Q_i}{\pi_i} - \frac{g_{sj} x_j f_j Q_j}{\pi_j} \right) \right]$$

$$+\frac{\sum' x_i^2 f_i Q_i^2}{\left(\sum' x_k^2 Q_k\right)^2} \sum\sum' \Delta_{ij}\left(\frac{g_{si}x_i}{\pi_i} - \frac{g_{sj}x_j}{\pi_j}\right)^2.$$

In our simulation studies that follow we illustrate only four choices of Q_i namely equal to $(1-\pi_i)/\pi_i x_i$, $1/\pi_i X_i$ respectively adopted by Brewer (1979) and Hájek (1971), $1/x_i^2$ and $1/x_i$. Corresponding t_g will be denoted respectively as t_B, t_H, t_S and $t_{S'}$.

The simulation study here is similar to that in section 3 of chapter one, the central interest shifting from $\bar{t}$ to t_g, everything else remaining same.

Since for the calculations of m_5, m_6, K_1 and K_2 one has to fix $f_i = x_i^g$ i.e. know the value of g, it is of interest to allow a chosen g say, g_0 to be different from the true g in $\sigma_i^2 = \sigma^2 x_i^g$ of the model $\underline{M}(g)$ and examine the consequences. For this we calculate CI with various g_0 in [0,2] and if the characteristics above remain more or less stable then we regard the procedures as robust; to extend this study of robustness we allow a non-zero intercept term θ in (1.1.1) and in that case denote the model by $\underline{M}_\theta$. Further, regarding

$$(1)\sum' \frac{x_i}{\pi_i} \quad \text{and} \quad (2)\sum' \frac{x_i}{\pi_i} \Big/ \sum' \frac{1}{\pi_i}$$

as ancillaries it is of interest to see how the CI's behave across samples with the values of (1) and (2) respectively fixed at certain levels. Numerical findings are not reported for (2).

Numerical findings are summarized in tables below in Appendix-C. We give ACP for τ and also for t_{31} with $\alpha = 0.05$, the latter within parentheses.

2.3 CONCLUDING REMARKS.

When a model (1.1.1) seems plausible so that one may legitimately employ a greg predictor to estimate Y it seems useful to reckon with a variance estimator that also takes consideration of the model. Thus model-assisted variance estimators m_1, m_2 in particular turn out quite effective competitors against the traditional ones namely v_j, K_j (j=1,2) even irrespective of the assignment of Q_i's. If one intends to employ a more restrictive variance estimator like K_1, K_2, m_5, m_6 that require preassigned g_0's, there is not much risk of mis-specification — procedures remain rather robust. If there underlies, however, a non-zero intercept term in the model which is unsuspected to begin with the situation is not so secure. If we consider group-wise comparison to take account of a reasonable ancillary statistic, even then model-based variance estimators like m_1, m_2, m_3, m_4 remain quite competitive against v_1 and even better than v_2, K_1, K_2 though the same cannot be said about m_5.

For further comments we refer to v_j, $K_j (j = 1, 2)$ as 'traditional' and the other variance estimators as 'new'. From the comments at the bottom of each table below one may note that the balance of relative advantages favours the 'new' rather than the 'traditional' varieties of variance estimators irrespective of choice of Q_i's.

Appendix B

Summary of numerical findings by simulation.

The tables B.1-B.5 presented below use the following abbreviations explained on p-15. ACP, ACV, PB, PS, PL, B(d), M(d), $\sqrt{\beta_1}$, $\beta_2 - 3$ relate respectively to coverage probability, coefficient of variation; bias, stability of variance estimator, length of CI; bias, MSE, 'root beta one' and 'excess measure' of the standardized statistic $d = (e - Y)/\sqrt{v}$.

Table B. 1

Performances of (e,v) by several criteria, under $\underline{M}$. Especially good (bad) values are under-scored (starred). $\beta = 1.0$, g=1.1, h=1.6, N=150, n=32, R=1000 for the model $\underline{M}$ of (1.1.1), p-5.

e v	10^4PCV	ACP	10^5ACV	-10^2PB	PS	PL
t_B v_1	4192	93.8(94.8)	4258	.54	.42	.98
t_B v_2	4268	93.8(94.7)	4261	.31	.43	.98
t_B K_1	4239	93.8(94.6)	4258	.49	.42	.98
t_B K_2	4241	93.8(94.6)	4258	.48	.42	.98
t_B m_1	4150	93.8(94.8)	4255	.75*	.41	.98
t_B m_2	4151	93.8(94.8)	4255	.75	.41	.98
t_B m_3	4178	93.8(94.7)	4256	.66	.42	.98
t_B m_4	4178	93.8(94.7)	4256	.66	.42	.98
t_B m_5	4282	94.0(95.2)	4280*	.58	.43	.98
t_B m_6	4390*	93.9(94.9)	4281*	.77*	.44	.98
t_H v_1	4190	93.8(94.7)	4257	.58	.42	.98
t_H v_2	4263	93.8(94.6)	4260	.36	.42	.98
t_H K_1	4236	93.8(94.6)	4258	.49	.42	.98
t_H K_2	4236	93.8(94.6)	4258	.49	.42	.98
t_H m_1	4150	93.8(94.8)	4255	.75*	.41	.98
t_H m_2	4151	93.8(94.8)	4255	.75*	.41	.98
t_H m_3	4178	93.8(94.8)	4256	.66	.41	.98
t_H m_4	4178	93.8(94.8)	4256	.66	.42	.98
t_H m_5	4282	94.0(95.2)	4280*	.59	.43	.98
t_H m_6	4290*	93.0(94.8)	4280*	.77*	.44	.98
t_S v_1	4192	93.8(94.8)	4258	.55	.42	.98
t_S v_2	4268	93.8(94.7)	4261	.31	.43	.98
t_S K_1	4240	93.8(94.6)	4258	.49	.42	.98
t_S K_2	4241	93.8(94.6)	4258	.48	.42	.98
t_S m_1	4150	93.9(94.8)	4255	.75*	.41	.98
t_S m_2	4151	93.8(94.8)	4255	.75*	.41	.98
t_S m_3	4178	93.8(94.7)	4256	.66	.42	.98

Table B.1 (continued)

e v	10^4PCV	ACP	10^5ACV	-10^2PB	PS	PL
t_S m_4	4178	93.8(94.7)	4256	.66	.42	.98
t_S m_5	4282	94.0(95.2)	4280*	.58	.43	.98
t_S m_6	4390*	93.9(94.9)	4280*	.76*	.44	.98
$t_{S'}$ v_1	4188	93.8(94.7)	4255	.67	.42	.98
$t_{S'}$ v_2	4247	93.8(94.5)	4257	.50	.42	.98
$t_{S'}$ K_1	4228	93.8(94.6)	4258	.48	.42	.98
$t_{S'}$ K_2	4221	93.8(94.6)	4258	.49	.42	.98
$t_{S'}$ m_1	4150	93.8(94.8)	4255	.73	.41	.98
$t_{S'}$ m_2	4151	93.8(94.9)	4256	.72	.41	.98
$t_{S'}$ m_3	4178	93.8(94.8)	4256	.63	.42	.98
$t_{S'}$ m_4	4178	93.8(94.8)	4256	.63	.42	.98
$t_{S'}$ m_5	4282	94.0(95.2)	4280	.61	.43	.98
$t_{S'}$ m_6	4390*	93.9(94.8)	4280	.79*	.44	.98

Table B.1 (continued)

e v	10^5 B(d)	10^3 M(d)	10^4 $\sqrt{\beta_1}$(d)	10^3 E(d)	e v	10^5 B(d)	10^3 M(d)	10^4 $\sqrt{\beta_1}$(d)	10^3 E(d)
t_B v_1	638	1093	815	57	t_H v_1	646	1094	814	58
t_B v_2	629	1093	824	51	t_H v_2	637	1094	823	52
t_B K_1	635	1094	819	53	t_H K_1	643	1094	818	53
t_B K_2	635	1094	819	52	t_H K_2	643	1094	818	53
t_B m_1	661	1094	802	61	t_H m_1	665	1094	802	62
t_B m_2	661	1094	802	61	t_H m_2	665	1094	802	62
t_B m_3	662	1094	802	58	t_H m_3	665	1094	802	58
t_B m_4	662	1094	802	58	t_H m_4	666	1094	802	58
t_B m_5	286	1089	946*	106*	t_H m_5	290	1089	946*	106*
t_B m_6	304	1089	936*	78	t_H m_6	308	1089	936*	78
t_S v_1	640	1093	815	58	$t_{S'}/v_1$	666	1095	810	59
t_S v_2	630	1093	824	51	$t_{S'}/v_2$	660	1096	819	54
t_S K_1	636	1094	819	53	$t_{S'}/K_1$	663	1095	812	55
t_S K_2	636	1094	819	53	$t_{S'}/K_2$	666	1095	813	56

Table [illegible] (continued)

e v	10^5 B(d)	10^3 M(d)	10^4 $\sqrt{\beta_1}$(d)	10^3 E(d)	e v	10^5 B(d)	10^3 M(d)	10^4 $\sqrt{\beta_1}$(d)	10^3 E(d)
t_S m_1	662	1094	819	62	t_S/m_1	673	1094	813	62
t_S m_2	662	1094	803	62	t_S/m_2	673	1094	<u>802</u>	62
t_S m_3	663	1094	<u>802</u>	58	t_S/m_3	675	1094	<u>802</u>	58
t_S m_4	663	1094	<u>802</u>	58	t_S/m_4	675	1094	<u>802</u>	58
t_S m_5	<u>287</u>	<u>1089</u>	946*	106*	t_S/m_5	<u>297</u>	<u>1089</u>	946*	106*
t_S m_6	<u>305</u>	<u>1089</u>	936*	78	t_S/m_6	<u>317</u>	<u>1089</u>	936*	78

Comments: For every Q_i in terms of criteria ACP, ACV PB, PS, PL the procedures compete keenly but by criterion PCV each $m_j (j = 1, \ldots, 4)$ is better than the traditional ones among which v_1 is the best, but m_5, m_6 are poorer. But m_5, m_6 are best by B(d), M(d) criteria for each Q_i.

Table B.2

Robustness of CI's by some criteria, under $\underline{M}_0$. β=1.0, g=1.2, h=1.9, N=150, n=32, R=1000. Consecutive values for $f_0 = x^{g_0}$ given for g_0=.4,.8,1.2. ACP values for τ and t_{31} separated by slashes.

e v	10^4 PCV	ACP	10^5 ACV
t_B K_1	4397, 4416, 4438	93.1/94.1, 93.1/93.9, 93.0/93.8	4799, 4797, 4796
t_B K_2	4397, 4417, 4438	93.1/94.1, 93.1.93.9, 93.0/93.8	4799, 4797, 4796
t_B m_5	4055, 4067, 4106	93.6/94.4, 93.5/94.5, 93.1/94.4	4825, 4797, 4772
t_B m_6	4235, 4235, 4235	93.1/94.2, 93.1/94.0, 92.8/93.9	4807, 4788, 4768
t_H K_1	4395, 4415, 4437	93.1/94.1, 93.1/93.9, 93.0/93.8	4799, 4797, 4796
t_H K_2	4395, 4415, 4437	93.1/94.1, 93.1/93.9, 93.0/93.8	4799, 4797, 4796
t_H m_5	4055, 4067, 4106	93.6/94.4, 93.5/94.4, 93.1/94.4	4825, 4797, 4772
t_H m_6	4235, 4235, 4235	93.1/94.2, 93.1/94.0, 92.8/93.9	4807, 4788, 4768
t_S K_1	4396, 4415, 4437	93.1/94.1, 93.1/93.9, 93.0/93.8	4799, 4797, 4796
t_S K_2	4396, 4415, 4437	93.1/94.1, 93.1/93.9, 93.0/93.8	4799, 4797, 4796
t_S m_5	4055, 4067, 4106	93.6/94.4, 93.5/94.5, 93.1/94.4	4825, 4797, 4772
t_S m_6	4235, 4235, 4235	93.1/94.2, 93.1/94.0, 92.8/93.9	4807, 4788, 4768
t_S/K_1	4391, 4411, 4434	93.1/94.1, 93.1/94.1, 93.1/93.8	4799, 4797, 4796
t_S/K_2	4390, 4410, 4431	93.1/94.1, 93.1/94.1, 93.1/93.8	4799, 4797, 4796
t_S/m_5	4055, 4067, 4106	93.6/94.4, 93.6/94.4, 93.1/94.4	4825, 4797, 4771
t_S/m_6	4235, 4235, 4235	93.1/94.2, 93.1/94.2, 92.8/93.9	4807, 4788, 4768

Table B.2 (continued)

e v	−10PB	PS	PL
t_B K_1	.69,.70,.70	.42,.42,.42	.94,.94,.94
t_B K_2	.69,.70,.70	.42,.42,.42	.94,.94,.94
t_B m_5	.63,.74,.84	.39,.38,.39	.95,.94,.94
t_B m_6	.68,.76,.83	.40,.40,.40	.95,.94,.94
t_H K_1	.69,.70,.70	.41,.42,.42	.94,.94,.94
t_H K_2	.69,.70,.70	.41,.42,.42	.94,.94,.94
t_H m_5	.63,.74,.84	.39,.38,.39	.95,.94,.94
t_H m_6	.68,.76,.83	.40,.40,.40	.95,.94,.94

Table B.2 (continued)

e v	−10PB	PS	PL
t_S K_1	.69,.70,.70	.41,.42,.42	.94,.94,.94
t_S K_2	.69,.70,.70	.41,.42,.42	.94,.94,.94
t_S m_5	.63,.74,.84	.38,.38,.39	.95,.94,.94
t_S m_6	.68,.76,.83	.40,.40,.40	.95,.94,.94
t_S/K_1	.69,.70,.70	.41,.42,.42	.94,.94,.94
t_S/K_2	.69,.70,.70	.41,.42,.42	.94,.94,.94
t_S/m_5	.63,.74,.83	.39,.38,.39	.95,.94,.94
t_S/m_6	.68,.76,.83	.40,.40,.40	.95,.94,.94

Table B.2 (continued)

e v	$10^4 B(d)$	$10^3 M(d)$	$10^3 \sqrt{\beta}$ (d)	$10^3 E(d)$
t_B K_1	171,173,174	115,115,115	122,114,105	221,220,219
t_B K_2	171,173,175	115,115,115	122,114,105	222,220,210
t_B m_5	125,131,137	113,114,116	324,298,266	271,248,231
t_B m_6	149,150,150	114,115,116	194,194,194	210,210,210
t_H K_1	171,173,174	115,115,115	122,114,105	222,220,219
t_H K_2	171,172,174	115,115,115	122,114,105	222,220,219

Table B.2 (continued)

e v	10^4B(d)	10^3M(d)	$10^3\sqrt{\beta\ (d)}$	10^3E(d)
t_H m_5	125, 130, 137	113, 114, 116	323, 298, 265	271, 248, 231
t_H m_6	149, 150, 150	114, 115, 116	193, 193, 193	210, 210, 210
t_S K_1	171, 172, 174	115, 115, 115	122, 114, 102	222, 220, 219
t_S K_2	171, 172, 174	115, 115, 115	122, 114, 105	222, 220, 219
t_S m_5	125, 130, 137	113, 115, 116	323, 298, 266	271, 248, 231
t_S m_6	149, 150, 150	114, 115, 116	193, 193, 193	210, 210, 210
t_S/K_1	170, 172, 174	115, 115, 115	132, 115, 106	223, 221, 220
t_S/K_2	170, 171, 173	115, 115, 115	124, 115, 107	223, 221, 220
t_S/m_5	125, 130, 137	113, 115, 116	321, 115, 263	271, 248, 231
t_S/m_6	149, 149, 150	114, 115, 116	191, 115, 191	211, 211, 211

Comments: In respect of PCV only m_6 is robust and there is little variation with respect to Q_i. Every procedure is robust in terms of ACP. In terms of ACV the traditional estimators are better and they are only robust. By other criteria m_5, m_6 are preferable. Variation with Q_i is negligible.

Table B.3

Performance of (e, v) by d-criterion.
β=1.0, g=1.1, h=1.6, G=10, R=1000, N=150, n=32.

e	v_1	v_2	K_1	K_2	m_1	m_2	m_3	m_4	m_5	m_6
t_B	3.582	3.659	3.615	3.616	3.600	3.599	3.578	3.578	3.985*	3.610
t_H	3.582	3.661	3.618	3.618	3.596	3.595	3.578	3.578	3.969	3.610
t_S	3.583	3.660	3.616	3.616	3.600	3.599	3.579	3.579	3.984*	3.610
$t_S/$	3.585	3.668	3.626	3.623	3.584	3.584	3.577	3.577	3.918	3.612

Comment: The best performers are m_3, m_4, the worst is m_5 and among the traditional ones v_1 is the best.

Table B.4

Robustness of CI's under $\underline{M}_\theta$ by several criteria, β=1.0, g=1.3, h=1.7, θ=5.0 and 10.0, R=1000, N=150, n=32. Values for θ=5.0, 10.0 given consecutively. ACP for τ and t_{31} separated by slashes.

e v	10^4PCV	ACP	10^5ACV	-10PB
t_B v_1	2993, 2648	92.3/93.6, 91.9/92.8	4054, 3955	.50, .34
t_B v_2	3071, 2739	92.3/93.4, 91.7/92.8	4056, 3956	.48, .32
t_B K_1	3042, 2709	92.3/93.6, 91.8/92.9	4054, 3955	.49, .33
t_B K_2	3044, 2711	92.3/93.6, 91.8/92.8	4054, 3955	.49, .33

Table B.4 (continued)

e v	10^4PCV	ACP	10^5ACV	-10PB
t_B m_1	2946,2586	92.1/93.6,91.9/92.7	4049,3944	.53,.41
t_B m_2	2947,2587	92.1/93.6,91.9/92.7	4049,3944	.53,.41
t_B m_3	2965,2616	92.1/93.6,91.9/92.6	4049,3945	.53,.40
t_B m_4	2965,2616	92.1/93.6,91.9/92.6	4049,3945	.53,.40
t_B m_5	2843,2444	92.2/93.7,92.1/93.8	4072,3983	.44,.24
t_B m_6	3003,2641	92.3/93.8,91.9/93.1	4073,3985	.41,.20
t_H v_1	2996,2641	92.3/93.6,91.9/92.8	4050,3946	.50,.35
t_H v_2	3069,2728	92.3/93.4,91.7/92.8	4052,3947	.48,.33
t_H K_1	3043,2700	92.3/93.6,91.8/92.8	4051,3946	.49,.34
t_H K_2	3043,2700	92.3/93.6,91.8/92.8	4051,3946	.49,.34
t_H m_1	2946,2586	92.1/93.6,91.9/92.9	4049,3944	.52,.37
t_H m_2	2947,2587	92.1/93.6,91.9/92.9	4049,3944	.52,.37
t_H m_3	2965,2616	92.1/93.6,91.9/92.8	4049,3945	.51,.36
t_H m_4	2965,2616	92.1/93.6,91.9/92.8	4049,3945	.51,.36
t_H m_5	2863,2440	92.3/93.7,92.3/92.8	4072,3984	.41,.21
t_H m_6	3003,2641	92.3/93.8,91.9/92.1	4073,3985	.40,.16
t_S v_1	2993,2648	92.3/93.6,91.9/92.8	4053,3952	.50,.34
t_S v_2	3071,2793	92.3/93.4,91.7/92.8	4054,3953	.48,.32
t_S K_1	3042,2709	92.3/93.6,91.8/92.9	4053,3952	.49,.33
t_S K_2	3044,2711	92.3/93.6,91.8/92.8	4053,3952	.49,.33
t_S m_1	2946,2586	92.1/93.6,91.9/92.8	4049,3944	.53,.40
t_S m_2	2947,2587	92.1/93.6,91.9/92.8	4049,3944	.53,.40
t_S m_3	2965,2616	92.1/93.6,91.9/92.6	4049,3945	.52,.39
t_S m_4	2965,2616	92.1/93.6,91.9/92.6	4049,3945	.52,.39
t_S m_5	2863,2440	92.2/93.7,92.1/93.8	4072,3983	.43,.23
t_S m_6	3003,2641	92.3/93.8,91.9/93.1	4073,3985	.41,.19
t_S/v_1	2996,2641	92.3/93.6,91.7/92.6	4037,3916	.52,.37
t_S/v_2	3069,2728	92.3/93.6,91.6/92.7	4038,3918	.50,.35
t_S/K_1	3043,2700	92.4/93.7,91.7/92.7	4039,3919	.50,.35
t_S/K_2	3043,2700	92.4/93.7,91.7/92.8	4039,3919	.50,.35
t_S/m_1	2946,2586	92.1/93.6,91.9/93.1	4049,3944	.47,.24
t_S/m_2	2947,2587	92.1/93.6,91.9/93.1	4049,3944	.47,.24

Table B.4 (continued)

e v	10^4PCV	ACP	10^5ACV	-10PB
t_S/m_3	2965, 2616	92.2/93.6, 91.9/93.1	4050, 3945	.46, .23
t_S/m_4	2965, 2616	92.2/93.6, 91.9/93.1	4050, 3945	.46, .23
t_S/m_5	2863, 2440	92.4/93.9, 92.5/94.0	4072, 3984	.38, .28
t_S/m_6	3003, 2641	92.4/93.9, 92.0/93.4	4073, 3985	.35, .24

Table B.4 (continued)

e v	PS	PL	-10B(d)	10^3M(d)	$-\sqrt{\beta}$ (d)	E(d)
t_B v_1	.29, .26	.96, .97	.69/1.02	115/117	.19/.36	.12/.06
t_B v_2	.30, .27	.96, .97	.73/1.09	115/117	.21/.39	.11/.09
t_B K_1	.29, .26	.96, .97	.72/1.07	115/117	.21/.38	.12/.08
t_B K_2	.29, .26	.96, .97	.72/1.07	115/117	.21/.38	.12/.08
t_B m_1	.28, .25	.96, .97	.66/ .97	115/117	.18/.33	.12/.03
t_B m_2	.28, .25	.96, .97	.66/ .97	115/117	.18/.33	.12/.03
t_B m_3	.29, .25	.96, .97	.67/ .99	115/117	.18/.34	.12/.04
t_B m_4	.29, .25	.96, .97	.67/ .99	115/117	.18/.34	.12/.04
t_B m_5	.28, .24	.97, .98	.55/ .82	114/114	.13/.25	.13/.01
t_B m_6	.29, .26	.97, .98	.66/ .98	114/115	.18/.34	.12/.05
t_H v_1	.29, .26	.96, .97	.69/1.02	115/117	.19/.36	.12/.06
t_H v_2	.30, .27	.96, .97	.73/1.08	116/117	.21/.39	.11/.09
t_H K_1	.29, .26	.96, .97	.72/1.06	115/117	.21/.38	.12/.08
t_H K_2	.29, .26	.96, .97	.72/1.06	115/117	.21/.38	.12/.08
t_H m_1	.28, .25	.96, .97	.66/ .97	115/117	.18/.33	.12/.04
t_H m_2	.28, .25	.96, .97	.66/ .98	115/117	.18/.33	.12/.04
t_H m_3	.29, .25	.96, .97	.67/ .99	115/117	.18/.34	.12/.04
t_H m_4	.29, .25	.96, .97	.67/ .99	115/117	.18/.34	.12/.04
t_H m_5	.28, .24	.97, .98	.55/ .82	114/114	.13/.26	.12/.01
t_H m_6	.29, .26	.97, .98	.66/ .98	114/115	.18/.34	.12/.05
t_S v_1	.29, .26	.96, .97	.69/1.02	115/117	.19/.36	.12/.06
t_S v_2	.30, .29	.96, .97	.73/1.09	116/117	.21/.39	.11/.09
t_S K_1	.29, .26	.96, .97	.72/1.06	115/117	.21/.38	.12/.08

Table B.4 (continued)

e v	PS	PL	-10B(d)	10^3M(d)	$-\sqrt{\beta}$ (d)	E(d)
t_S K_2	.29,.26	.96,.97	.72/1.07	115/117	.21/.38	.12/.08
t_S m_1	.28,.25	.96,.97	.66/ .97	115/117	.18/.33	.12/.03
t_S m_2	.28,.25	.96,.97	.66/ .97	115/117	.18/.33	.12/.03
t_S m_3	.29,.25	.96,.98	.67/ .99	115/117	.18/.34	.12/.04
t_S m_4	.29,.25	.96,.97	.67/1.00	115/117	.18/.34	.12/.04
t_S m_5	.28,.24	.97,.98	.55/ .82	114/114	.13/.26	.12/.01
t_S m_6	.29,.26	.97,.98	.66/ .98	114/115	.18/.34	.12/.05
t_S/v_1	.29,.25	.96,.97	.69/1.02	116/117	.19/.35	.12/.06
t_S/v_2	.30,.26	.96,.97	.72/1.07	116/117	.21/.38	.11/.08
t_S/K_1	.29,.26	.96,.97	.71/1.05	116/117	.20/.37	.11/.08
t_S/K_2	.29,.26	.96,.97	.71/1.05	116/117	.20/.37	.11/.07
t_S/m_1	.28,.25	.97,.98	.67/1.05	115/115	.18/.34	.12/.04
t_S/m_2	.28,.25	.97,.98	.67/ .98	115/115	.18/.34	.12/.05
t_S/m_3	.29,.26	.97,.98	.68/1.00	115/115	.19/.35	.12/.05
t_S/m_4	.29,.26	.97,.98	.68/1.00	115/115	.19/.35	.12/.05
t_S/m_5	.28,.24	.97,.99	.56/ .83	113/112	.13/.26	.12/.04
t_S/m_6	.29,.26	.97,.99	.67/ .99	114/113	.18/.35	.12/.06

Comments: Performance of every procedure is clearly affected by variation in θ in respect of each criterion except possibly PS and PL.

Table B.5

Conditional performances of the CI's under $\underline{M}_0$ with ancillary $\Sigma' x_i/\pi_i$. β=1.0, g=1.1, h=1.6, R=1000, Number of groups=10, N=150, n=32. For (t_g, v_g), (ACP, 10^5ACV, 10^4PCV) values given for successive groups

t_B v_1	93,4723,4416	97,4343,3577	96,4339,3831	93,4251,4586	91,4232,3470
	93,4295,4595	92,4259,3976	94,4133,3788	96,4188,4039	93,3888,3920
t_B v_2	93,4821,4451	97,4385,3576	96,4382,3836	93,4264,4572	91,4233,3471
	93,4296,4593	92,4239,4012	94,4092,3782	96,4079,4037	93,3815,3953
t_B K_1	93,4784,7735	97,4369,3575	96,4365,3831	93,4258,4575	91,4230,3472
	93,4294,4591	92,4244,4002	94,4105,3786	96,4691,4041	93,3839,3947

Table B.5 (continued)

t_B K_2	93, 4786, 4441	97, 4369, 3577	96, 4355, 3835	93, 4258, 4576	91, 4230, 3472
	93, 4294, 4591	92, 4244, 4002	94, 4104, 3786	96, 4090, 4041	93, 3838, 3947
t_B m_1	93, 4670, 4401	97, 4319, 3574	96, 4313, 3826	93, 4242, 4587	91, 4228, 3465
	93, 4293, 4584	92, 4267, 3950	94, 4153, 3776	96, 4137, 4030	93, 3930, 3889
t_B m_2	93, 4671, 4401	97, 4320, 3575	96, 4313, 3826	93, 4242, 4587	91, 4558, 3465
	93, 4293, 4584	92, 4267, 3951	94, 4152, 3776	96, 4137, 4030	93, 3929, 3890
t_B m_3	93, 4712, 4413	97, 4338, 3574	96, 4332, 3825	93, 4248, 4578	91, 4229 3469
	93, 4293, 4578	92, 4259, 3969	94, 4135, 3773	96, 4120, 4034	93, 3897, 3910
t_B m_4	93, 4712, 4413	97, 4338, 3574	96, 4332, 3825	93, 4248, 4578	91, 4229, 3469
	93, 4293, 4578	92, 4259, 3969	94, 4135, 3773	96, 4120 4034	93, 3897, 3910
t_B m_5	91, 4540, 4531	96, 4259, 3760	96, 4258, 4030	92, 4242, 4875	91, 4248, 3660
	94, 4331, 4862	94, 4329, 4108	94, 4255, 4054	96, 4242, 4229	96, 4097, 3995
t_B m_6	93, 4750, 4577	97, 4357, 3751	96, 4357, 4020	93, 4272, 4817	91, 4250, 3673
	94, 4327, 4826	92, 4281, 4199	94, 4156, 4031	96, 4143, 4236	93, 3908, 4095
t_H v_1	93, 4722, 4466	97, 4343, 3576	96, 4338, 3831	93, 4250, 4586	91, 4231, 3469
	93, 4295, 4593	92, 4258, 3976	94, 4132, 3787	96, 4118, 4037	93, 3888, 3918
t_H v_2	93, 4816, 4444	97, 4383, 3572	96, 4380, 3832	93, 4263, 4571	91, 4232, 3470
	93, 4295, 4591	92, 4239, 4012	94, 4093, 3786	96, 4080, 4040	93, 3817, 3956
t_H K_1	93, 4782, 4434	97, 4368, 3574	96, 4364, 3831	93, 4258, 4575	91, 4233, 3472
	93, 4294, 4589	92, 4245, 4008	94, 4106, 3785	96, 4092, 4040	93, 3841, 3944
t_H K_2	93, 4782, 4434	97, 4368, 3574	96, 4364, 3831	93, 4258, 4575	91, 4230, 3472
	93, 4294, 4589	92, 4245, 4001	94, 4106, 3785	96, 4092, 4040	93, 3841, 3944
t_H m_1	93, 4670, 4401	97, 4319, 3574	96, 4313, 3826	93, 4242, 4587	91, 4228, 3465
	93, 4293, 4584	92, 4267, 3950	94, 4153, 3777	96, 4138, 4030	93, 3930, 3890
t_H m_2	93, 4671, 4404	97, 4320, 3575	96, 4313, 3826	93, 4242, 4587	91, 4228, 3465
	93, 4293, 4584	92, 4267, 3951	94, 4152, 3776	96, 4137, 4030	93, 3929, 3890
t_H m_3	93, 4712, 4413	97, 4338, 3574	96, 4332, 3825	93, 4248, 4576	91, 4229, 3469
	93, 4293, 4578	92, 4259, 3969	94, 4135, 3773	96, 4120, 4034	93, 3897, 3910
t_H m_4	93, 4712, 4413	97, 4338, 3574	96, 4332, 3825	93, 4248, 4578	91, 4229, 3469
	93, 4293, 4578	92, 4259, 3969	94, 4135, 3773	96, 4120, 4034	93, 3897, 3910

Table B.5 (continued)

t_H m_5	91,4540,4531	96,4259,3760	96,4258,4030	92,4242,4875	91,4248,3660
	94,4331,4862	94,4329,4108	94,4255,4254	96,4242,4229	96,4097,3995
t_H m_6	93,4750,4577	97,4357,3751	96,4357,4020	93,4272,4817	91,4250,3673
	94,4327,4826	92,4281,4199	94,4156,4031	96,4243,4236	93,3908 4095
t_S v_1	93,4723,4416	97,4343,3577	96,4339,3832	93,4251,4586	91,4323,3470
	93,4296,4595	92,4259,3977	94,4133,3788	96,4118,4039	93,3889,3920
t_S v_2	93,4821,4451	97,4385,3576	96,4382,3836	93,4264,4573	91,4232 3471
	93,4296,4593	92,4239,4012	94,4092,3782	96,4079,4037	93,3815,3953
t_S K_1	93,4784,4435	97,4369,3575	96,4365,3831	93,4258,4575	91,4230,3472
	93,4294,4591	92,4244,4002	94,4105,3786	96,4091,4041	93,3839,3947
t_S K_2	93,4786,4441	97,4369,3578	96,4365,3835	93,4258,4277	91,4230,3472
	93,4294,4591	92,4244,4001	94,4104,3781	96,4091,4038	93,3839,3941
t_S m_1	93,4670,4401	97,4319,3574	96,4313,3826	93,4242,4587	91,4228,3465
	93,4293,4584	92,4267,3950	94,4153,3776	96,4137,4030	94,3930,3890
t_S m_2	93,4671,4401	97,4320,3575	96,4313,3826	93,4242,4587	91,4228,3465
	93,4293,4584	92,4267,3951	94,4152,3776	96,4237,4030	93,3929,3890
t_S m_3	93,4712,4413	97,4338,3574	96,4332,3825	93,4248,4578	91,4229,3469
	93,4293,4578	92,4259,3969	94,4135,3773	96,4120,4034	93,3897,3910
t_S m_4	93,4712,4423	97,4338,3574	96,4332,3825	93,4248,4878	91,4229,3469
	93,4293,4578	92,4259,3969	94,4135,3773	96,4120,4030	93,3897,3910
t_S m_5	91,4540,4531	96,4259,3760	96,4258,4030	92,4242,4875	91,4248,3660
	94,4331,4862	94,4329,4108	94,4255,4054	96,4242,4229	96,4097,3995
t_S m_6	93,4750,4577	97,4357,3751	96,4357,4020	93,4272,1817	91,4259,3673
	94,4327,4826	92,4281,4199	94,4156,4031	96,4143,4236	93,3908,4095
$t_{S'}$ v_1	93,4719,4414	97,4341,3571	96,4335,3829	93,4248,4583	91,4229,3468
	93,4292,4588	92,4256,3976	94,4129,3785	96,4116,4034	93,3887,3910
$t_{S'}$ v_2	93,4798,4418	97,4375,3558	96,4372,3818	93,4259,4267	91,4229,3468
	93,4292,4587	92,4239,4012	94,4096,3799	96,4083,4048	93,3824,3963
$t_{S'}$ K_1	93,4773,4430	97,4365,3570	96,4361,3829	93,4275,4574	91,4231,3470
	93,4293,4584	92,4247,3998	94,4109,3783	96,4095,4037	93,3847,3935

Table B.5 (continued)

t_S/K_2	93, 4767, 4409	97, 4364, 3560	96, 4359, 3817	93, 4256, 4571	91, 4231, 3469
	93, 4293, 4585	92, 4247, 4002	94, 4111, 3797	96, 4098, 4048	96, 6850, 3452
t_S/m_1	93, 4669, 4401	97, 4319, 3574	96, 4313, 3826	93, 4241, 4587	91, 4228, 3465
	93, 4293, 4584	92, 4267, 3950	94, 4153, 3776	96, 4138, 4030	93, 3930, 3890
t_S/m_2	93, 4671, 4401	97, 4320, 3575	96, 4313, 3826	93, 4242, 4587	91, 4228, 3465
	93, 4293, 4584	92, 4267, 3951	94, 4152, 3776	96, 4137, 4030	93, 3929, 3890
t_S/m_3	93, 4711, 4413	97, 4338, 3574	96, 4332, 3825	93, 4248, 4578	91, 4229, 3469
	93, 4293, 4578	92, 4259, 3969	94, 4135, 3773	96, 4120, 4034	93, 3898, 3910
t_S/m_4	93, 4711, 4413	97, 4338, 3574	96, 4332, 3825	93, 4248, 4578	91, 4229, 3469
	93, 4293, 4578	92, 4259, 3969	94, 4135, 3773	96, 4120, 4034	93, 3898, 3910
t_S/m_5	91, 4540, 4531	96, 4259, 3760	96, 4258, 4030	92, 4242, 4875	91, 4248, 3660
	94, 4331, 4862	94, 4329, 4108	94, 4255, 4054	96, 4242, 4229	96, 4098, 3995
t_S/m_6	93, 4749, 4579	97, 4357, 3751	96, 4357, 4020	93, 4272, 4817	91, 4049, 3673
	94, 4327, 4826	92, 4281, 4199	94, 4156, 4031	96, 4144, 4236	93, 3909, 4095

Comments: Irrespective of Q_i, every procedure seems to be affected by changes in the ancillary statistic in respect of the three criteria chosen. But no clear pattern is discernible.

CHAPTER THREE

INTERVAL ESTIMATION BY RATIO ESTIMATOR AND MODEL-CUM-DESIGN-BASED VARIANCE ESTIMATORS

3.0 SUMMARY.

For the special case of super-population linear regression model with the model-variance proportional to the regressor variable the ratio estimator is known to be appropriate. With the ratio estimator as the point estimator, confidence intervals for Y are constructed deriving model-cum-asymptotic design-based variance estimators. The variance estimators themselves however are derived postulating the general model and not the above special case. Simulation studies are resorted to for comparing the confidence intervals. The newly emerged variance estimators are demonstrated to fare as good competitors against those well-known in the literature. We restrict to simple random sampling without replacement.

3.1 INTRODUCTION.

We consider the model $\underline{M}$ of (1.1.1) in Chapter One. Usually with $g = 1$, the ratio estimator is taken as the point estimator for Y, given by

$$t = X(\bar{y}/\bar{x})$$

Here $\bar{x}$, $\bar{y}$ are sample means of x, y. Various alternative variance estimators v of t are well-known, including those studied by Royall and Eberhardt (1975), Royall and Cumberland (1978a,b, 1981a,b, 1985), Cumberland and Royall (1988), Wu (1982), Wu and Deng (1983), Särndal (1982,1984), Särndal, Swensson and Wretman (1989,1992), Kott (1990a,b) among others, some of which are motivated by consideration of super-population modelling as pointed out above. We shall derive a few more alternative variance estimators for t. Here we shall consider both model (1.1.1) and an asymptotic approach of Brewer (1979) to hit upon a few more alternative variance etimators for t utilizing the model (1.1.1) of Chapter One and Brewer's asymptotic design-based approach explained in Chapter One. Since it is difficult to have a reasonable comparative evaluation, analytically, of these variance estimators and of associated CI's, we attempt at a numerical evaluation on taking observations through simulations. Details of theory are given in section 3.2, numerical findings are summarized by tables **in** Appendix-C at the end of this chapter and comments and remarks in section 3.3.

3.2 VARIANCE ESTIMATORS.

We throughout assume that the sample-size is large and the model $\underline{M}$ of Chapter One is tenable. For the ratio estimator t, well-known Cochran's (1977) approximate variance formula is

$$V_a = N^2 \frac{1-f}{n} \frac{1}{N-1} \Sigma (y_i - Rx_i)^2$$

admitting two well-known estimators

$$v_0 = N^2 \frac{1-f}{n} \frac{1}{n-1} \Sigma' (y_i - rx_i)^2$$

$$v_2 = \left(\frac{\bar{X}}{\bar{x}} \right)^2 v_0.$$

Here $R=Y/X$, $r=\bar{y}/\bar{x}$, $f=n/N$. Kott's (1990a,b) variance estimator is taken as

$$K(v) = \frac{E_m(t-Y)^2}{E_m(v)} v,$$

with v as v_j denoted by v_{Kj} $(j=0,2)$. Writing $\bar{x}_c$ as the mean of x's for units outside s, $s_x^2 = \Sigma' (x_i - \bar{x})^2/(n-1)$, $c_x^2 = s_x^2 / \bar{x}^2$ it may be noted that

$$E_m(t-Y)^2 = \sigma^2 \frac{N^2(1-f)}{n} \frac{\bar{X}\,\bar{x}_c}{\bar{x}}$$

$$E_m(v_0) = \sigma^2 \frac{N^2(1-f)}{n} \bar{x} \left(1- \frac{c_x^2}{n}\right) \text{ leading to}$$

$$v_{K0} = \frac{\bar{X}\,\bar{x}_c}{\bar{x}^2} \left(1- \frac{c_x^2}{n}\right)^{-1} v_0 = v_{K2}$$ which happens to coincide with the one earlier given by Royall and Eberhardt (1975), denoted by v_H.

Two other variance estimators for t available in the literature already mentioned are

$$v_D = \frac{N^2(1-f)}{n} \frac{\bar{X}\,\bar{x}_c}{\bar{x}^2} \frac{1}{n} \sum{}' \frac{(y_i - rx_i)^2}{(1-x_i/n\bar{x})^2}$$

and, $$v_J = \frac{N^2(1-f)}{n} (n-1)\, \bar{X}^2\, \Sigma'(d_i - \bar{d})^2,$$

the Jack-knife estimator, where $d_i = (n\bar{y}-y_i)/(n\bar{x}-x_i)$, i in s, $\bar{d} = \Sigma' d_i/n$.

To derive new variance estimators utilizing the model and adopting Brewer's (1979) asymptotic approach we proceed as follows.

First we consider estimating

$$E_m(V_a) = \sigma^2 \frac{N^2(1-f)}{n} \bar{X} \left(1- \frac{C_0^2}{N}\right) = M(x), \text{ say,}$$

where,

$$C_0^2 = S_x^2 / \bar{X}^2, \quad S_x^2 = \frac{1}{N-1} \Sigma(y_i - Rx_i)^2,$$

by a statistic v, say, for which

$$\lim E_p E_m(v) = M(x). \tag{3.2.1}$$

A few such alternative choices of v satisfying (3.2.1) follow as :

$$v_{01} = \frac{E_m(V_a)}{\lim E_p E_m(v_0)} v_0 = \frac{(1 - C_0^2/N)}{(1 - C_0^2/n)} v_0$$

noting that $\lim E_p E_m(v_0) = \sigma^2 \frac{N^2(1-f)}{n} \bar{X} \left(1- \frac{C_0^2}{n}\right)$,

$$v_{21} = \left(\frac{\bar{X}}{\bar{x}}\right)^2 v_{01}$$

$$v_{02} = \frac{E_m(V_a)}{E_m(v_0)} v_0 = \left(\frac{\bar{X}}{\bar{x}}\right) \frac{(1 - C_0^2 /N)}{(1 - c_x^2 /n)} v_0,$$

noting that $E_m(v_{02}-V_a)=0$.

Also, incidentally, one may note that,

$$v_{22} = \frac{E_m(V_a)}{E_m(v_a)} v_2 = v_{02},\ K(v_{01}) = K(v_{02}) = K(v_{22})$$

$$= K(v_2) = K(v_0),$$

so that Kott's (1990a,b) method does not yield any new variance estimator.

A second use of Brewer's (1979) approach is to first note that

$$\lim E_p E_m (t-Y)^2 = \sigma^2 \frac{N^2(1-f)}{n} \bar{X} = M'(x), \text{ say,}$$

and then seek a statistic v such that

$$\lim E_p E_m(v)=M'(x).$$

This approach easily yields the following alternatives, namely,

$$v_{03} = \frac{M'(x)}{\lim E_p E_m(v_0)} v_0 = \left(1 - \frac{C_0^2}{n}\right)^{-1} v_0$$

$$v_{23} = \frac{M'(x)}{\lim E_p E_m(v_2)} v_2 = \left(\frac{\bar{X}}{\bar{x}}\right)^2 v_{03}$$

$$v_{04} = \frac{M'(x)}{E_m(v_0)} v_0 = \left(\frac{\bar{X}}{\bar{x}}\right)\left(1 - \frac{c_x^2}{n}\right)^{-1} v_0,$$

on observing that

$$E_m(v_{04}-M'(x))=0.$$

Further one may note that

$$v_{24} = \frac{M'(x)}{E_m(v_2)} v_2 = v_{04}$$

and, $K(v_{03}) = K(v_{23}) = K(v_{04}) = K(v_0)$.

To find more alternatives we consider variance estimators of the form

$$t(\alpha) = \sum{}' \alpha_i \left(\frac{y_i}{x_i} - \frac{1}{n} \sum{}' \frac{y_k}{x_k} \right)^2,$$

with α_i's as assignable constants to be determined on solving

$$\lim E_p E_m(t(\alpha)) = M'(x). \qquad (3.2.2)$$

As mentioned in section 1.2 many other choices might be tried.

Two sets of α_i's that result from this turn out to be

$$\alpha_i(1) = \frac{N^2(1-f)}{n(n-2)} \left[x_i^2 - \frac{\Sigma x_k^2}{N(n-1)} \right]$$

and, $$\alpha_i(2) = \frac{N^2(1-f)}{n(n-2)} \left[x_i^2 - \frac{\Sigma' x_k^2}{n(n-1)} \right],$$

leading to the following four variance estimators :

$$m_1 = \frac{N^2(1-f)}{n(n-2)} \sum{}' \left[x_i^2 - \frac{\Sigma x_k^2}{N(n-1)} \right] \left(\frac{y_i}{x_i} - \frac{1}{n} \sum{}' \frac{y_k}{x_k} \right)^2$$

$$m_2 = \frac{N^2(1-f)}{n(n-2)} \sum{}' \left[x_i^2 - \frac{\Sigma' x_k^2}{n(n-1)} \right] \left(\frac{y_i}{x_i} - \frac{1}{n} \sum{}' \frac{y_k}{x_k} \right)^2$$

$$m_3 = \frac{E_p[\Sigma' \alpha_i(1)]}{\Sigma' \alpha_i(1)} m_1 = \frac{n-2}{n-1} \frac{\frac{1}{N}\Sigma x_k^2}{\frac{1}{n}\Sigma' x_k^2 - \frac{1}{n-1} \frac{1}{N}\Sigma x_k^2} m_1$$

$$m_4 = \frac{E_p[\Sigma' \alpha_i(2)]}{\Sigma' \alpha_i(2)} m_2 = \frac{n}{N} \frac{\Sigma x_k^2}{\Sigma' x_k^2} m_2.$$

Two more estimators satisfying (3.2.2) are

$$m_5 = \frac{N^2(1-f)\bar{X}}{n} \sum{}' \left(\frac{y_i}{x_i} - \frac{1}{n} \sum{}' \frac{y_k}{x_k} \right)^2 \Big/ \left(\frac{n-1}{N} \sum \frac{1}{x_i} \right)$$

$$m_6 = \frac{N^2(1-f)\bar{X}}{n} \sum{}' \left(\frac{y_i}{x_i} - \frac{1}{n} \sum{}' \frac{y_k}{x_k} \right)^2 \Big/ \left(\frac{n-1}{n} \sum{}' \frac{1}{x_i} \right).$$

Obviously, it is quite difficult to discriminate among so many alternative variance estimators purely on theoretical considerations, especially because many of them are proposed because of their asymptotic properties. So we resort to simulations to study performances of CI's based on t and these alternative variance estimators.

3.3 SIMULATION STUDY.

For simulations we proceed essentially as in Chapters One and Two. But we repeat the process to help the readership. We take N=150, draw x_i's as random samples from the density

$$f(x) = \frac{1}{\lambda} e^{-x/\lambda}, \quad x>0, \quad \lambda=8.5,$$

take $\sigma=1.0$, draw ε_i's randomly from the normal distribution N(0,1), take $\beta=1.0$, $g=1.0$, and $y_i=\beta x_i+\sigma x_i^{g/2}\varepsilon_i$, sample-size n is taken as 32 and α as 0.05. We use tables of both normal distribution and Student's distribution of t-statistic t_{n-1} with (n-1) degrees of freedom. Number of replicates F is taken as 1000. Sum over replication is denoted as Σ_r. We write $A = \frac{1}{F} \Sigma_r v$ and $P = \frac{1}{F} \Sigma_r (t-Y)^2$.

To discriminate among the CI's we consider the following criteria in accordance with usual practices, vide Rao and Wu (1983) :

I ACP (Actual coverage percentage) ≡ the percent of F replicates for which the CI covers Y — the closer it is to the nominal confidence coefficient .95, the better.

II ACV (Average coeffiecient of variation) ≡ the average of $\sqrt{v}/t$ over the replicates — this reflects the length of CI relative to t.

III Pseudo relative bias (PB) ≡ $PB(v) = \frac{1}{P}\left[\frac{1}{F} \Sigma_r (v-P)\right]$

IV Pseudo relative stability (PS) ≡ $PS(v) = \frac{1}{P}\left[\frac{1}{F} \Sigma_r (v-P)^2\right]^{1/2}$

V Pseudo standardized length (PL) ≡ $PL(v) = \left[\frac{1}{F} \Sigma_r \sqrt{v}\right] \Big/ \sqrt{P}$

VI Bias of d $\equiv B(d) = \frac{1}{F} \Sigma_r d$

VII Mean square error (MSE) of d $\equiv M(d) = \frac{1}{F} \Sigma_r (d-B(d))^2$

VIII Root beta one of d $\equiv \sqrt{\beta_1(d)} = \frac{1}{F} \Sigma_r \left[\frac{(d-B(d))}{\sqrt{M(d)}}\right]^3$

$\underline{\text{IX}}$ Excess $\equiv \beta_2(d)-3 = \frac{1}{F} \Sigma_r \left[\frac{(d-B(d))}{\sqrt{M(d)}}\right]^4 - 3$

$\underline{\text{X}}$ Pseudo coefficient of variation (PCV) $= \frac{1}{A}\left[\frac{1}{F} \Sigma_r (v-A)^2\right]^{1/2}$.

The smaller the magnitudes of II — $\underline{\text{X}}$ the better the pair (t,v). In some of the choices of v, knowledge of g_0 in

$$\sigma_i^2=\sigma^2 x_i^{g_0},\ i \text{ in } U,\ g_0 \text{ in } [0,2]$$

is required. But if the choice is wrong and true form of σ_i^2 is

$$\sigma_i^2=\sigma^2 x_i^{g},\ i \text{ in } U,\ g \text{ in } [0,2],\ g \neq g_0,$$

then the procedure may or may not remain good — if it remains good then the procedure is robust, otherwise not. To examine robustness by simulation we examine the above criteria allowing variation in g_0 around g. Further we also examine robustness allowing change in model $\underline{M}$ to $\underline{M}_\theta$, where for $\underline{M}_\theta$ everything else in $\underline{M}$ remains intact except that

$$y_i = \theta + \beta x_i + \sigma x_i^{g/2}\varepsilon_i,\ i\in U \text{ with } \theta \neq 0. \qquad (3.3.1)$$

Choosing such y_i's subject to (3.3.1) we examine the CI in terms of the above ten criteria. Finally we note that $\bar{x}$ may be regarded as an ancillary statistic and, to see how the CI's behave with variation in $\bar{x}$, we make a conditional study as mentioned in Chaudhuri and Stenger (1992). For this we divide the F = 1000 replicates into 10 equal groups of F_k=100, (k=1,...,10) sub-replicates taking the first group as one consisting of the replicates with the lowest 100 $\bar{x}$-values, the next group consists of those 100 replicates with the next higher $\bar{x}$'s and so on. Then we calculate CI's within respective groups and examine the above ten criteria group-wise. As an over-all measure of comparison we consider the new d-criterion, namely,

$$d = \left[\frac{1}{10} \sum_k \left(\sqrt{\frac{1}{F_k} \Sigma_{r_k} v} - \sqrt{\frac{1}{F_k} \Sigma_{r_k} (t-Y)^2} \right)^2 \right]^{1/2}$$

where Σ_k is sum over the groups and Σ_{r_k} the sum over the units in the k-th group of F_k replicates. Findings are summarized and tabulated in Appendix-C.

3.4 COMMENTS AND CONCLUSIONS.

In this chapter we have proposed 12 alternative variance estimators for the ratio estimator of a finite population total as possible competitors against 5 traditional ones. The former group we denote by $\underline{A}$ and the latter by $\underline{B}$. Our plan is to numerically compare the suitabilities of CI's based respectively on them when values are generated according to a postulated model that suits the use of a ratio estimator. From Table 1 we see that better performances are more in evidence when we use one from $\underline{A}$ rather than from $\underline{B}$, though very bad performances also follow with use of m_5 and m_6. From Table 2 we see that there is not much robustness in allowing variation in g_0 especially when it is very large. From Table 3 we see of course that there is robustness in respect of β for all criteria except for ACV but there is little robustness in respect of θ. So, if there is a surreptitious intercept term then it may not be wise to use ratio estimator at all and there is hardly any clue from the presented data about how to choose among procedures in $\underline{A}$ or in $\underline{B}$. But if blindly a ratio estimator is used even if $\theta \neq 0$, then for each fixed θ, better results are expected for those in $\underline{A}$. From Table 4 we first see that $\bar{x}$ in fact serves as a useful ancillary, performances showing appreciable changes across $\bar{x}$. Here also better results are discernible with the use of those in $\underline{A}$ though the best **d**-values are produced by those in $\underline{B}$. Taking everything into consideration we would rather recommend that m_1 and m_2 should serve as the variance estimators with the highest potentials, use of m_5 and m_6 may often become unsafe but the original Cochran's (1977) v_0 may yet be taken as one with enough strength to continue as a challenging competitor against every one else.

APPENDIX C

SUMMARY OF FINDINGS.

The abbreviated symbols ACP, ACV, PB, PS, PL, B(d), M(d), $\sqrt{\beta_1(d)}$, E(d) are as explained on pages 45-46, relate respectively to coverage probability, coefficient of variation, bias, stability of variance estimator, length of CI, bias, MSE, 'root beta one' and 'excess measure' of the standardized statistic $d = (e - Y)/\sqrt{v}$.

Table C.1

Performances of CI by several criteria : λ=8.5, β=1.0, σ=1.0, g=1.0, N=150, n=32, F=1000, α=0.05. ACP values for τ and t_{31} are separated by slashes. Especially good (bad) values are under-scored(starred).

v	ACP	10^5 ACV	10^4 PCV	10^2 PB	10^2 PS	10^2 PL	-10^3 B(d)	10^2 M(d)	-10 $\sqrt{\beta_1(d)}$	E(d)
v_0	93.7/94.7	3697	2922	384	31	101	.02	106	.27	.20
v_2	94.0/94.9	3701	3110	430	33	101	.15	105	.27	.19
v_H	94.2/95.2	3709	3171	485	34	101	.19	105	.28	.19
v_D	94.2/95.2	3704	3160	454	33	101	.76	105	.31	.19
v_J	94.0/95.0	3706	3088	457	33	101	1.30	105	.34	.19
v_{01}	93.8/94.7	3703	2922	417	31	101	.02	106	.27	.20
v_{21}	94.0/95.0	3707	3110	463	33	101	.15	105	.27	.19
v_{02}	94.0/94.7	3704	2956	425	31	101	.07	105	.27	.19
v_{03}	93.8/94.8	3704	2922	426	31	101	.02	106	.27	.20
v_{23}	94.0/95.0	3708	3110	472	33	101	.15	105	.27	.19
v_{04}	94.0/94.7	3705	2956	434	31	101	.07	105	.27	.19
m_1	93.8/94.9	3687	2903	331	30	101	1.65	107	.38	.20
m_2	93.9/94.9	3688	2900	333	30	101	1.63	106	.38	.20
m_3	94.3/94.7	3698	3092	413	32	101	1.75	105	.37	.19
m_4	94.3/94.8	3699	3100	416	33	101	1.74	105	.37	.19
m_5	95.3/96.1	3986	4321*	230	58*	108	3.49*	94	1.51*	.38*
m_6	95.4/96.2	3981	4371*	228	58*	108	3.45*	94	1.48*	.38*

Comments: The new procedures and the traditional ones have similar ACP values. Better ACV is yielded by $m_j (j = 1, \ldots, 4)$ and v_0 ; better PCV is realized by m_1, m_2, v_0 and v_{01}. Better PB is obtained for m_1, m_2, m_5, m_6 and v_0. Better B(d) is ensured by v_0, v_{01}, v_{03}. In respect of PCV, PS, B(d), $\sqrt{\beta_1(d)}$, performances of m_5 and m_6 are poor. The balance seems to favour the new procedures though v_0 competes well against them.

Table C.2

Robustness of CI's under $\underline{M}_0$. Model : $y_i = x_i + x_i^{g/2}\varepsilon_i$, ε_i is distributed as N(0,1). Values for g_0= .4, .8, 1.2, 1.6 given consecutively downwards. Values for τ and t_{31} are separated by slashes.

		10^5	10^4	10^4	10^2	10^2	-10^3	10^2	-10	
v	ACP	ACV	PCV	PB	PS	PL	B(d)	M(d)	$\sqrt{\beta_1}$(d)	E(d)
v_0	93.9/95.3	1853	3241	361	34	100	19	106	.54	.14
	93.8/94.9	2928	3010	377	31	101	6	106	.00	.18
	93.8/94.7	4681	2854	389	30	101	-6	106	-.56	.22
	94.0/94.8	7587	2782	395	29	101	-18	108	-1.14	.27
v_2	94.4/95.6	1857	3524	444	37	101	19	104	.56	.14
	94.1/95.1	2933	3235	436	30	101	6	105	.00	.12
	94.3/94.8	4685	3001	423	32	101	-6	105	-.56	.21
	94.4/94.9	7589	2835	401	30	101	-18	107	-1.14	.24
v_H	94.4/95.7	1861	3593	504	38	101	19	104	.56	.14
	94.2/95.2	2939	3299	493	35	101	6	104	.01	.17
	94.2/95.1	4696	3057	476	32	101	-6	105	-.56	.21
	94.3/94.9	7605	2881	451	30	101	-18	106	-1.14	.24
v_D	94.2/95.7	1857	3578	448	38	101	18	104	.53	.14
	94.1/95.1	2934	3286	453	35	101	5	105	-.02	.17
	94.2/95.0	4691	3048	453	32	101	-7	105	-.60	.21
	94.3/94.9	7603	2876	445	30	101	-18	106	-1.18	.24
v_J	4.1/95.7	1856	3493	422	37	101	18	104	.50	.14
	94.2/95.2	2934	3209	446	34	101	5	104	-.05	.17
	94.4/94.9	4695	2983	466	32	101	-7	105	-.63	.21
	94.4/94.9	7617	2826	479	30	101	-19	106	-1.21	.24
01	94.3/95.4	1856	3241	393	34	101	19	105	.54	.14
	93.9/94.9	2933	3010	410	32	101	6	105	.01	.18
	93.8/94.8	4689	2854	422	30	101	-6	106	-.56	.22
	94.0/94.9	7599	2782	427	29	101	-18	107	-1.14	.27
21	94.4/95.8	1860	3524	477	37	101	19	104	.56	.14
	94.1/95.1	2937	3235	469	34	101	6	104	.01	.17
	94.3/94.8	4693	3001	456	32	101	-6	105	-.56	.21

Table C.2 (continued)

v	ACP	10^5 ACV	10^4 PCV	10^2 PB	10^2 PS	10^2 PL	-10^3 B(d)	10^2 M(d)	-10 $\sqrt{\beta_1(d)}$	E(d)
	94.4/94.9	7601	2835	434	30	101	−18	106	−1.13	.24
v_{02}	94.2/95.4	1858	3329	420	35	101	19	104	.56	.13
	94.0/94.9	2934	3064	425	32	101	6	105	.01	.17
	94.0/94.5	4689	2866	424	30	101	−6	105	−.56	.21
	94.2/94.8	7598	2744	416	39	101	−18	107	−1.14	.25
v_{03}	94.3/95.4	1857	3241	402	34	101	19	105	.54	.14
	93.9/94.9	2934	3010	419	32	101	6	105	.01	.18
	93.8/94.8	4691	2854	431	30	101	−6	106	−.56	.22
	94.0/95.0	7602	2782	436	29	101	−18	107	−1.14	.27
v_{23}	94.5/95.8	1861	3524	486	37	101	19	104	.56	.14
	94.1/95.1	2938	3235	478	34	101	6	104	.01	.17
	94.3/94.8	4695	3001	464	32	101	−6	105	−.56	.21
	94.4/94.9	7604	2835	443	30	101	−18	106	−1.14	.24
v_{04}	94.2/95.4	1858	3329	429	35	101	19	104	.56	.13
	94.0/94.9	2935	3064	434	32	101	6	104	.01	.17
	94.0/94.5	4691	2866	433	30	101	−6	105	−.56	.21
	92.4/94.8	7601	2744	425	29	101	−18	106	−1.13	.25
m_1	93.9/95.4	1844	3200	249	33	100	17	106	.41	.12
	93.8/94.9	2918	2983	304	31	100	5	106	−.11	.18
	93.9/94.8	4674	2842	355	30	101	−8	107	−.66	.23
	93.9/94.9	7588	2783	399	29	101	−19	108	−1.23	.28
m_2	93.9/95.4	1844	3203	253	33	100	17	106	.42	.12
	93.9/95.0	2919	2982	307	31	100	5	106	−.10	.17
	93.9/94.8	4674	2837	352	30	101	−8	107	−.66	.23
	94.0/94.8	7589	2775	400	29	101	−19	108	−1.23	.27
m_3	94.2/95.8	1851	3484	367	36	100	17	104	.44	.12
	94.1/94.8	2928	3210	399	34	101	4	105	−.09	.17
	94.3/94.8	4686	2990	425	31	101	−8	105	−.66	.21
	94.4/94.8	7603	2835	442	30	101	−19	106	−1.23	.25
m_4	94.2/95.7	1851	3497	371	36	100	17	104	.44	.12
	94.1/94.9	2928	3219	402	24	101	4	105	−.09	.17
	94.3/94.8	4686	2996	427	32	101	−8	105	−.66	.21

Table C.2 (continued)

v	ACP	10^5 ACV	10^4 PCV	10^2 PB	10^2 PS	10^2 PL	-10^3 B(d)	10^2 M(d)	-10 $\sqrt{\beta_1(d)}$	E(d)
	94.4/94.8	7603	2838	443	30	101	-19	106	-1.23	.25
m_5	96.0/97.1	2129	4854	404	79	115	49	84	.22	.40
	95.7/96.4	3226	4499	287	65	111	39	91	1.74	.39
	95.2/95.7	4938	4145	174	52	106	30	98	1.28	.37
	94.6/95.3	7653	3807	668	41	101	20	108	.81	.35
m_6	96.4/97.3	2128	4922	404	80	115	48	84	.22	.41
	95.8/96.7	3223	4556	286	65	111	39	91	1.71	.39
	95.1/95.8	4931	4187	172	52	106	30	98	1.25	.37
	94.7/95.1	7641	3827	636	41	101	19	107	.77	.35

Comments: ACP's are good throughout with little variation for differences in g_0. The ACV's increase and PCV's decrease along with g_0 and so do respectively PB and PS. The procedures $m_j (j = 1, \ldots, 4)$ outperform the traditional ones except v_0.

Table C.3

Robustness of CI under $\underline{M}_\theta$. Model : $y_1 = \theta + \beta x_1 + \sigma x_1^{g/2}$. $\lambda=8.5$, $g=1.0$, $\sigma=1.0$, $\alpha=0.05$. Values for ($\theta=0$, $\beta=1$), ($\theta=2.5$, $\beta=1$), ($\theta=5$, $\beta=1$), ($\theta=2.5$, $\beta=2$) and ($\theta=5$, $\beta=2$) respectively successively downwards. ACP for τ and t_{31} separated by slashes.

v	ACP	10^5 ACV	10^4 PCV	10^2 PB	10^2 PS	10^2 PL	-10^3 B(d)	10^2 M(d)	-10 $\sqrt{\beta_1(d)}$	E(d)
v_0	93.7/94.7	3697	2922	384	31	101	.02	106	.27	.20
	93.2/94.5	3146	2482	162	25	100	-2.95	107	-.57	.06
	93.5/94.4	3063	2156	6	22	99	-3.76	107	-.86	.02
	93.2/94.5	1734	2482	162	25	100	-2.95	107	-.56	.06
	93.5/94.4	1821	2156	6	22	99	-3.75	107	-.86	.02
v_2	94.0/94.9	3701	3110	430	33	101	.15	105	.27	.19
	93.6/94.4	3149	2707	216	28	100	17.45	106	.51	.07
	93.4/94.8	3065	2456	76	25	100	31.17	107	1.03	.01
	93.6/94.4	1735	2707	216	28	100	17.45	106	.51	.07
	93.4/94.8	1823	2456	76	25	100	31.18	107	1.03	.01
v_H	94.2/95.2	3709	3171	485	34	101	.19	105	.28	.19
	93.6/94.2	3156	2776	271	29	100	20.22	106	.66	.07
	93.4/94.8	3072	2537	133	26	100	35.90	107	1.29	.02

Table C.3 (continued)

v	ACP	10^5 ACV	10^4 PCV	10^2 PB	10^2 PS	10^2 PL	-10^3 B(d)	10^2 M(d)	-10 $\sqrt{\beta_1(d)}$	E(d)
	93.6/94.2	1739	2776	271	29	100	20.22	106	.66	.07
	93.4/94.8	1827	2537	133	26	100	35.91	107	1.29	.02
v_D	94.2/95.2	3704	3160	454	33	101	.76	105	.31	.19
	93.6/94.3	3152	2770	244	28	100	20.73	106	.68	.07
	93.4/94.8	3069	2543	117	26	100	36.36	107	1.31	.02
	93.6/94.3	1737	2770	244	28	100	20.74	106	.68	.07
	93.4/94.8	1825	2543	117	26	100	36.37	107	1.31	.02
v_J	94.0/95.0	3706	3088	457	33	101	1.29	105	.34	.19
	93.6/94.4	3155	2697	250	28	100	18.46	106	.56	.07
	93.4/94.9	3073	2471	131	25	100	32.05	106	1.07	.01
	93.6/94.4	1738	2697	250	28	100	18.46	106	.56	.07
	93.4/94.9	1827	2471	131	25	100	32.06	106	1.07	.01
v_{01}	93.8/94.7	3703	2922	417	31	101	.02	106	.27	.20
	93.3/94.5	3151	2489	194	25	100	2.94	106	-.57	.06
	93.5/94.4	3067	2156	38	22	100	3.75	107	-.86	.02
	93.3/94.5	1736	2482	194	25	100	2.94	106	-.57	.06
	93.5/94.4	1823	2156	38	22	100	3.74	107	-.86	.02
v_{21}	94.0/95.0	3707	3110	463	33	101	.15	105	.27	.19
	93.6/94.4	3154	2707	248	28	100	17.43	106	.51	.07
	93.4/94.9	3070	2456	108	25	100	31.12	106	1.03	.01
	93.6/94.4	1728	2707	248	28	100	17.43	106	.51	.07
	93.4/94.9	1826	2456	108	25	100	31.13	106	1.03	.01
v_{02}	94.0/94.7	3704	2956	425	31	101	.07	105	.27	.19
	93.0/94.5	3152	2529	207	26	100	7.27	106	-.30	.05
	93.5/94.8	3068	2239	59	23	100	13.75	106	.09	-.01
	93.0/94.5	1737	2529	207	26	100	7.28	106	-.30	.05
	93.5/94.8	1824	2239	59	23	100	13.76	106	.09	-.01
v_{03}	93.8/94.8	3704	2922	426	31	101	.02	106	.27	.20
	93.3/94.5	3153	2482	202	25	100	-2.94	106	-.57	.06
	93.5/94.5	3069	2156	46	22	100	-3.75	107	-.86	.02
	93.3/94.5	1737	2482	202	25	100	-2.94	106	-.57	.06
	93.5/94.5	1824	2156	46	22	100	-3.74	107	-.86	.02

Table C.3 (continued)

v	ACP	10^5 ACV	10^4 PCV	10^2 PB	10^2 PS	10^2 PL	-10^3 B(d)	10^2 M(d)	-10 $\sqrt{\beta_1(d)}$	E(d)
v_{23}	94.0/95.0	3708	3110	472	33	101	.15	105	.27	.19
	93.6/94.4	3155	2707	257	28	100	17.42	106	.51	.07
	93.4/94.9	3071	2456	116	25	100	31.11	106	1.03	.01
	93.6/94.4	1739	2707	257	28	100	17.42	106	.51	.07
	93.4/94.9	1826	2456	116	25	100	31.11	106	1.03	.01
v_{04}	94.0/94.7	3705	2956	434	31	101	.07	105	.27	.19
	93.0/94.5	3153	2529	215	26	100	7.27	106	-.03	.05
	93.5/94.8	3069	2239	68	23	100	13.75	106	.09	-.01
	93.0/94.5	1738	2529	215	26	100	7.27	106	-.03	.05
	93.5/94.8	1825	2239	68	23	100	13.76	106	.09	-.01
m_1	93.8/94.8	3687	2903	330	30	101	1.65	107	.38	.20
	93.9/94.8	3218	2427	623	27	102	2.32	102	-.36	.08
	94.6/95.9	3295	2369	1611	32	107	1.55	94	-.61	.08
	93.9/94.8	1773	2427	623	27	102	2.32	102	-.36	.08
	94.6/95.9	1959	2369	1611	32	107	1.56	94	-.61	.08
m_2	93.9/94.9	3688	2900	333	30	101	1.63	106	.38	.20
	93.9/94.8	3218	2423	626	26	102	3.25	102	-.31	.08
	94.6/95.9	3295	2366	1611	32	107	3.02	94	-.52	.07
	93.9/94.8	1773	2423	626	26	102	3.25	102	-.31	.08
	94.6/95.9	1959	2366	1611	32	107	3.03	94	-.52	.07
m_3	94.3/94.7	3698	3092	413	32	101	1.75	105	.37	.19
	94.0/95.2	3226	2628	713	29	103	22.11	101	.73	.08
	95.3/96.0	3302	2563	1711	35	107	33.94	93	1.32	.06
	94.0/95.2	1778	2628	713	29	103	22.11	101	.73	.08
	95.3/96.0	1964	2563	1711	35	107	33.95	93	1.32	.06
m_4	94.3/94.8	3699	3100	416	33	101	1.74	105	.37	.19
	94.0/95.1	3226	2635	715	29	103	22.39	101	.75	.08
	95.3/96.0	3302	2568	1712	35	107	34.35	93	1.35	.06
	94.0/95.1	1778	2635	715	29	103	22.39	101	.75	.08
	95.3/96.0	1964	2568	1712	35	107	34.35	93	1.35	.06
m_5	95.3/96.1	3985	4321	2300	58	108	34.91	94	-1.51	.38
	95.1/96.6	3445	3384	2301	48	109	-28.11	91	-1.81	.07
	95.9/97.3	3401	2472	2366	39	110	-20.57	87	-1.74	-.01

Table C.3 (continued)

v	ACP	10^5 ACV	10^4 PCV	10^2 PB	10^2 PS	10^2 PL	-10^3 B(d)	10^2 M(d)	$-10\ \sqrt{\beta_1(d)}$	E(d)
	95.1/96.6	1897	3384	2301	48	109	-28.11	91	-1.81	.07
	95.9/97.3	2021	2472	2366	39	110	-20.56	87	-1.74	-.01
m_6	95.4/96.2	3981	4371	2283	58	108	34.55	94	-1.48	.38
	95.2/96.3	3440	3476	2295	49	109	-18.57	91	-1.26	.08
	96.1/96.5	3397	2625	2371	40	110	-4.90	88	-.79	-.02
	95.2/96.3	1895	3476	2295	49	109	-18.57	91	-1.26	.08
	96.1/96.5	2019	2625	2371	40	110	-4.90	88	-.79	-.02

Comments: ACP remains good throughout and best for m_5, m_6 showing little variation with changing parameters. Since only non-negative θ is illustrated ACV's naturally decrease with positive θ without showing a pattern. Fluctuations are also pronounced in respect of other criteria. The procedures $m_1, \ldots, m_4$ seem to outperform the traditional ones except possibly v_1.

Table C.4

Conditional comparison of CI under $\underline{M}_0$

N=150, n=32, λ=8.5, g=1.0, α=0.05, β=1.0, F=1000, Ancillary : $\bar{x}$.

Normal ACP, 10^5ACV and 10^4PCV values given downwards in succession.

v	groups 1	2	3	4	5	6	7	8	9	10	10^4 d-value
v_0	94	97	94	86	94	94	92	93	95	98	6153
	3704	3624	3609	3608	3663	3730	3706	3763	3818	3742	
	3073	3393	2786	3212	2880	2726	2925	2812	2675	2545	
v_2	96	98	95	86	94	94	92	93	94	99	4909
	4110	3845	3740	3682	3688	3702	3621	3618	3603	3400	
	3097	3397	2786	3201	2884	2733	2920	2816	2668	2597	
v_H	97	99	95	86	94	94	92	93	94	98	4921
	4174	3883	3766	3700	3699	3707	3617	3606	3580	3359	
	3106	3396	2785	3199	2885	2733	2919	2817	2667	2612	
v_D	97	99	95	86	94	94	92	93	94	98	4893
	4168	3877	3760	3696	3694	3701	3612	3599	3576	3355	
	3096	3384	2776	3187	2871	2726	2908	2805	2651	2597	
v_J	96	98	95	86	94	94	92	93	94	98	4926
	4117	3849	3744	3690	3694	3707	3625	3621	3608	3405	
	3076	3370	2766	3176	2855	2716	2897	2793	2636	2569	

Table C.4 (continued)

v	groups 1	2	3	4	5	6	7	8	9	10	10^4 d-value
v_{01}	94	97	95	86	94	94	92	93	95	98	6175
	3710	3630	3615	3614	3669	3736	3712	3769	3824	3748	
	3073	3393	2786	3212	2880	2726	2925	2812	2675	2545	
v_{21}	96	98	95	86	94	94	92	93	94	98	4937
	4117	3851	3746	3688	3694	3708	3627	3624	3608	3405	
	3097	3397	2785	3201	2884	2733	2920	2816	2668	2597	
v_{02}	95	98	95	86	94	94	92	93	95	98	5323
	3908	3739	3681	3651	3682	3722	3669	3696	3714	3572	
	3071	3393	2785	3206	2881	2727	2922	2814	2670	2563	
v_{03}	94	97	95	86	94	94	92	93	95	98	6181
	3711	3631	3616	3616	3670	3737	3714	3770	3825	3750	
	3073	3393	2786	3212	2880	2726	2925	2812	2675	2545	
v_{23}	96	98	95	86	94	94	92	93	94	98	4944
	4119	3853	3747	3690	3696	3710	3628	3625	3610	3406	
	3097	3397	2785	3201	2884	2733	2920	2816	2669	2597	
v_{04}	95	98	95	86	94	94	92	93	95	98	5330
	3910	3741	3682	3653	3683	3724	3671	3697	3716	3574	
	3071	3393	2785	3206	2881	2727	2922	2814	2670	2563	
m_1	93	97	95	86	94	94	93	93	95	98	6225
	3675	3601	3592	3597	3657	3721	3698	3758	3825	3750	
	3073	3367	2728	3211	2874	2719	2891	2784	2604	2545	
m_2	94	97	95	86	94	94	93	93	95	98	6131
	3695	3612	3599	3601	3659	3720	3695	3751	3815	3735	
	3079	3372	2731	3212	2876	2721	2889	2781	2599	2533	
m_3	97	98	95	86	95	94	92	93	95	98	4957
	4093	3827	3720	3675	3691	3697	3621	3620	3620	3418	
	3124	3421	2753	3210	2900	2782	2889	2788	2622	2575	
m_4	97	98	95	86	95	94	92	93	95	98	4954
	4100	3832	3722	3676	3692	3697	3620	3617	3817	3413	
	3132	3425	2755	3212	2901	2783	2888	2785	2617	2565	

Table C.4 (continued)

v	groups										10^4 d-value
	1	2	3	4	5	6	7	8	9	10	
m_5	96	99	95	89	95	96	92	95	97	99	8473
	4101	4042	4009	3846	3932	4005	3992	4083	4003	3846	
	4131	4517	4027	4551	4233	3862	4461	4282	4534	4389	
m_6	96	99	95	90	95	96	92	95	97	99	8027
	4295	4152	4078	3880	3937	3986	3942	4000	3881	3659	
	4104	4527	4028	4556	4209	3836	4472	4279	4512	4393	

Comments: It is well-known that v_H, v_D, v_J which approximate v_2 are better in tracking the conditional variance given x than is v_0. This is confirmed with the d-values for the former turning out less than that for v_0. Among the new procedures $m_1, m_2, m_5, m_6, v_{01}, v_{03}$ are also poor like v_0 in this respect while the others appear to compete well with v_H, v_D, v_J. Since the intercept term is absent, the ACP's turn out good throughout. ACV and PCV values show considerable fluctuations without displaying any clear pattern. In these two respects the new procedures m_1, m_2 and the traditional v_0 seem to fare better than the others.

CHAPTER FOUR

CONFIDENCE INTERVAL ESTIMATION FOR DOMAIN TOTALS IN COMPLEX SURVEYS

4.0 SUMMARY.

We consider the population divisible into non-overlapping domains of known sizes and every unit assignable on inspection to the domain to which it belongs. The problem is to estimate the respective domain totals on drawing a sample from the entire population. Again an auxiliary variable with known values is available to which the variable of interest bears a super-population linear regression relation through the origin. Since the regression may vary across the domains separate generalized regression predictors may be appropriate for respective domain totals. For them we consider traditional variance estimators and also derive new alternatives using linear regression models and applying Brewer's asymptotic design-based approach. Postulating again that the domains of differing sizes may be alike to the extent of permitting the slopes to be identical in the regression model alternative greg predictors and corresponding variance estimators are also considered. The latter that borrow strength across the domains are really 'synthetic' versions in contrast with the former which may be called 'non-synthetic'. Confidence intervals for domain totals are then constructed as usual. Analytic comparison among them is rather impracticable. So, we resort to simulations for a numerical evaluation. The non-synthetic approach here should naturally be infructuous for many domains with small sizes because sample-sizes for them should also be quite small in practice. So, only the 'synthetic' approach seems reasonable unless the 'assumption' of a common regression slope is grossly untenable. For simulations we postulate a 'common' slope and compare the above two sets of confidence intervals employing both the synthetic and non-synthetic "estimators, variance estimators" combinations. In

addition we also consider composite estimators combining these two types of greg predictors and deriving their variance estimators.

4.1 INTRODUCTION.

Suppose the population $U=(1,\ldots,i,\ldots,N)$ is divisible into a number D of disjoint domains U_d of sizes N_d, $\sum_{d=1}^{D} N_d=N$. For y, x let the domain totals be Y_d, X_d, $d=1,\ldots,D$. We persist with the model $\underline{M}$ of earlier chapters taking

$$y_i = \beta x_i + \varepsilon_i, \quad i\in U, \qquad (4.1.1)$$

and consider the problem to estimate Y_d, $d=1,\ldots,D$, on surveying a sample s of size n taken from U with probability p(s), admitting as usual positive inclusion probabilities π_i, π_{ij} of the first two orders. We shall need the indicator variable I_{di} valued 1 for i in U_d and 0, else, for simplified analysis. In case β is permitted to vary across the domains, say, taken as β_d for i in U_d then $\underline{M}$ will be written as $\underline{M}_d$, and $\underline{M}(f)$ as $\underline{M}_d(f)$.

Unassisted by model $\underline{M}$ a traditional estimator for Y_d is the direct Horvitz-Thompson's (HT, 1952) estimator

$$\bar{t} = \sum{}' \frac{y_i}{\pi_i} I_{di} ,$$

to be denoted in tables by HTE, admitting Yates and Grundy's (YG, 1953) variance estimator, to be denoted in tables by YGE,

$$v_{YG} = \sum{}'\sum{}' \Delta_{ij}\left(\frac{y_i I_{di}}{\pi_i} - \frac{y_j I_{dj}}{\pi_j} \right)^2$$

writing,

$$\Delta_{ij} = (\pi_i\pi_j-\pi_{ij})/\pi_{ij}.$$

With Q_i's (>0) as assignable constants let,

$$\hat{\beta}_{Qd} = \frac{\Sigma' y_i x_i Q_i I_{di}}{\Sigma' x_i^2 Q_i I_{di}}, \quad e_{di} = y_i - x_i \hat{\beta}_{Qd},$$

$$\hat{\beta}_Q = \frac{\Sigma' y_i x_i Q_i}{\Sigma' x_i^2 Q_i}, \quad e_i = y_i - x_i \hat{\beta}_Q.$$

Following Särndal (1980, 1992), assisted by faith in model $\underline{M}_d$ one gets for Y_d the greg predictor

$$t_{gd}(1) = X_d \hat{\beta}_{Qd} + \sum{}' \frac{e_{di} I_{di}}{\pi_i}$$

$$= \sum{}' \frac{y_i I_{di}}{\pi_i} + \left(X_d - \sum{}' \frac{x_i I_{di}}{\pi_i} \right) \hat{\beta}_{Qd}$$

$$= \sum{}' \frac{y_i}{\pi_i} \left[I_{di} + \left(X_d - \sum{}' \frac{x_k I_{dk}}{\pi_k} \right) \frac{x_i Q_i \pi_i I_{di}}{\Sigma' x_k^2 Q_k I_{dk}} \right]$$

$$= \sum{}' \frac{y_i}{\pi_i} g_{sdi}$$

writing,

$$g_{sdi} = I_{di} + \left(X_d - \sum{}' \frac{x_k I_{dk}}{\pi_k} \right) \frac{x_i Q_i \pi_i I_{di}}{\Sigma' x_k^2 Q_k I_{dk}}.$$

Following Särndal (1982) two variance estimators for $t_1 = t_{gd}(1)$ follow as

$$v_2(t_1) = \sum{}' \sum{}' \Delta_{ij} \left(\frac{g_{sdi} e_{di} I_{di}}{\pi_i} - \frac{g_{sdj} e_{dj} I_{dj}}{\pi_j} \right)^2$$

and $v_1(t_1)$ from $v_2(t_1)$ by putting $g_{sdi} = I_{di}$ in it, to be denoted respectively by Tay2 and Tay.

Motivated by faith in $\underline{M}$ one gets the alternative greg predictor, 'borrowing' strength across the domains, as

$$t_2 = t_{gd}(2) = X_d\hat{\beta}_Q + \sum{}' \frac{e_i I_{di}}{\pi_i}$$

$$= \sum{}' \frac{y_i I_{di}}{\pi_i} + \left(X_d - \sum{}' \frac{x_i I_{di}}{\pi_i} \right) \hat{\beta}_Q$$

$$= \sum{}' \frac{y_i}{\pi_i} \left[I_{di} + \left(X_d - \sum{}' \frac{x_k I_{dk}}{\pi_k} \right) \frac{x_i Q_i \pi_i}{\Sigma' x_k^2 Q_k} \right]$$

$$= \sum{}' \frac{y_i}{\pi_i} g'_{sdi}$$

writing,

$$g'_{sdi} = I_{di} + \left(X_d - \sum{}' \frac{x_k I_{dk}}{\pi_k} \right) \frac{x_i Q_i \pi_i}{\Sigma' x_k^2 Q_k} .$$

It is reasonable to call t_1 a 'non-synthetic' greg predictor and t_2 a 'synthetic' greg prdictor. For t_2, Särndal's two variance estimators follow as

$$v_2(t_2) = \sum{}' \sum{}' \Delta_{ij} \left(\frac{g'_{sdi} e_i I_{di}}{\pi_i} - \frac{g'_{sdj} e_j I_{dj}}{\pi_j} \right)^2$$

and $v_1(t_2)$ follows from $v_2(t_2)$ replacing g'_{sdi}'s by I_{di}'s.

Kott's (1990 a,b) variance estimators for t_1 follow under $\underline{M}_d(f)$ as

$$K_j(t_1) = \frac{E_m(t_1 - Y_d)^2}{E_m[v_j(t_1)]} v_j(t_1) , \; j=1,2,$$

respectively to be denoted in tables by KT and KT2, and for t_2 under $\underline{M}(f)$ as

$$K_j(t_2) = \frac{E_m(t_2 - Y_d)^2}{E_m[v_j(t_2)]} v_j(t_2) , \; j=1,2,$$

again to be denoted in tables by KT and KT2 respectively.

Under $\underline{M}_d(f)$ we work out $E_m(t_1-Y_d)^2$, $E_m[v_j(t_1)]$ and under $\underline{M}(f)$ we work out $E_m(t_2-Y_d)^2$, $E_m[v_j(t_2)]$, j=1,2 as follows

$$\frac{1}{\sigma^2} E_m(t_1-Y_d)^2 = \sum{}' \left(\frac{1}{\pi_i} - 1 + c_{s1}Q_i x_i \right)^2 f_i I_{di}$$

$$+ \sum f_i I_{di} - \sum{}' f_i I_{di}$$

where,

$$c_{s1} = (X_d - \Sigma' x_i I_{di}/\pi_i) / (\Sigma' Q_i x_i^2 I_{di}) ,$$

$$\frac{1}{\sigma^2} E_m[v_2(t_1)]$$

$$= \sum{}'\sum{}' \Delta_{ij}\left(\frac{g_{sdi}^2 f_i I_{di}}{\pi_i^2} + \frac{g_{sdj}^2 f_j I_{dj}}{\pi_j^2} \right)$$

$$+ \frac{\Sigma' Q_i^2 x_i^2 f_i I_{di}}{(\Sigma' Q_i x_i^2 I_{di})^2} \sum{}'\sum{}' \Delta_{ij}\left(\frac{g_{sdi} x_i I_{di}}{\pi_i} - \frac{g_{sdj} x_j I_{dj}}{\pi_j} \right)^2$$

$$- \frac{2}{(\Sigma' Q_i x_i^2 I_{di})} \sum{}'\sum{}' \Delta_{ij}\left(\frac{g_{sdi} x_i I_{di}}{\pi_i} - \frac{g_{sdj} x_j I_{dj}}{\pi_j} \right)$$

$$\cdot \left(\frac{g_{sdi} Q_i x_i f_i I_{di}}{\pi_i} - \frac{g_{sdi} Q_j x_j f_j I_{dj}}{\pi_j} \right)$$

and, $\frac{1}{\sigma^2} E_m[v_1(t_1)]$ follows from $\frac{1}{\sigma^2} E_m[v_2(t_1)]$ by replacing g_{sdi}'s by I_{di}'s.

Similarly,

$$\frac{1}{\sigma^2} E_m(t_2-Y_d)^2 = \sum{}' \left(I_{di}(\frac{1}{\pi_i} - 1) + c_{s2}Q_i x_i \right)^2 f_i I_{di}$$

$$+ \sum f_i I_{di} - \sum{}' f_i I_{di}$$

where,

$$c_{s2} = (X_d - \Sigma' x_i I_{di}/\pi_i) / (\Sigma' Q_i x_i^2)$$

and,

$$\frac{1}{\sigma^2} E_m[v_2(t_2)]$$

$$= \sum\nolimits'\sum\nolimits' \Delta_{ij}\left(\frac{g'^2_{sdi} f_i I_{di}}{\pi_i^2} + \frac{g'^2_{sdj} f_j I_{dj}}{\pi_j^2} \right)$$

$$+ \frac{\Sigma' Q_i^2 x_i^2 f_i}{(\Sigma' Q_i x_i^2)^2} \sum\nolimits'\sum\nolimits' \Delta_{ij}\left(\frac{g'_{sdi} x_i I_{di}}{\pi_i} - \frac{g'_{sdj} x_j I_{dj}}{\pi_j} \right)^2$$

$$- \frac{2}{(\Sigma' Q_i x_i^2)} \sum\nolimits'\sum\nolimits' \Delta_{ij}\left[\left(\frac{g'_{sdi} x_i I_{di}}{\pi_i} - \frac{g'_{sdj} x_j I_{dj}}{\pi_j} \right)\right.$$

$$\left. \cdot \left(\frac{g'_{sdi} Q_i x_i f_i I_{di}}{\pi_i} - \frac{g'_{sdi} Q_j x_j f_j I_{dj}}{\pi_j} \right)\right] ,$$

and, $\frac{1}{\sigma^2} E_m[v_1(t_2)]$ follows from $\frac{1}{\sigma^2} E_m[v_2(t_2)]$ on replacing g'_{sdi}'s by I_{di}'s.

In order to derive further variance estimators for t_j, $j=1,2$, we restrict only to models $\underline{M}(f)$ or $\underline{M}_d(f)$ and follow Brewer's (1979) asymptotic design - based approach. Noting $\lim E_p g_{sdi} = I_{di} = \lim E_p g'_{sdi}$, it follows that t_j ($j=1,2$) are both ADU for Y_d.

First, postulating either $\underline{M}_d(f)$ or $\underline{M}(f)$ we find

$$\lim E_p E_m (t_1 - Y_d)^2 = \sigma^2 \sum \left(\frac{1-\pi_i}{\pi_i} \right) f_i I_{di}$$

$$= \lim E_p E_m (t_2 - Y_d)^2 = M_d, \text{ say.}$$

So, for either t_1 or t_2 we deem it reasonable to seek a variance estimator m_d, say, which satisfies

$$\lim E_p E_m (m_d) = M_d. \tag{4.1.2}$$

For this, we consider assigning constants α_i in $t(\alpha) = \Sigma' \alpha_i (r_i - \bar{r})^2$, where $r_i = y_i/x_i$, $\bar{r} = \Sigma' r_i/n$, so as to equate $\lim E_p E_m t(\alpha) = M_d$. Two sets of α_i as $\alpha_i(1)$ and $\alpha_i(2)$ given below are available, on noting

$$t_2(\alpha) = \lim E_p E_m t(\alpha) = \sigma^2 \sum \left(\frac{n-2}{nx_i^2} \alpha_i + \frac{\Sigma \alpha_k \pi_k}{n^2 x_i^2} \right) \pi_i f_i ,$$

hence solving

$$\left(\frac{n-2}{nx_i^2} \alpha_i + \frac{\Sigma' \alpha_k \pi_k}{n^2 x_i^2} \right) \pi_i f_i = \frac{1-\pi_i}{\pi_i} I_{di} , \quad i \in U,$$

and noting

$$t_1(\alpha) = E_m t(\alpha) = \sigma^2 \sum{}' \left(\frac{n-2}{nx_i^2} \alpha_i + \frac{\Sigma' \alpha_k}{n^2 x_i^2} \right) f_i ,$$

hence solving

$$\left(\frac{n-2}{nx_i^2} \alpha_i + \frac{\Sigma' \alpha_k}{n^2 x_i^2} \right) f_i = \frac{1-\pi_i}{\pi_i^2} I_{di} , \quad i \in s, \text{ as}$$

$$\alpha_i(1) = \frac{n}{n-2} \left[\frac{x_i^2}{\pi_i} \left(\frac{1}{\pi_i} - 1 \right) I_{di} - \frac{1}{n(n-1)} \sum x_k^2 \left(\frac{1}{\pi_k} - 1 \right) I_{dk} \right]$$

and,

$$\alpha_i(2) = \frac{n}{n-2} \left[\frac{x_i^2}{\pi_i} \left(\frac{1}{\pi_i} - 1 \right) I_{di} - \frac{1}{n(n-1)} \sum{}' \frac{x_k^2}{\pi_k} \left(\frac{1}{\pi_k} - 1 \right) I_{dk} \right].$$

Keeping $\underline{M}(f)$ or $\underline{M}_d(f)$ and (1.2) throughout in mind, we then get the following six new variance estimators for t_j $(j=1,2)$

$$m_{d1} = \frac{n}{n-2} \sum{}' \left[\frac{x_i^2}{\pi_i} \left(\frac{1}{\pi_i} - 1 \right) I_{di} - \frac{1}{n(n-1)} \sum x_k^2 \left(\frac{1}{\pi_k} - 1 \right) I_{dk} \right] (r_i - \bar{r})^2,$$

$$m_{d2} = \frac{n}{n-2} \sum{}' \left[\frac{x_i^2}{\pi_i} \left(\frac{1}{\pi_i} - 1 \right) I_{di} \right.$$

$$- \frac{1}{n(n-1)} \sum' \frac{x_k^2}{\pi_k}\left(\frac{1}{\pi_k} - 1\right) I_{dk}\Bigg] (r_i - \bar{r})^2,$$

$$m_{d3} = \frac{E_p[\Sigma' \alpha_i(1)]}{\Sigma' \alpha_i(1)} m_{d1}$$

$$= \frac{\sum x_k^2 \left(\frac{1}{\pi_k} - 1\right) I_{dk}}{\sum' \frac{x_k^2}{\pi_k}\left(\frac{1}{\pi_k} - 1\right) I_{dk} - \frac{1}{n-1} \sum \frac{x_k^2}{\pi_k}\left(\frac{1}{\pi_k} - 1\right) I_{dk}} m_{d1}$$

$$m_{d4} = \frac{E_p[\Sigma' \alpha_i(2)]}{\Sigma' \alpha_i(2)} m_{d2} = \frac{\sum x_k^2 \left(\frac{1}{\pi_k} - 1\right) I_{dk}}{\sum' \frac{x_k^2}{\pi_k}\left(\frac{1}{\pi_k} - 1\right) I_{dk}} m_{d2}$$

$$m_{d5} = \frac{n}{n-1} \frac{\Sigma' (r_i - \bar{r})^2}{\Sigma' f_i / x_i^2} \sum \frac{1-\pi_k}{\pi_k} f_k I_{dk}$$

and,

$$m_{d6} = \frac{n}{n-1} \frac{\Sigma' (r_i - \bar{r})^2}{\Sigma f_i \pi_i / x_i^2} \sum \frac{1-\pi_k}{\pi_k} f_k I_{dk} .$$

For simplicity m_{dj} will be denoted in tables by m_j (j=1 (1) 6).

For t_1, t_2 we then have ten variance estimators each namely $v_j(t_i)$, $K_j(t_i)$, i=1,2, j=1,2 and m_{dk}, k=1,...,6. Another estimator for Y_d may be taken as $t_\theta = \theta t_1 + (1-\theta) t_2$, $\theta \in (0,1)$. The optimum chioce of θ is ofcourse $\bar{\theta} = [V_p(t_2) - C_p(t_1,t_2)] \; / \; [V_p(t_1) + V_p(t_2) - 2C_p(t_1,t_2)]$, which being unknown, θ in t_θ may be taken as $\hat{\theta}$ with each unknown parameter in $\bar{\theta}$ being replaced by a suitable estimator for it. To avoid $\hat{\theta}$ from ranging beyond [0,1] we drop, following Schaible (1979) and Schaible (1992),
$C_p(\cdot,\cdot)$ in $\bar{\theta}$ and take $\hat{\theta}$ as $\bar{\theta}$ with $C_p(\cdot,\cdot)$ dropped and $V_p(\cdot)$ terms replaced by their estimators and denote the resulting t_θ by $t_{\hat{\theta}}$. The variance estimators for $t_{\hat{\theta}}$ are taken as

$$v(t_{\hat{\theta}}) = \hat{\theta}^2 v(t_1) + (1-\hat{\theta})^2 v(t_2) + 2\hat{\theta}(1-\hat{\theta})c(t_1, t_2)$$

with $v(t_j)$ as variance estimators for t_j (j=1,2) and $c(t_1, t_2)$ as an estimator for $C_p(t_1, t_2)$ of the form

$$c(t_1, t_2) = \sum\nolimits' \sum\nolimits' \Delta_{ij}\left(\frac{a_{si}I_{di}}{\pi_i} - \frac{a_{sj}I_{dj}}{\pi_j}\right)\left(\frac{b_{si}I_{di}}{\pi_i} - \frac{b_{sj}I_{dj}}{\pi_j}\right)$$

with (a_{si}, b_{si}) as (e_{di}, e_i), $(e_{di}, g'_{sdi}e_i)$, $(g_{sdi}e_{di}, e_i)$ or $(g_{sdi}e_{di}, g'_{sdi}e_i)$ yielding 4 alternative forms of $c(t_1, t_2)$ denoted respectively as c_1, c_2, c_3 and c_4. The corresponding predictors are denoted respectively by TH11, TH12, TH21 and TH22 and the corresponding variance estimators by VT11, VT12, VT21 and VT22 in the tables.

Writing e_d for an estimator for Y_d based on s of a large size n with v_d as its non-negative variance estimator it is usual to regard as discussed in the earlier chapters too, the distribution of $\delta = (e_d - Y_d)/\sqrt{v_d}$ as approximately that of the standard normal deviate τ or Student's t_{n-1} distribution with (n-1) degrees of freedom (df). Consequently $(e_d \pm c_{\alpha/2}\sqrt{v_d})$ is taken as a $100(1-\alpha)$ percent confidence interval for Y_d with $c_{\alpha/2}$ as the $100\alpha/2$ percent point on the right tail of the distribution of τ or t_{n-1}, with α in (0,1), recognizing $100(1-\alpha)$ as the nominal confidence coefficient of CI. Taking e_d as $\bar{t}$, t_1, t_2, $t_{\hat{\theta}}$ and v_d as their various alternative variance estimators noted above, our plan is to study the performances of the respective CI's. As it is very difficult, obviously, to study their relative efficacies theoretically, we then proceed to undertake a simulation study as in the earlier chapters, to examine their efficacies numerically. The simulation study in detail, is given in the next section presenting the numerical findings in the tables and conclusions in a series of comments and remarks.

4.2 SIMULATION STUDY.

We take N=752, D=33. Domain U_1 consists of the first N_1 units of

U, domain U_2 consists of the next N_2 units and so on. The domains and their respective sizes are $(U_1,2)$, $(U_2,2)$, $(U_3,3)$, $(U_4,3)$, $(U_5,3)$, $(U_6,4)$, $(U_7,4)$, $(U_8,5)$, $(U_9,5)$, $(U_{10},5)$, $(U_{11},8)$, $(U_{12},8)$, $(U_{13},9)$, $(U_{14},9)$, $(U_{15},10)$, $(U_{16},10)$, $(U_{17},12)$, $(U_{18},13)$, $(U_{19},13)$, $(U_{20},14)$, $(U_{21},17)$, $(U_{22},19)$, $(U_{23},19)$, $(U_{24},20)$, $(U_{25},25)$, $(U_{26},30)$, $(U_{27},32)$, $(U_{28},49)$, $(U_{29},55)$, $(U_{30},65)$, $(U_{31},83)$, $(U_{32},91)$, $(U_{33},105)$.

Domains are divided into G = 4 groups, group 1 consiting of domains (4, 5, 9, 21, 22, 26 and 33), group 2 consists of domains (3, 6, 10, 15, 20, 23, 27 and 32), group 3 consists of domains (2, 7, 11, 14, 19, 24, 28 and 31) and group 4 consists of domains (1, 8, 12, 13, 16, 17, 18, 25, 29 and 30).

First we consider the model connecting y and x as :

$$y_i = \theta(g) + \beta(d)\, x_i + \sigma\, x_i^{h/2}\, \varepsilon_i, \quad i \in U_d,$$

for U_d in g-th group, g = 1 (1) G. We generate ε_i's, i = 1 (1) 752, from the standard normal distribution. The auxiliary x_i, i = 1 (1) 752, are generated from the distribution with density

$$f(x) = \frac{1}{8.5} \exp\left[-(x - 5.0) / 8.5 \right], \quad x \geq 5.0.$$

We throughout take $\sigma = 1.0$ and $\beta(d) = \beta = 2.0$ for each d, choose h = 0.8, 1.4 and two sets of values of $\theta(g)$ namely (i) $\theta(g) = 0$, g = 1 (1) 4 and (ii) $\theta(1) = 0.4$, $\theta(2) = 1.0$, $\theta(3) = 1.5$ and $\theta(4) = 2.9$. The parameters relating to the four different $\underline{Y}$ -vectors thus obtained are given in table 4.1 below :

Table 4.1

Model parameters of $\underline{Y}$-populations

Pop. Id.	$\theta(1)$	$\theta(2)$	$\theta(3)$	$\theta(4)$	σ	β	h
I	0.0	0.0	0.0	0.0	1.0	2.0	0.8
II	0.0	0.0	0.0	0.0	1.0	2.0	1.4
III	0.4	1.0	1.5	2.9	1.0	2.0	0.8
IV	0.4	1.0	1.5	2.9	1.0	2.0	1.4

Samples from $U = (1,\ldots,N)$ employing Hartley-Rao (1962) sampling scheme are drawn using size-measures w_i $(i=1,\ldots,N)$ with values generated from

$$f(w) = \frac{1}{15.0} \exp\left[- (w - 20.0) / 15.0 \right], \quad w \geq 20.0.$$

Samples of three sizes n = 38, 105, 150 are drawn each replicated R=1000 times.

The following measures are considered for evaluations of comparative performances of confidence intervals of nominal confidence coefficients $100(1-\alpha)$ percent based on (e,v) taking $\alpha = .01, .05, .10$.

(1) R(d,e,v) = number of replicated samples admitting values of both e and v.

(2) Pseudo Mean Square Error of e for U_d when v is the variance estimator :

$$PMSE(d,e,v) = \frac{1}{R(d,e,v)} \sum\nolimits_r (e - Y_d)^2$$

where, $\sum_r \equiv$ the sum over samples admitting both e and v.

(3) Pseudo Relative Bias of v corresponding to e for U_d :

$$PRB(d,e,v) = \frac{1}{R(d,e,v)} \sum\nolimits_r \frac{v}{PMSE(d,e,v)} - 1 .$$

(4) Pseudo Relative Stability of v corresponding to e for U_d :

$$PRS(d,e,v) = \left[\frac{1}{R(d,e,v)} \sum\nolimits_r \left(\frac{v}{PMSE(d,e,v)} - 1 \right)^2 \right]^{1/2} .$$

(5) Pseudo Standardized Length of confidence interval using (e,v) for U_d :

$$PSL(d,e,v) = \frac{1}{R(d,e,v)} \sum\nolimits_r \sqrt{v} \Big/ \sqrt{PMSE(d,e,v)} .$$

(6) Pseudo Average Coefficient of Variation :

$$ACV_1(d,e,v) = \frac{1}{R(d,e,v)} \sum_r \left| \frac{\sqrt{v}}{e} \right| .$$

(7) Average Coefficient of Variation :

$$ACV_2(d,e,v) = \frac{1}{R(d,e,v)} \sum_r \left| \frac{\sqrt{v}}{Y_d} \right| .$$

(8) Absolute Relative Bias of e :

$$ARB(d,e,v) = \left| \left(\frac{1}{R(d,e,v)} \sum_r e \right) - Y_d \right| \Big/ Y_d .$$

(9) Absolute Relative Error of e :

$$ARE(d,e,v) = \frac{1}{R(d,e,v)} \sum_r \frac{|e - Y_d|}{Y_d} .$$

(10) Actual Coverage Percentage :

$$ACP(\alpha,d,e,v) = \frac{1}{R(d,e,v)} \sum_r I(\alpha,d,e,v),$$

where, $I(\alpha,d,e,v) = 1$ if $Y_d \in (e_d \pm z_{\alpha/2}\sqrt{v_d})$
$= 0$ else.

(10a) Pseudo Coefficient of variation (PCV) : $PCV(d,e,v) = \{\frac{1}{R(d,e,v)}\Sigma r(\frac{v}{A(d,e,v)} - 1)^2\}^{\frac{1}{2}}$, $A(d,e,v) = \frac{1}{R(d,e,v)}\Sigma r v$.

Several overall measures averaged across the domains are :

(11) $$\overline{MSE}(e,v) = \frac{1}{D} \sum_{d=1}^{D} PMSE(d,e,v) .$$

(12) $$PRB(e,v) = \frac{1}{D} \sum_{d=1}^{D} PRB(d,e,v) .$$

(13) $$PRS(e,v) = \frac{1}{D} \sum_{d=1}^{D} PRS(d,e,v) .$$

(14) $$PSL(e,v) = \frac{1}{D}\sum_{d=1}^{D} PSL(d,e,v) .$$

(15) $$ACV_1(e,v) = \frac{1}{D}\sum_{d=1}^{D} ACV_1(d,e,v) .$$

(16) $$ACV_2(e,v) = \frac{1}{D}\sum_{d=1}^{D} ACV_2(d,e,v) .$$

(17) Average Absolute Relative Bias :

$$AARB(e,v) = \frac{1}{D}\sum_{d=1}^{D} ARB(d,e,v) .$$

(18) Average Absolute Relative Error :

$$AARE(e,v) = \frac{1}{D}\sum_{d=1}^{D} ARE(d,e,v) .$$

(19) $$ACP(\alpha,e,v) = \frac{1}{D}\sum_{d=1}^{D} ACP(\alpha,d,e,v) .$$

(20) $$\overline{EFF}(e,v) = \left[\frac{\overline{MSE}\ (\ HTE,\ YGE\ )}{\overline{MSE}\ (\ e,\ v\ )} \right]^{1/2} .$$

The greg predictors are calculated for four choices of Q_i as $(1-\pi_i)/\pi_i x_i$, $1/\pi_i x_i$, $1/x_i^2$ and $1/x_i$ but since they do not differ much among themselves we show in tables in Appendix-D those only for $1/\pi_i x_i$ due to Hájek (1971) and hence denoted by H.

For a good pair (e, v) we desire

(a) R(d, e, v) to be high,

(b) $\left| ACP(\alpha,e,v) - 100(1-\alpha) \right|$ to be small and,

(c) measures (11) — (19) to be small and (20) to be high.

4.3 RESULTS OF SIMULATION STUDY.

(1) The non-synthetic greg predictor, denoted by 'NSY' in tables below, is practically useless for U_d with very small N_d — the value of R(d,e,v) itself being too small and as low as 2 in many cases though R=1000. When it is bad, the composite predictor need not be tried. Calculations concerning the NSY and composite predictors are not shown in tables in case R(d,e,v) is less than 50. The corresponding synthetic greg predictor, marked 'SY' in tables, however is found quite serviceable even in such cases.

(2) Our main observation is that the newly proposed variance estimators $m_1,..,m_6$ (denoted respectively by M1,..,M6 in tables) perform as impressively good competitors against the traditional ones even in domain estimation as in estimating population totals and especially so even when N_d's are small and synthetic greg predictors are employed and also when composite ones are used in case N_d's are not too small.

A few less important observations are also worth noting, e.g.,

(3) The observation (2) persists when $\theta(g) = 0$, g=1,..,4 and h=0.8,1.4. In particular, M1,..,M4 perform better when N_d's are pretty small while M5 and M6 seem preferable for larger N_d's.

(4) For $\theta(g) \neq 0$, but h=0.8, M1,..,M6 do well but when h=1.4, no clear picture is discernible.

We present in the Appedix D below five tables : Table D.1 showing the R(d,e,v) values, Table D.2, D.3 and D.4 showing domainwise figures only for N_d=9, 49 and 105 and Table D.5 presents over-all figures averaging over all the 33 domains.

APPENDIX D

Table D.1

Domain sizes and the domain-wise R(d,e,v)- values

$R(d,e,v)$ = Number of replicated samples yielding values of both e and v related to a domain U_d.

Domain no. :		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
Domain Size :		2	2	3	3	3	4	4	5	5	5	8	8	9	9	10	10	12	13	13	14	17	19	19	20	25	30	32	49	55	65	83	91	105
e	v																																	
HTE	YGE	229	256	356	321	289	378	360	566	501	446	711	729	808	768	820	742	815	864	833	915	959	956	936	951	991	984	989	1000	1000	1000	1000	1000	1000
NSY	TAY	14	13	37	32	35	68	53	173	120	95	311	341	439	410	463	367	449	552	469	689	787	750	698	802	960	911	952	993	999	1000	1000	1000	1000
NSY	TAY2	14	13	37	32	35	68	53	173	120	95	311	341	439	410	463	367	449	552	469	689	787	750	698	802	960	911	952	993	999	1000	1000	1000	1000
NSY	KT	14	13	37	32	35	68	53	173	120	95	311	341	439	410	463	367	449	552	469	689	787	750	698	802	960	911	952	993	999	1000	1000	1000	1000
NSY	KT2	14	13	37	32	35	68	53	173	120	95	311	341	439	410	463	367	449	552	469	689	787	750	698	802	960	911	952	993	999	1000	1000	1000	1000
NSY	M1	228	256	356	269	264	355	360	362	437	272	685	553	678	635	768	655	733	754	800	824	893	878	883	919	894	933	969	990	989	1000	999	1000	1000
NSY	M2	226	256	356	231	248	350	355	382	441	345	688	579	735	689	814	713	772	844	821	870	938	911	922	934	965	959	980	999	999	1000	1000	1000	1000
NSY	M3	228	256	356	269	264	355	360	362	437	272	685	553	678	635	768	655	733	754	800	824	915	878	883	919	898	933	969	991	989	1000	999	1000	1000
NSY	M4	226	256	356	231	248	350	355	382	441	345	688	579	735	689	814	713	772	844	821	870	938	911	922	934	965	959	980	999	999	1000	1000	1000	1000
NSY	M5	229	256	356	321	289	378	360	566	501	446	711	729	808	768	820	742	815	864	833	915	959	956	936	951	991	984	989	1000	1000	1000	1000	1000	1000
NSY	M6	229	256	356	321	289	378	360	566	501	446	711	729	808	768	820	742	815	864	833	915	959	956	936	951	991	984	989	1000	1000	1000	1000	1000	1000
SY	TAY	229	256	356	321	289	378	360	566	501	446	711	729	808	768	820	742	815	864	833	915	959	956	936	951	991	984	989	1000	1000	1000	1000	1000	1000
SY	TAY2	229	256	356	321	289	378	360	566	501	446	711	729	808	768	820	742	815	864	833	915	959	956	936	951	991	984	989	1000	1000	1000	1000	1000	1000
SY	KT	229	256	356	321	289	378	360	566	501	446	711	729	808	768	820	742	815	864	833	915	959	956	936	951	991	984	989	1000	1000	1000	1000	1000	1000
SY	KT2	229	256	356	321	289	378	360	566	501	446	711	729	808	768	820	742	815	864	833	915	959	956	936	951	991	984	989	1000	1000	1000	1000	1000	1000
SY	M1	228	256	356	269	264	355	360	362	437	272	685	553	678	635	768	655	733	754	800	824	893	878	883	919	894	933	969	990	989	1000	999	1000	1000
SY	M2	226	256	356	231	248	350	355	382	441	345	688	579	735	689	814	713	772	844	821	870	938	911	922	934	965	959	980	999	999	1000	1000	1000	1000
SY	M3	999	1000	1000	948	975	977	1000	796	936	826	974	824	870	867	948	913	918	890	967	909	956	922	947	968	907	949	980	991	989	1000	999	1000	1000
SY	M4	226	256	356	231	248	350	355	382	441	345	688	579	735	689	814	713	772	844	821	870	938	911	922	934	965	959	980	999	999	1000	1000	1000	1000
SY	M5	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
SY	M6	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
TH11	VT11	14	13	37	32	35	68	53	173	120	95	311	341	439	410	463	367	449	552	469	689	787	750	698	802	960	911	952	993	999	1000	1000	1000	1000
TH12	VT12	14	13	37	32	35	68	53	173	120	95	311	341	439	410	463	367	449	552	469	689	787	750	698	802	960	911	952	993	999	1000	1000	1000	1000
TH21	VT21	14	13	37	32	35	68	53	173	120	95	311	341	439	410	463	367	449	552	469	689	787	750	698	802	960	911	952	993	999	1000	1000	1000	1000
TH22	VT22	14	13	37	32	35	68	53	173	120	95	311	341	439	410	463	367	449	552	469	689	787	750	698	802	960	911	952	993	999	1000	1000	1000	1000

HTE = Horvitz-Thompson estimator; YGE = Yates-Grundy estimator; NSY = non-synthetic estimator; SY = synthetic estimator; TH = Composite estimator; TAY and TAY2 are Särndal's Taylor series based estimators; KT, KT2 = Kott's estimators.

Comments: To ensure that 70% or more replicates may yield a CI, domain-size should not be too small if a non-synthetic estimator is employed. The situation is better if a synthetic estimator or the HTE is used.

TABLE D.2

Domain-wise Statistics for each pair (e,v)

DOMAIN NO. : 13 DOMAIN SIZE : 9

Q	e	v	PRB	PRS	PSL	ACV(1)	ACV(2)	PCV	ACP	ARB	ARE
H	HTE	YGE	0.2100	1.5294	0.8958	0.7779	0.9444	1.2520	84.901	0.2961	0.7951
H	NSY	TAY	-0.0776	1.1010	0.7794	0.0599	0.0610	1.1906	72.893	0.0157	0.0546
H	NSY	TAY2	-0.5391	0.9009	0.5212	0.0390	0.0408	1.5659	64.920	0.0157	0.0546
H	NSY	KT	0.3327	2.2336	0.8725	0.0655	0.0683	1.6572	83.599	0.0157	0.0546
H	NSY	KT2	0.3327	2.2336	0.8725	0.0655	0.0683	1.6572	83.599	0.0157	0.0546
H	NSY	M1	-0.6297	0.7174	0.5259	0.0722	0.0741	0.9281	68.289	0.0407	0.0898
H	NSY	M2	-0.6417	0.7343	0.5081	0.0674	0.0693	0.9961	65.170	0.0390	0.0873
H	NSY	M3	0.3598	2.9271	0.7988	0.1002	0.1126	2.1364	93.953	0.0407	0.0898
H	NSY	M4	0.2567	2.6170	0.8001	0.0981	0.1092	2.0725	97.279	0.0390	0.0873
H	NSY	M5	0.3939	0.4673	1.1758	0.1498	0.1533	0.1803	93.317	0.0363	0.0815
H	NSY	M6	0.3799	0.4552	1.1699	0.1490	0.1525	0.1817	93.317	0.0363	0.0815
H	SY	TAY	0.0077	1.0676	0.8365	0.0668	0.0663	1.0594	89.851	0.0052	0.0633
H	SY	TAY2	-0.0129	1.0311	0.8298	0.0662	0.0658	1.0445	89.480	0.0052	0.0633
H	SY	KT	0.1034	1.1480	0.8831	0.0701	0.0700	1.0362	91.832	0.0052	0.0633
H	SY	KT2	0.1034	1.1480	0.8831	0.0701	0.0700	1.0362	91.832	0.0052	0.0633
H	SY	M1	-0.0010	0.9272	0.8637	0.0749	0.0741	0.9281	89.086	0.0040	0.0712
H	SY	M2	-0.0182	0.9782	0.8410	0.0700	0.0693	0.9961	91.020	0.0040	0.0666
H	SY	M3	3.9589	9.6627	1.6915	0.1261	0.1288	1.7775	88.966	0.0027	0.0585
H	SY	M4	2.4430	7.5423	1.3243	0.1059	0.1092	2.0725	88.980	0.0040	0.0666
H	SY	M5	3.6028	3.6974	2.1367	0.1534	0.1531	0.1805	100.000	0.0039	0.0538
H	SY	M6	3.5557	3.6512	2.1256	0.1525	0.1523	0.1820	100.000	0.0039	0.0538
H	TH11	[illegible]	-0.1197	1.0239	0.7701	0.0596	0.0606	1.1552	68.793	0.0145	0.0607
H	TH12	[illegible]	-0.1380	1.0007	0.7630	0.0592	0.0602	1.1498	68.565	0.0145	0.0609
H	TH21	[illegible]	-0.4017	0.9097	0.6119	0.0406	0.0419	1.3641	65.376	0.0117	0.0522
H	TH22	[illegible]	-0.4029	0.9088	0.6112	0.0407	0.0419	1.3643	65.376	0.0117	0.0523

As explained on pp 67-69, PRB, PRS, PSL measure respectively bias and stability of w and length of confidence interval (CI) based on (e,v); ACV(1), ACV(2) and PCV relate to coefficient of variation; ARB, ARE relate to bias and error of e and ACP denotes coverage percentage. For other acronyms please see page 71.

Comments: Since the domain size is small the 'direct' estimators have poor coverage probability except when our newly proposed M_3, M_4, M_5 are used along with a non-synthetic greg predictor. The indirect synthetic estimator is better but the composite estimator [illegible]

TABLE D.3
Domain-wise Statistics for each pair (e,v)
DOMAIN NO. : 28   DOMAIN SIZE : 49

Q	e	v	PRB	PRS	PSL	ACV(1)	ACV(2)	PCV	ACP	ARB	ARE
H	HTE	YGE	0.0475	0.9701	0.9396	0.4358	0.4230	0.9250	89.200	0.0130	0.3558
H	NSY	TAY	-0.4246	0.7300	0.6684	0.0524	0.0521	1.0319	73.011	0.0096	0.0596
H	NSY	TAY2	-0.4206	0.7323	0.6878	0.0533	0.0536	1.0347	80.765	0.0096	0.0596
H	NSY	KT	-0.1085	1.0721	0.8404	0.0651	0.0655	1.1964	87.613	0.0096	0.0596
H	NSY	KT2	-0.1085	1.0721	0.8404	0.0651	0.0655	1.1964	87.613	0.0096	0.0596
H	NSY	M1	-0.2432	0.8337	0.7676	0.0628	0.0619	1.0537	81.212	0.0101	0.0605
H	NSY	M2	-0.2461	0.8342	0.7654	0.0624	0.0616	1.0573	81.782	0.0102	0.0604
H	NSY	M3	0.1711	6.8862	0.8946	0.0718	0.0721	5.8785	96.165	0.0101	0.0605
H	NSY	M4	-0.0620	0.9359	0.8783	0.0704	0.0706	0.9956	95.696	0.0102	0.0604
H	NSY	M5	-0.2344	0.2721	0.8714	0.0697	0.0701	0.1805	92.300	0.0102	0.0603
H	NSY	M6	-0.2423	0.2788	0.8669	0.0694	0.0697	0.1820	92.400	0.0102	0.0603
H	SY	TAY	0.0083	1.0370	0.8903	0.0623	0.0608	1.0284	92.400	0.0007	0.0541
H	SY	TAY2	-0.0222	0.9763	0.8822	0.0617	0.0603	0.9982	92.800	0.0007	0.0541
H	SY	KT	0.0059	0.9689	0.9024	0.0631	0.0617	0.9632	94.500	0.0007	0.0541
H	SY	KT2	0.0059	0.9689	0.9024	0.0631	0.0617	0.9632	94.500	0.0007	0.0541
H	SY	M1	0.0449	1.1019	0.9020	0.0635	0.0619	1.0537	92.121	0.0006	0.0545
H	SY	M2	0.0437	1.1044	0.9006	0.0631	0.0616	1.0573	92.693	0.0007	0.0541
H	SY	M3	0.6166	9.5233	1.0511	0.0728	0.0721	5.8785	95.257	0.0007	0.0544
H	SY	M4	0.2986	1.3269	1.0334	0.0713	0.0706	0.9956	94.695	0.0007	0.0541
H	SY	M5	0.0597	0.2004	1.0252	0.0703	0.0701	0.1805	96.300	0.0007	0.0541
H	SY	M6	0.0489	0.1971	1.0199	0.0700	0.0697	0.1820	96.200	0.0007	0.0541
H	TH11	VT11	-0.3405	0.7671	0.7148	0.0541	0.0535	1.0423	77.341	0.0063	0.0585
H	TH12	VT12	-0.3583	0.7491	0.7072	0.0538	0.0532	1.0251	77.442	0.0063	0.0588
H	TH21	VT21	-0.3292	0.7012	0.7468	0.0507	0.0504	0.9229	81.672	0.0051	0.0534
H	TH22	VT22	-0.3301	0.6990	0.7469	0.0510	0.0507	0.9198	81.873	0.0051	0.0537

For the acronyms please see the previous page 72.

Comments: The domain size is moderate. So, the HTE and the non-synthetic greg predictor are no longer quite bad; combined with M_3, M_4 or M_5 the latter yields good coverage probabilities. The synthetic greg predictor is quite good but the composite one is still inadequate, because the non-synthetic component is poor.

TABLE D.4

Domain-wise Statistics for each pair (e,v)

DOMAIN NO. : 33 DOMAIN SIZE : 105

Q	e	v	PRB	PRS	PSL	ACV(1)	ACV(2)	PCV	ACP	ARB	ARE
H	HTE	YGE	-0.0296	0.4906	0.9528	0.2992	0.2977	0.5046	90.600	0.0175	0.2499
H	NSY	TAY	-0.1652	0.7934	0.8281	0.0501	0.0496	0.9296	83.200	0.0010	0.0480
H	NSY	TAY2	-0.2479	0.5985	0.8161	0.0494	0.0489	0.7243	87.300	0.0010	0.0480
H	NSY	KT	-0.0951	0.7444	0.8887	0.0538	0.0532	0.8159	89.800	0.0010	0.0480
H	NSY	KT2	-0.0951	0.7444	0.8887	0.0538	0.0532	0.8159	89.800	0.0010	0.0480
H	NSY	M1	0.0481	1.0138	0.9147	0.0556	0.0548	0.9662	87.900	0.0010	0.0480
H	NSY	M2	0.0478	1.0105	0.9155	0.0556	0.0549	0.9633	88.200	0.0010	0.0480
H	NSY	M3	-0.0434	0.6703	0.9167	0.0557	0.0549	0.6992	92.600	0.0010	0.0480
H	NSY	M4	-0.0431	0.6672	0.9173	0.0557	0.0550	0.6957	92.600	0.0010	0.0480
H	NSY	M5	-0.3671	0.3845	0.7923	0.0476	0.0475	0.1805	89.200	0.0010	0.0480
H	NSY	M6	-0.3736	0.3906	0.7882	0.0474	0.0472	0.1820	88.800	0.0010	0.0480
H	SY	TAY	0.0377	0.9737	0.9166	0.0549	0.0541	0.9376	91.800	0.0008	0.0474
H	SY	TAY2	-0.0082	0.8792	0.9043	0.0542	0.0534	0.8864	92.800	0.0008	0.0474
H	SY	KT	0.0062	0.8737	0.9147	0.0548	0.0540	0.8683	92.900	0.0008	0.0474
H	SY	KT2	0.0062	0.8737	0.9147	0.0548	0.0540	0.8683	92.900	0.0008	0.0474
H	SY	M1	0.0804	1.0469	0.9287	0.0557	0.0548	0.9662	92.100	0.0008	0.0474
H	SY	M2	0.0800	1.0435	0.9295	0.0558	0.0549	0.9633	92.200	0.0008	0.0474
H	SY	M3	-0.0139	0.6896	0.9307	0.0557	0.0549	0.6992	92.700	0.0008	0.0474
H	SY	M4	-0.0136	0.6864	0.9313	0.0557	0.0550	0.6957	92.700	0.0008	0.0474
H	SY	M5	-0.3477	0.3671	0.8044	0.0476	0.0475	0.1805	88.900	0.0008	0.0474
H	SY	M6	-0.3543	0.3733	0.8002	0.0474	0.0472	0.1820	88.600	0.0008	0.0474
H	TH11	VT11	-0.1083	0.8367	0.8535	0.0514	0.0508	0.9305	86.700	0.0009	0.0478
H	TH12	VT12	-0.1389	0.7833	0.8431	0.0510	0.0504	0.8952	86.800	0.0007	0.0481
H	TH21	VT21	-0.1859	0.6462	0.8396	0.0489	0.0484	0.7602	88.300	0.0019	0.0462
H	TH22	VT22	-0.1887	0.6379	0.8398	0.0491	0.0486	0.7512	88.600	0.0018	0.0464

For the acronyms please see pp 60-65 and also pp 71-72.

Comments: Even though the domain size is quite large the non-synthetic greg predictor except when combined with M_3, M_4, M_5 is not quite serviceable and so is the synthetic greg combined with M_5. The combined estimator turns out quite poor probably because the covariance term is indiscriminately ignored.

TABLE D.5

Over-all performances for each pair (e,v)

Q	e	v	PRB	PRS	PSL	ACV(1)	ACV(2)	PCV	ACP	AARB	AARE	Eff
H	HTE	YGE	0.2081	1.1471	0.9679	0.7025	1.2296	0.9280	89.863	0.6561	0.9903	1.0000
H	NSY	M1	6.3947	9.7413	1.5729	0.2097	0.1923	1.0945	81.327	0.0199	0.1005	6.3768
H	NSY	M2	5.4068	8.4886	1.4987	0.2077	0.1897	1.1187	80.322	0.0219	0.1001	6.4052
H	NSY	M3	1.8663	3.8419	1.2636	0.2106	0.1686	1.3576	91.935	0.0199	0.1004	6.3821
H	NSY	M4	1.6181	3.3053	1.2279	0.2054	0.1659	1.1769	92.434	0.0219	0.1001	6.4052
H	NSY	M5	1.4920	1.8202	1.3662	0.1654	0.1566	0.1805	94.135	0.0192	0.0983	6.4358
H	NSY	M6	1.4674	1.8020	1.3593	0.1646	0.1558	0.1820	94.054	0.0192	0.0983	6.4358
H	SY	TAY	0.0392	1.2122	0.8351	1.3297	0.1770	1.1622	85.648	0.0627	0.1685	6.5857
H	SY	TAY2	0.0119	1.1599	0.8272	1.3219	0.1753	1.1432	85.682	0.0627	0.1685	6.5857
H	SY	KT	0.0119	1.1273	0.8364	1.2756	0.1725	1.1118	87.326	0.0627	0.1685	6.5857
H	SY	KT2	0.0119	1.1273	0.8364	1.2756	0.1725	1.1118	87.326	0.0627	0.1685	6.5857
H	SY	M1	0.0661	1.1737	0.8654	1.3634	0.1923	1.0945	88.353	0.0754	0.1787	6.4393
H	SY	M2	0.0549	1.1849	0.8546	1.3696	0.1897	1.1187	87.786	0.0751	0.1773	6.5106
H	SY	M3	1.3660	3.8023	1.1771	0.4388	0.1668	1.2597	90.581	0.0046	0.1106	6.8969
H	SY	M4	0.4955	2.2578	0.9264	0.9338	0.1659	1.1769	84.618	0.0751	0.1773	6.5106
H	SY	M5	0.9886	1.3529	1.2422	0.2781	0.1563	0.1805	92.985	0.0028	0.1076	6.9699
H	SY	M6	0.9683	1.3382	1.2358	0.2769	0.1555	0.1820	92.903	0.0028	0.1076	6.9699
H	TH11	VT11	1.2814	2.6467	1.1657	0.1774	0.1670	1.0985	80.260	0.0362	0.0986	6.9713
H	TH12	VT12	1.2036	2.5434	1.1489	0.1766	0.1661	1.0867	80.274	0.0367	0.0996	6.9374
H	TH21	VT21	16.7477	17.8151	1.6857	0.0656	0.0625	1.0039	74.515	0.0114	0.0667	7.4917
H	TH22	VT22	16.7654	17.8329	1.6867	0.0657	0.0626	1.0040	74.497	0.0114	0.0668	7.4653

The acronyms are as explained on pp 60-65. HTE, YGE denote Horvitz - Thompson's and Yates - Grundy's estimators; NSY and SY denote non-synthetic and synthetic greg predictors and TH the composite predictor. TAY, TAY2 are Särndal's variance estimators; KT, KT2 are Kott's variance estimators; PRB, PRS denote bias and stability of v, ARB and ARE the bias and error of e; PCV, ACV(1) and ACV(2) relate to coefficients of variation; ACP the coverage percentages of confidence intervals and EFF the efficiency of (e, v) relative to (HTE,YGE).

Comments: The direct HTE ensures a good coverage probability but needs a too wide CI. The direct NSY combined with M_3, M_4, M_5, M_6 seems serviceable and even better than the synthetic combined with Särndal's and Kott's variance estimators. But the synthetic greg coupled with M_3, M_4, M_5, M_6 seem better. The composite is poor and may be kept out of reckoning.

CHAPTER FIVE

RATIO ESTIMATION BY RANDOMIZED RESPONSE

5.0 SUMMARY.

Supposing that only randomized response (RR) instead of direct response (DR) is available, modifications are considered on the ratio estimator for a survey population total of a sensitive variable based on a simple random sample taken without replacement (SRSWOR) and on its DR-based variance estimators. A simulation-based numerical comparison is presented on the relative efficacies of confidence intervals involving the respective modified variance estimators.

5.1 INTRODUCTION AND THE MAIN RESULTS.

The variable of interest y is supposed to relate to stigmatizing matters like the amount lost in gambling or spent on drug etc. An SRSWOR of size n is taken to estimate Y. Since y is sensitive we suppose values y_i for i in a sample s are unavailable. Instead, following Chaudhuri (1987), we suppose that RR's, say, z_i are available for any sampled individual as

$$z_i = a_j y_i + b_k. \tag{5.1.1}$$

Here out of a vector $\underline{A}=(a_1,\ldots,a_j,\ldots,a_J)$ of pre-assigned real numbers a number a_j is chosen at random by a sampled respondent when interviewed in a survey, and so is b_k out of another pre-assigned vector $\underline{B}=(b_1,\ldots,b_k,\ldots,b_K)$. Though $\underline{A}$ with mean θ_a and variance σ_a^2, say, and $\underline{B}$ with mean θ_b and variance σ_b^2, say, are known both to the respondent and the interviewer, the three elements on the right hand side of (5.1.1) are known only to the former but not to the latter to whom only the value z_i is reported. Writing E_R (V_R) as operator of expectation (variance) with respect to this 'randomization' experiment done independently by each respondent and a_j being chosen independently of b_k, it follows that

$$E_R(z_i) = \theta_a y_i + \theta_b \text{ and } V_R(z_i) = \sigma_a^2 y_i^2 + \sigma_b^2. \qquad (5.1.2)$$

We shall write $r_i=(z_i-\theta_b)/\theta_a$ and $V_R(r_i)=V_i$. Then

$$\hat{V}_i = \frac{\theta_a^2}{\theta_a^2+\sigma_a^2}\left[\frac{\sigma_a^2}{\theta_a^2} r_i^2 + \frac{\sigma_b^2}{\theta_a^2}\right] \text{ satisfies } E_R(\hat{V}_i) = V_i.$$

If DR were available, then a commonly employed estimator for Y is the ratio estimator

$$t = X\frac{\Sigma' y_i}{\Sigma' x_i} = X\cdot r \text{ with } r = \frac{\Sigma' y_i}{\Sigma' x_i},$$

denoting by Σ' the sum over i in s. The following variance estimators for t are well-known, vide Chaudhuri and Stenger (1992), Cochran (1977), Royall and Eberhardt (1975) and Royall and Cumberland (1978 a), namely,

$$v_0 = \frac{N^2(1-f)}{n(n-1)}\Sigma'(y_i-rx_i)^2,\ v_2 = \left(\frac{\bar{X}}{\bar{x}}\right)^2 v_0,$$

writing $\bar{X}$, $\bar{x}$ as population and sample means of x,

$$v_H = \frac{\bar{X}\bar{x}_c}{\bar{x}^2}\left(1-\frac{c_x^2}{n}\right)^{-1} v_0$$

and,
$$v_D = \frac{N^2(1-f)}{n}\frac{\bar{X}\bar{x}_c}{\bar{x}^2}\frac{1}{n}\sum{}'\frac{(y_i-rx_i)^2}{(1-x_i/n\bar{x})}$$

writing $\bar{x}_c$ as the mean of non-sampled values of x and $c_x^2 = \dfrac{\Sigma'(x_i-\bar{x})^2}{(n-1)\bar{x}^2}$. These are purported to provide estimators for $V = E_p(t-Y)^2$, writing E_p for expectation operator with respect to sampling design p. This V in practice is approximated by $V_a = \dfrac{N^2(1-f)}{n(n-1)}\Sigma(y_i-Fx_i)^2$, $f=n/N$, $F=Y/X$. Since y_i is not available we assume r_i for $i \in s$ are available and we take $e = X\Sigma' r_i/\Sigma' x_i = Xr'$ with $r' = \Sigma' r_i/\Sigma' x_i$, as a natural substitute for t with which we may proceed to estimate Y recognizing $V' = E_pE_R(e-Y)^2$ as a measure of its error. We shall assume that E_p and E_R

commute and consider estimators for V', to be called variance estimators for e, as quantities v' which satisfy the condition

$$E_R(v') = v + a_s \tag{5.1.3}$$

writing v as a particular variance estimator for t, like v_0, v_2, v_H, v_D above or any other, and $a_s = (\Sigma' V_i)(X/\Sigma' x_k)^2$.

Let, $e_i = y_i - rx_i$ and $e_i' = r_i - r' x_i$; then

$$E_R(e_i'^2) = e_i^2 + (1-2x_i/n\bar{x})V_i + (x_i/n\bar{x})^2 \Sigma' V_k.$$

So,

$$E_R \left[\Sigma' e_i'^2 - \Sigma' \{ 1 - (2x_i/n\bar{x}) + (\Sigma' x_k^2)/(n\bar{x})^2 \} \hat{V}_i \right]$$

$$= E_R(\Sigma' b_{si}) = E_R(b_s) = \Sigma' e_i^2$$

where, $b_s = \Sigma' b_{si}$, and,

$$b_{si} = e_i'^2 - \left(1 - \frac{2x_i}{n\bar{x}} + \frac{\Sigma' x_k^2}{(n\bar{x})^2}\right)\hat{V}_i \text{ , say.}$$

Let $a_s = (\Sigma' V_i)(X/\Sigma' x_k)^2$ and $\hat{a}_s = (\Sigma' \hat{V}_i)(X/\Sigma' x_k)^2$. Then corresponding to v_0, v_2, v_H, v_D above, we propose the following 4 variance estimators v' for e which may be checked to satisfy (5.1.3), namely

$$v_0' = \hat{a}_s + \frac{N^2(1-f)}{n(n-1)} \left[\sum' e_i'^2 - \sum' \left(1 - \frac{2x_i}{n\bar{x}} + \frac{\Sigma' x_k^2}{(n\bar{x})^2}\right)\hat{V}_i \right]$$

$$v_2' = \hat{a}_s + (\bar{X} / \bar{x})^2 (v_0' - \hat{a}_s)$$

$$v_H' = \hat{a}_s + \frac{\bar{X}\,\bar{x}_c}{\bar{x}^2} \left(1 - \frac{c_x^2}{n} \right)^{-1} (v_0' - \hat{a}_s) \text{ and}$$

$$v_D' = \hat{a}_s + \frac{N^2(1-f)}{n} \frac{\bar{X}\,\bar{x}_c}{\bar{x}^2} \frac{1}{n} \sum' \frac{b_{si}}{(1-x_i/n\bar{x})} .$$

Then we postulate the linear regression model of Chapter Three with error variance proportional to the regressor as is appropriate for the choice of the ratio estimator. Also we adopt Brewer's (1979) asymptotic approach.

Writing $M(x)=E_m(V_a)$, $M'(x)=\lim E_pE_m(t-Y)^2$ we propose first the following alternative forms of v, namely

$$v_{01} = \frac{M(x)}{\lim E_pE_m(v_0)} v_0 = \left(\frac{(1-C_0^2/N)}{(1-C_0^2/n)}\right) v_0$$

on writing Σ for sum over i in U, $C_0^2 = S_x^2 / \bar{X}^2$, $S_x^2 = \frac{1}{N-1}\Sigma(y_i-Fx_i)^2$,

$$v_{21} = \left(\frac{\bar{X}}{\bar{x}}\right)^2 v_{01},\quad v_{02} = \frac{M(x)}{E_m(v_0)} v_0 = \left(\frac{\bar{X}}{\bar{x}}\right)\left(\frac{(1-C_0^2/N)}{(1-c_x^2/n)}\right) v_0$$

$$v_{03} = \frac{M'(x)}{\lim E_pE_m(v_0)} v_0 = \left(1-\frac{C_0^2}{n}\right)^{-1} v_0$$

$$v_{23} = \frac{M'(x)}{\lim E_pE_m(v_0)} v_2 = \left(\frac{\bar{X}}{\bar{x}}\right)^2 v_{03},$$

$$v_{04} = \frac{M'(x)}{E_m(v_0)} v_0 = \left(\frac{\bar{X}}{\bar{x}}\right)\left(1-\frac{c_x^2}{n}\right)^{-1} v_0.$$

For each of these new v's, $\lim E_pE_m(v)$ equals either $M(x)$ or $M'(x)$.

Further we consider a few more v's of the form

$$t(\alpha) = \sum{}' \alpha_i\left(\frac{y_i}{x_i} - \frac{1}{n}\sum{}'\frac{y_k}{x_k}\right)^2$$

with α_i's chosen to satisfy $\lim E_pE_m[t(\alpha)] = M'(x)$ and they turn out as

$$m_1 = \frac{N^2(1-f)}{n(n-1)}\sum{}'\left(x_i^2 - \frac{\Sigma x_k^2}{N(n-1)}\right)\left(\frac{y_i}{x_i} - \frac{1}{n}\sum{}'\frac{y_k}{x_k}\right)^2$$

$$m_2 = \frac{N^2(1-f)}{n(n-1)}\sum{}'\left(x_i^2 - \frac{\Sigma' x_k^2}{n(n-1)}\right)\left(\frac{y_i}{x_i} - \frac{1}{n}\sum{}'\frac{y_k}{x_k}\right)^2$$

$$m_3 = \frac{\frac{n-2}{n-1}\cdot\frac{1}{N}\Sigma x_k^2}{\frac{1}{n}\Sigma' x_k^2 - \frac{1}{n-1}\cdot\frac{1}{N}\Sigma x_k^2} m_1$$

and, $$m_4 = \frac{\frac{1}{N}\Sigma x_k^2}{\frac{1}{n}\Sigma' x_k^2} m_2.$$

From these, v''s subject to (5.1.3) are derived respectively as :

$$v'_{01} = \hat{a}_s + \left(\frac{(1-C_0^2/N)}{(1-C_0^2/n)}\right)(v'_0-\hat{a}_s), \quad v'_{21} = \hat{a}_s + \left(\frac{\bar{X}}{\bar{x}}\right)^2 (v'_{01}-\hat{a}_s),$$

$$v'_{02} = \hat{a}_s + \left(\frac{\bar{X}}{\bar{x}}\right)\left(\frac{(1-C_0^2/N)}{(1-c_x^2/n)}\right)(v'_0-\hat{a}_s), \quad v'_{03} = \hat{a}_s + \left(1-\frac{C_0^2}{n}\right)^{-1}(v'_0-\hat{a}_s)$$

$$v'_{23} = \hat{a}_s + \left(\frac{\bar{X}}{\bar{x}}\right)^2 (v'_{03}-\hat{a}_s), \quad v'_{04} = \hat{a}_s + \left(\frac{\bar{X}}{\bar{x}}\right)\left(1-\frac{c_x^2}{n}\right)^{-1}(v'_0-\hat{a}_s).$$

$$m'_1 = \hat{a}_s + \frac{N^2(1-f)}{n(n-1)} \sum{}' \left\{ \left(x_i^2 - \frac{\Sigma x_k^2}{N(n-1)} \right) \times \left[\left(\frac{r_i}{x_i} - \frac{1}{n}\sum{}' \frac{r_k}{x_k} \right)^2 - \hat{a}_i \right] \right\}$$

$$m'_2 = \hat{a}_s + \frac{N^2(1-f)}{n(n-1)} \sum{}' \left\{ \left(x_i^2 - \frac{\Sigma' x_k^2}{n(n-1)} \right) \times \left[\left(\frac{r_i}{x_i} - \frac{1}{n}\sum{}' \frac{r_k}{x_k} \right)^2 - \hat{a}_i \right] \right\}$$

writing $\hat{a}_i = \frac{n-2}{n}\frac{\hat{V}_i}{x_i^2} + \frac{1}{n^2}\sum{}' \frac{\hat{V}_k}{x_k^2}$

and,

$$m'_3 = \hat{a}_s + \frac{\frac{n-2}{n-1}\cdot\frac{1}{N}\Sigma x_k^2}{\frac{1}{n}\Sigma' x_k^2 - \frac{1}{n-1}\cdot\frac{1}{N}\Sigma x_k^2}(m'_1-\hat{a}_s) \quad \text{and,}$$

$$m'_4 = \hat{a}_s + \frac{\frac{1}{N}\Sigma x_k^2}{\frac{1}{n}\Sigma' x_k^2}(m'_2-\hat{a}_s).$$

As the variance estimators for e are pretty complicated a theoretical comparison of their relative efficacies is difficult to carry out. So we consider setting up confidence intervals (CI) for Y of the form $e \pm \psi_{\alpha/2}\sqrt{v'}$ with the postulation that $d=(e-Y)/\sqrt{v'}$ is distributed as either a standard normal deviate or as Student's t statistic with (n-1) degrees of freedom with $\psi_{\alpha/2}$ as the $100\alpha/2$ percent point on the right tail area of the corresponding distribution, taking α in (0,1) with $100(1-\alpha)$ denoting the nominal confidence coefficient. Taking for v' each of the above-noted alternative forms we consider corresponding CI's and compare the relative efficacies of the latter. For this we resort to a numerical exercise through a simulation study reported next. With (t, v) in place of (e, v') a parallel exercise was done earlier, among others, by Royall and Cumberland (1978 b) and Wu and Deng (1983).

5.2 SIMULATION.

We take N=150, $\sigma=1$, $\beta=1$, $\alpha=0.05$, take x_i's as a random sample from the density $f_\lambda(u) = \frac{1}{\lambda} e^{-u/\lambda}$, $u>0$, $\lambda=8.5$, take τ_i's as a random sample from standard normal distribution N(0,1), $\varepsilon_i=\tau_i\sqrt{x_i}$ and $y_i = x_i+\varepsilon_i$, $i=1,\ldots,150$. Then we draw a replicate of R=1000 SRSWOR's from U of size n=32 each, write Σ_r as sum over replicates, $A = \frac{1}{R}\Sigma_r v'$ and $P = \frac{1}{R}\Sigma_r (e-Y)^2$. Next we take J=20, K=25 and consider arbitrary choices of $\underline{A}$ and $\underline{B}$ denoted respectively $\underline{A}_j$, $\underline{B}_j$, $(j=1,\ldots,5)$ given after the tables that follow. To discriminate among the CI's we consider the following criteria, following Rao and Wu's (1983) well-known convention.

(1) ACP (Actual coverage percentage) ≡ the percentage of replicates for which CI covers Y — the closer it is to the nominal confidence coefficient 95% the better. The ACP's calculated referring to Student's t-table are given in the table after a slash following those by normal table.

(2) ACV (Average coefficient of variation) ≡ the average of $\sqrt{v'}/e$ over the replicates — this reflects the length of CI relative to e.

(3) Pseudo relative bias ≡ $PB(v') \equiv \frac{1}{P}\frac{1}{R}\Sigma_r(v'-P)$

(4) Pseudo relative stability ≡ $PS(v') = \frac{1}{P}\left[\frac{1}{R}\Sigma_r(v'-P)^2\right]^{1/2}$

(5) Pseudo standardized length ≡ $PL(v') = \frac{1}{R}\Sigma_r\sqrt{v'}\,/\sqrt{P}$

(6) Bias of d ≡ $B(d) = \frac{1}{R}\Sigma_r d.$

(7) Mean square error of d ≡ $M(d) = \frac{1}{R}\Sigma_r(d-B(d))^2$

(8) Root beta one of d ≡ $\sqrt{\beta_1(d)} = \frac{1}{R}\sum_r\left(\frac{d-B(d)}{\sqrt{M(d)}}\right)^3$

(9) Excess $\equiv E(d) = \beta_2(d) - 3 = \frac{1}{R} \sum_r \left(\frac{d-B(d)}{\sqrt{M(d)}} \right)^4 - 3.$

(10) PCV (Pseudo coefficient of variation) $\equiv \frac{1}{A} \left[\frac{1}{R} \Sigma_r (v' - A)^2 \right]^{1/2}.$

The smaller the magnitudes of (2) — (10) the better the CI and better the choice (e, v').

The numerical findings are presented in the table below. The five sections I–V of the table relate respectively to five choices of $\underline{A}_j$, $\underline{B}_j$, j=1,..5.

Finally, to make the notations in the table easier, we represent v' by v throughout for all v''s.

5.3 CONCLUSION.

From the five sections of the table presented in Appendix-E at the end of this chapter we find that compared to the last section which duplicates the DR situation the first three do not fare too badly and the performances deteriorate across the sections as σ_A^2 and σ_B^2 increase. For the preservation of confidentiality, only high σ_A^2, σ_B^2 will be acceptable to the respondents. An RR situation as in section IV with too large σ_A^2, σ_B^2 may be unsuitable to the survey designer but the first three situations seem quite effective if RR's could be procured with restricted σ_A^2, σ_B^2 as contained therein. Among the 14 alternative variance estimators it is difficult to identify the most effective ones but m_1', m_2' seem to beat most of the others.

APPENDIX E

The acronyms used are as explained on pp 81-82. ACP denotes coverage percentage of confidence interval (CI), ACV, PCV relate to coefficients of variation; PB, PS relate to bias and stability of v, PL the length of CI and B(.), M(.), $\sqrt{\beta_1(.)}$ and E(.) the bias, MSE, skewness and excess of $d = (e - Y)/\sqrt{v}$.

Table

Performances of CI by several criteria. Especially good (bad) values are under-scored (marked by asterisks).

v	ACP	10^5 ACV	10^4 PCV	-10^4 PB	10^2 PS	10^2 PL	-10^3 B(d)	10^2 M(d)	-10 $\sqrt{\beta_1(d)}$	E(d)
					Section I					
v_0	93.0/93.8	3686	3127	644	30	95	2.4	122	.50	.78
v_2	92.8/93.8	3691	3312	598	32	96	3.1	121	.32	.77
v_H	93.0/93.8	3699	3371	547	32	96	3.2	120	.30	.77
v_D	93.0/93.7	3693	3363	576	32	96	3.8	120	.34	.77
v_{01}	93.0/93.8	3692	3127	615	30	96	2.4	121	.50*	.78
v_{21}	92.8/93.8	3696	3312	568	32	96	3.1	120	.32	.77
v_{02}	93.0/93.7	3693	3162	605	30	96	2.7	121	.41	.77
v_{03}	93.0/93.8	3693	3127	607	30	96	2.4	121	.50*	.78
v_{23}	92.8/93.8	3698	3312	560	32	96	3.1	120	.32	.77
v_{04}	93.0/93.7	3695	3162	597	30	96	2.7	120	.41	.77
m_1	93.0/93.7	3677	3121	691*	30	95	4.4	122	.64*	.80
m_2	92.9/93.8	3677	3119	689*	30	95	4.4	122	.63*	.80
m_3	92.8/93.7	3688	3302	613	32	96	4.9*	121	.41	.79
m_4	92.8/93.7	3688	3309	611	32	96	4.9*	121	.41	.79

v	ACP	10^5 ACV	10^4 PCV	-10^4 PB	10^2 PS	10^2 PL	-10^3 B(d)	10^2 M(d)	-10 $\sqrt{\beta_1(d)}$	E(d)
					Section II					
v_0	89.1/90.1	3633	3587	2653	37	84	27.1	164	-.07	1.77
v_2	88.9/90.2	3688	3745	2615	38	84	26.8	163	-.59	1.78
v_H	89.0/90.1	3676	3796*	2575	38	85	26.7	162	-.65	1.78
v_D	89.0/90.2	3671	3795*	2597	38	84	27.4	163	-.60	1.79
v_{01}	89.1/90.1	3669	3587	2630	37	84	27.1	163	-.07	1.77
v_{21}	89.0/90.3	3674	3745	2591	38	85	26.7	162	-.59	1.78
v_{02}	89.5/90.2	3670	3616	2621	37	84	26.9	162	-.33	1.76
v_{03}	89.1/90.1	3670	3587	2623	37	84	27.1	163	-.07	1.77
v_{23}	89.0/90.3	3675	3745	2585	38	85	26.7	162	-.59	1.78

Section II (continued)

v_{04}	89.5/90.3	3672	3616	2615	37	84	26.9	162	-.33	1.76
m_1	89.1/90.0	3653	3590	2691	38	84	29.3*	165	.17	1.79
m_2	89.1/90.0	3654	3588	2689	38	84	29.2*	165	.14	1.79
m_3	89.0/90.2	3665	3742	2627	38	84	28.7	163	-.43	1.84
m_4	89.0/90.2	3665	3749	2626	38	84	28.7	163	-.44	1.84

Section III

v_0	75.6/77.4	3616	5227	5678	61	63	105	424	23.6	37
v_2	76.3/78.0	3619	5317	5661	61	63	107	426	24.1	38
v_H	76.6/78.0	3627*	5354	5638	61	63	107	425	24.2	39
v_D	76.4/78.0	3620	5376	5652	61	63	112	460*	35.8*	61*
v_{01}	75.7/77.4	3621	5227	5665	61	63	104	423	23.7	37
v_{21}	76.4/78.0	3625*	5317	5647	61	63	107	425	24.2	39
v_{02}	76.1/77.4	3622	5230	5662	61	63	105	423	24.0	38
v_{03}	75.7/77.4	3623	5227	5661	61	63	104	422	23.7	37
v_{23}	76.4/78.0	3627*	5317	5644	61	63	107	424	24.2	39
v_{04}	76.3/77.4	3623	5230	5659	61	63	105	423	24.0	38
m_1	75.2/77.1	3604	5285	5702	61	63	67	412	-7.2	12
m_2	75.3/77.0	3605	5283	5701	61	63	67	410	-7.0	12
m_3	75.8/77.9	3615	5377*	5668	61	63	69	407	-7.2	12
m_4	75.8/77.9	3615	5380*	5667	61	63	69	407	-7.1	12

Section IV

v_0	56.5/58.4	4623	6908	8004	81	42	296	2156	-32.2	44.7
v_2	56.6/58.3	4623	6899	8005	81	42	276	2333	-44.4	62.8
v_H	56.5/58.3	4633	6918	7996	81	42	268	2419*	-50.0	72.7*
v_D	56.5/58.5	4635	6950	7991	81	42	436*	2122	-13.8	52.3
v_{01}	56.8/58.4	4630	6909	7999	81	42	294	2158	-32.5	45.2
v_{21}	56.6/58.3	4631	6899	7999	81	42	274	2343	-45.3	64.3
v_{02}	56.5/58.4	4629	6871	8002	81	42	287	2211	-36.6	50.4
v_{03}	56.8/58.4	4632	6909	7997	81	42	294	2159	-32.6	45.4

Section IV (continued)

v_{23}	56.6/58.3	4633	6899	7997	81	42	273	2346	-45.5*	64.7
v_{04}	56.5/58.5	4631	6871	8000	81	42	286	2212	-16.7	50.6
m_1	56.4/57.8	4594	7087	8009	81	41	305	1778	-9.5	11.2
m_2	56.4/57.8	4601	7065	8008	81	42	339	1687	-7.8	11.0
m_3	55.8/58.0	4603	7069	8003	81	42	314	1706	-8.9	11.4
m_4	55.8/57.8	4604	7065	8002	81	42	309	1711	-9.2	11.2

Section V

v_0	93.7/94.7	3697	2922	384	31	101	.02	106	.27	.20
v_2	94.0/94.9	3701	3110	430	33	101	.15	105	.27	.19
v_H	94.2/95.2	3709	3171	485*	34	101	.19	105	.28	.19
v_D	94.2/95.2	3704	3160	454	33	101	.76	105	.31	.19
v_{01}	93.8/94.7	3703	2922	417	31	101	.02	106	.27	.20
v_{21}	94.0/95.0	3707	3110	463	33	101	.15	105	.27	.19
v_{02}	94.0/94.7	3704	2956	425	31	101	.07	105	.27	.19
v_{03}	93.8/94.8	3704	2922	426	31	101	.02*	106	.27	.20
v_{23}	94.0/95.0	3708	3110	472	33	101	.15	105	.27	.19
v_{04}	94.0/94.7	3705	2956	434	31	101	.07	105	.27	.19
m_1	93.8/94.9	3687	2903	330	30	101	1.65	107	.38	.20
m_2	93.9/94.9	3688	2900	333	30	101	1.63	106	.38	.20
m_3	94.3/94.7	3698	3093	413	32	101	1.75*	105	.37	.19
m_4	94.3/94.8	3699	3100	416	33	101	1.74*	105	.37	.19

N.B. Five sets of choices of $\underline{A}$ and $\underline{B}$ are as follows :

$\underline{A}_1$ = (5.02, 4.99, 4.65, 5.44, 5.10, 5.06, 4.90, 5.04, 5.50, 5.35, 5.35, 4.85, 5.30, 4.51, 5.45, 4.80, 5.37, 4.68, 4.77, 4.56)

$\underline{B}_1$ = (-0.66, -0.76, -1.74, -0.58, -0.63, -0.41, -1.17, -0.93, -0.43, -1.30, -0.18, -0.06, -0.74, -1.77, -0.54, -0.44, -0.36, -1.94,

$\theta_a = 5.03$, $\sigma_a^2 = .09$; $\theta_b = -.94$, $\sigma_b^2 = .32$.

$\underline{A}_2$ = (5.90, 6.40, 5.26, 5.60, 6.42, 4.57, 5.45, 4.53, 5.53, 5.42, 5.11, 5.69, 5.86, 4.60, 6.36, 4.98, 6.29, 5.71, 4.85, 4.59)

$\underline{B}_2$ = (0.60, 0.55, -1.04, -1.95, -1.78, 0.56, 0.01, 1.24, 1.78, 0.98, -0.38, 1.58, -1.09, 0.56, 1.66, -1.61, 1.36, -1.56, 1.51, 0.21, 0.23, 0.19, -0.85, 0.13, 0.09)

$$\theta_a = 5.46,\ \sigma_a^2 = .38;\ \theta_b = .12,\ \sigma_b^2 = 1.25.$$

$\underline{A}_3$ = (9.21, 7.31, 8.83, 7.42, 10.49, 8.60, 5.59, 9.91, 7.39, 5.20, 8.74, 4.50, 8.67, 4.98, 7.34, 5.56, 7.72, 6.58, 5.08, 7.19)

$\underline{B}_3$ = (0.58, -0.27, 1.88, -1.08, -0.32, -0.54, -1.79, 0.90, 1.58, -1.22, -1.40, 0.36, -0.18, -1.21, 0.01, -1.23, 1.87, 0.09, -1.26, 0.38, 1.46, 1.17, 0.69, -1.59, -1.01)

$$\theta_a = 7.32,\ \sigma_a^2 = 2.89;\ \theta_b = -.09,\ \sigma_b^2 = 1.25.$$

$\underline{A}_4$ = (0.28, 5.15, 8.68, 8.05, 0.60, 3.63, 9.35, 0.96, 9.16, 3.45, 4.49, 7.51, 6.40, 2.13, 8.48, 2.20, 4.30, 8.70, 1.96, 3.06)
-0.22, -1.62, -1.17, -1.32, -1.66, -1.13 -1.84)

$\underline{B}_4$ = (-3.46, -4.43, -4.43, -2.10, 1.77, -2.49, 4.60, 0.19, 3.44, -2.77, 2.41, -1.80, -2.31, 0.27, 3.28, 3.57, 3.67, 3.72, -2.50, -2.22, 2.94, -3.69, 4.24, -2.32, -2.86)

$$\theta_a = 5.13,\ \sigma_a^2 = 8.15;\ \theta_b = -.13,\ \sigma_b^2 = 9.42.$$

$\underline{A}_5$ consists of all entries as 1.00 and $\underline{B}_5$ consists of all entries as 0.00 — i.e. this set corresponds to DR rather than RR and is considered for comparison with RR.

Comments: On applying badly designed RR techniques, for example with choices of $\underline{A_3}$, $\underline{B_3}$ or $\underline{A_4}$, $\underline{B_4}$ one cannot get results comparable to those available with DR if applicable. But if RR is properly implemented, for example, if $\underline{A_1}$, $\underline{B_1}$ or $\underline{A_2}$, $\underline{B_2}$ may be employed, the RR technique yields serviceable results. At any rate m_1, m_2 turn out the best in all the five situations illustrated.

CHAPTER SIX

INFERENCE IN RANDOMIZED RESPONSE SURVEYS WITH COMPLEX STRATEGIES

6.1 SUMMARY.

Modifications on the generalized regression predictor for Y and on its variance estimators discussed in Chapter Two are presented here when instead of direct responses only randomized responses are available essentially in the manners described in Chapter Five. Comparative efficacies of competing confidence intervals based on these modified estimators and variance estimators are examined numerically through simulation studies.

6.1 INTRODUCTION.

In this chapter also we regard y as a sensitive variable and consequently y_i values are unavailable but instead RR's are available as, say, r_i from each sampled individual adopting a suitable device as discussed in Chapter Five. Suppose, more generally than reported in Chapter Five, r_i be available satisfying

$$E_R(r_i) = y_i$$

and, $$V_R(r_i) = \alpha_i y_i^2 + \beta_i y_i + \theta_i = V_i, \text{ say, } i \in U, \quad (6.1.1)$$

with α_i, β_i, θ_i known. One possibility to get such an r_i as illustrated in Chaudhuri and Mukerjee (1988) is as follows. Each sampled individual i may be requested to report the true value y_i with a pre-assigned probability c $(0<c<1)$ and with a probability $(1-c)$ to report a value chosen out of a large number, say, K of given real values $(z_1,\ldots,z_K)$. Then the randomized response, say, w_i would satisfy

$$E_R(w_i) = c\ y_i + \frac{(1-c)}{K} \sum_{j=1}^{K} z_j = c\ y_i + Q, \text{ say.}$$

Then $r_i = (w_i - Q)/\ c$ would meet the requirements (6.1.1). It will then follow that

$$\hat{V}_i = \frac{1}{1+\alpha_i} (\alpha_i r_i^2 + \beta_i r_i + \theta_i)$$

satisfies, $E_R(\hat{V}_i) = V_i,\ i \in U$.

If DR were available, an appropriate estimator for Y under model $\underline{M}$ *of the earlier chaprters would be the well-known greg predictor of* Särndal (1980), namely,

$$t_g = \sum{}' \frac{y_i}{\pi_i} g_{si}$$

where, $$g_{si} = 1 + \left(X - \sum{}' \frac{x_k}{\pi_k} \right) \frac{x_i Q_i \pi_i}{\Sigma' x_k^2 Q_k}$$

taking Q_i as suitable positive numbers as mentioned in chapters One, Two and Four.

In the next section we consider a version of t_g when y_i is replaced by r_i and variance estimators *of the latter in terms of* r_i, $i \in s$ on adjusting the variance estimators of t_g in terms of y_i given in Chapter Two. Next we derive confidence intervals of Y and examine their relative efficacies through a simulation study. Again linear regression model and Brewer's (1979) asymptotic approach are used.

6.2 ESTIMATORS AND VARIANCE ESTIMATORS IN RR SET-UP.

Replacing y_i by r_i throughout in t_g our proposed obvious modification on t_g is

$$t_g(r) = \sum{}' \frac{r_i}{\pi_i} g_{si}.$$

Naturally, $E_R[t_g(r)] = t_g$. Noting that, $E_R[t_g(r)-Y]^2 = (t_g-Y)^2 + E_R(t_g(r)-t_g)^2 = (t_g-Y)^2 + \sum{}' V_i \left(\frac{g_{si}}{\pi_i}\right)^2$ and writing E_G for the expectation operator as E_p or $\lim E_p$ or $\lim E_p E_m$ which is used to choose a measure of error of t_g for Y as $E_G(t_g-Y)^2$, we naturally take

$$m_R = E_G E_R (t_g(r)-Y)^2$$

as a measure of error of $t_g(r)$ as an estimator of Y.

We note that $\lim E_p \sum' V_i \left(\frac{g_{si}}{\pi_i}\right)^2 = \sum \frac{V_i}{\pi_i}$ and so take

$$v(r) = \sum' \hat{V}_i/\pi_i^2 \quad \text{and} \quad v'(r) = \sum' \hat{V}_i \left(\frac{g_{si}}{\pi_i}\right)^2$$

as estimators for $\sum \frac{V_i}{\pi_i}$ because we have

$$\lim E_p E_R [v(r)] = \sum \frac{V_i}{\pi_i} = \lim E_p E_R [v'(r)].$$

Next in each of v_j, K_j (j=1,2) and m_j (j=1,..,6) we replace y_i throughout by r_i and make adjustments on them so as to derive a formula v', say, for which we may have $E_R(v') = v$ where v stands for these 10 variance estimators in turn. For variance estimators of $t_g(r)$ then we take $m'=v'+v(r)$ and $m''=v'+ v'(r)$. Denoting v' respectively corresponding to v_1, v_2, K_1, K_2, $m_1,\dots,m_6$ by v'_1, $v'_2,\dots,v'_{10}$ we have then the following 20 alternative variance estimators for $t_g(r)$, namely, $m'_j=v'_j+v(r)$, j=1,..,10 and $m''_j=v'_j+v'(r)$, j=1,..,10. These v'_j, j=1,..10 work out respectively, writing $e_i(r)$ for e_i replacing y_i in latter by r_i, as

$$v'_1 = \sum\sum' \Delta_{ij}\left[\left(\frac{e_i(r)}{\pi_i} - \frac{e_j(r)}{\pi_j}\right)^2 - a_{ij}\right], \text{ where}$$

$$a_{ij} = \left(\frac{\hat{V}_i}{\pi_i^2} + \frac{\hat{V}_j}{\pi_j^2}\right) + \left(\frac{x_i}{\pi_i} - \frac{x_j}{\pi_j}\right)^2 \left[\frac{\Sigma' x_i^2 Q_i^2 \hat{V}_i}{(\Sigma' x_i^2 Q_i)^2}\right]$$

$$- \frac{2}{\Sigma' x_i^2 Q_i}\left(\frac{x_i}{\pi_i} - \frac{x_j}{\pi_j}\right)\left[\frac{x_i Q_i \hat{V}_i}{\pi_i} - \frac{x_j Q_j \hat{V}_j}{\pi_j}\right]$$

$$v'_2 = \sum\sum' \Delta_{ij}\left[\left(\frac{g_{si} e_i(r)}{\pi_i} - \frac{g_{sj} e_j(r)}{\pi_j}\right)^2 - b_{ij}\right], \text{ where}$$

$$b_{ij} = \left(\frac{g_{si}^2 \hat{V}_i}{\pi_i^2} + \frac{g_{sj}^2 \hat{V}_j}{\pi_j^2}\right) + \left(\frac{g_{si} x_i}{\pi_i} - \frac{g_{sj} x_j}{\pi_j}\right)^2 \left[\frac{\Sigma' x_i^2 Q_i^2 \hat{V}_i}{(\Sigma'_i x^2_i Q\ )^2}\right]$$

$$- \frac{2}{\Sigma' x_i^2 Q_i}\left(\frac{g_{si}x_i}{\pi_i} - \frac{g_{sj}x_j}{\pi_j}\right)\left[\frac{g_{si}x_i\, Q_i\, \hat{V}_i}{\pi_i} - \frac{g_{sj}x_j\, Q_j\, \hat{V}_j}{\pi_j}\right]$$

$$v_3' = \frac{E_m(t_g-Y)^2}{E_mE_R(v_1')}\, v_1' \,, \quad v_4' = \frac{E_m(t_g-Y)^2}{E_mE_R(v_2')}\, v_2'$$

$$v_5' = \frac{n}{n-2}\sum{}'\left[\frac{x_i^2}{\pi_i^2}(1-\pi_i) - \frac{1}{n(n-1)}\sum\frac{x_k^2}{\pi_k}(1-\pi_k)\right]\left[(w_i - \bar{w})^2 - a_i'\right]$$

where, $a_i' = \frac{n-2}{n}\,\frac{\hat{V}_i}{x_i^2} + \frac{1}{n^2}\sum{}'\frac{\hat{V}_k}{x_k^2}$,

$$v_6' = \frac{n}{n-2}\sum{}'\left[\frac{x_i^2}{\pi_i^2}(1-\pi_i) - \frac{1}{n(n-1)}\sum{}'\frac{x_k^2}{\pi_k^2}(1-\pi_k)\right]\left[(w_i - \bar{w})^2 - a_i'\right]$$

$$v_7' = \frac{\frac{n-2}{n-1}\sum x_k^2\,\frac{1-\pi_k}{\pi_k}}{\sum{}' x_i^2\,\frac{1-\pi_i}{\pi_i^2} - \frac{1}{n-1}\sum x_k^2\,\frac{1-\pi_k}{\pi_k}}\, v_5' \,,$$

$$v_8' = \frac{\sum x_k^2\,\frac{1-\pi_k}{\pi_k}}{\sum{}' x_i^2\,\frac{1-\pi_i}{\pi_i^2}}\, v_6' \,,$$

$$v_9' = \frac{\Sigma f_i(1/\pi_i - 1)}{\frac{n}{n-1}\,\Sigma f_i \pi_i / x_i^2}\sum{}'\left[\, (w_i - \bar{w})^2 - a_i' \,\right]$$

and, $$v_{10}' = \frac{\Sigma f_i(1/\pi_i - 1)}{\frac{n}{n-1}\,\Sigma' f_i / x_i^2}\sum{}'\left[\, (w_i - \bar{w})^2 - a_i' \,\right] ,$$

Since obviously it is not easy to choose from among the m_j' and m_j'', $j=1,..,10$ by analytical considerations, we consider examining their efficacies in yielding confidence intervals (CI) for Y of the form $t_g(r) \pm \tau_{\alpha/2}\sqrt{\hat{v}}$, where $\hat{v}$ *stands for one of these* m_j', m_j'', $j=1,..,10$. Here, for large n, the distribution of $d = \left(t_g(r) - Y\right)/\sqrt{\hat{v}}$ is supposed to be approximately that of the standard normal variable τ and $\tau_{\alpha/2}$, for a chosen α in (0,1) is supposed to be the $100\alpha/2$ percent point on the right tail area of the N(0,1) distribution. Performances of the respective CI's are examined by us through a simulation study described below.

6.3 SIMULATION STUDY.

As in earlier chapters, we take N=150, $\sigma_i^2 = \sigma^2 x_i^g$, σ=1, g=1.2, β=1, draw x_i's at random from the density

$$f(t) = \frac{1}{\lambda} e^{-t/\lambda}, \quad t>0, \quad \lambda=8.5,$$

draw τ_i's at random from N(0,1), take $\varepsilon_i = x_i^{g/2}\tau_i$ and then generate y_i subject to $\underline{M}(f)$ with these stipulations. We take 5 sets of $\underline{A}$, $\underline{B}$ denoted I—$\overline{V}$ and given below, with means θ_a, θ_b and variances σ_a^2, σ_b^2 and apply RR device of Chapter Five.

$\underline{A}_1$ = (5.90, 6.40, 5.26, 5.60, 6.42, 4.57, 5.45, 4.53, 5.53, 5.42, 5.11, 5.69, 5.86, 4.60, 6.36, 4.98, 6.29, 5.71, 4.85, 4.59)

$\underline{B}_1$ = (0.60, 0.55, −1.04, −1.95, −1.78, 0.56, 0.01, 1.24, 1.78, 0.98, −0.38, 1.58, −1.09, 0.56, 1.66, −1.61, 1.36, −1.56, 1.51, 0.21, 0.23, 0.19, −0.85, 0.13, 0.09)

$\theta_a = 5.46$, $\sigma_a^2 = .38$; $\theta_b = .12$, $\sigma_b^2 = 1.25$.

$\underline{A}_2$ = (0.28, 5.15, 8.68, 8.05, 0.60, 3.63, 9.35, 0.96, 9.16, 3.45, 4.49, 7.51, 6.40, 2.13, 8.48, 2.20, 4.30, 8.70, 1.96, 3.06)

/ −0.22, −1.62, −1.17, −1.32, −1.66, −1.13 −1.84)

$\underline{B}_2$ = (−3.46, −4.43, −4.43, −2.10, 1.77, −2.49, 4.60, 0.19, 3.44, −2.77, 2.41, −1.80, −2.31, 0.27, 3.28, 3.57, 3.67, 3.72, −2.50, −2.22, 2.94, −3.69, 4.24, −2.32, −2.86)

$\theta_a = 5.13$, $\sigma_a^2 = 8.15$; $\theta_b = -.13$, $\sigma_b^2 = 9.42$.

$\underline{A}_4$ = (5.02, 4.99, 4.65, 5.44, 5.10, 5.06, 4.90, 5.04, 5.50, 5.35, 5.35, 4.85, 5.30, 4.51, 5.45, 4.80, 5.37, 4.68, 4.77, 4.56)

$\underline{B}_4$ = (−0.66, −0.76, −1.74, −0.58, −0.63, −0.41, −1.17, −0.93, −0.43, −1.30, −0.18, −0.06, −0.74, −1.77, −0.54, −0.44, −0.36, −1.94, $\theta_a = 5.03$, $\sigma_a^2 = .09$; $\theta_b = -.94$, $\sigma_b^2 = .32$.

$\underline{A}_5$ = (9.21, 7.31, 8.83, 7.42, 10.49, 8.60, 5.59, 9.91, 7.39, 5.20, 8.74, 4.50, 8.67, 4.98, 7.34, 5.56, 7.72, 6.58, 5.08, 7.19)

$\underline{B}_5$ = (0.58, -0.27, 1.88, -1.08, -0.32, -0.54, -1.79, 0.90, 1.58, -1.22, -1.40, 0.36, -0.18, -1.21, 0.01, -1.23, 1.87, 0.09, -1.26, 0.38, 1.46, 1.17, 0.69, -1.59, -1.01)

$$\theta_a = 7.32,\ \sigma_a^2 = 2.89;\ \theta_b = -.09,\ \sigma_b^2 = 1.25.$$

$\underline{A}_3$ consists of all entries as 1.00 and $\underline{B}_3$ consists of all entries as 0.00 — i.e. this set corresponds to DR rather than RR and is considered for comparison with RR.

Next, taking n=32, adopting Hartley and Rao's (HR, 1962) scheme of sampling using size measures as x_1^g, R=1000 replicates of samples are chosen. For each sample, values of d are calculated, CI's as $t_g(r) \pm 1.96\sqrt{\hat{v}}$, choosing α=0.05, i.e. with nominal confidence coefficients of 95% are computed. To discriminate among the CI's we consider the similar criteria following Rao and Wu (1983) among others, as in earlier chapters. We shall denote by Σ_r the sum over the R=1000 replicates of samples, $\frac{1}{R}\Sigma_r\hat{v}$ by A and pseudo mean square error by $P = \frac{1}{R}\Sigma_r[t_g(r)-Y]^2$.

(1) ACP (Actual coverage percentage) ≡ the number of samples out of R for which CI's cover Y — the closer it is to 95 the better the procedure.

(2) ACV (Average coefficient of variation) ≡ the average of $\sqrt{\hat{v}}/t_g(r)$ over the R replicates — this reflects the length of CI relative to $t_g(r)$.

To choose among the $\hat{v}$'s we further consider the criteria

(3) PB (Pseudo relative bias) : $B(v) = (\frac{1}{R}\Sigma_r\hat{v} - P)/P$.

(4) PS (Pseudo relative stability) : $S(v) = \left[\frac{1}{R} \Sigma_r (\hat{v}-P)^2\right]^{1/2} / P.$

(5) PL (Pseudo standardized length) : $L(v) = \frac{1}{R} \Sigma_r \sqrt{\hat{v}} / \sqrt{P}.$

(6) B(d) (Bias of d) $= \frac{1}{R} \Sigma_r d.$

(7) M(d) (Mean square error (MSE) of d) $= \frac{1}{R} \Sigma_r [d-B(d)]^2.$

(8) Root beta one $: \sqrt{\beta_1(d)} = \frac{1}{R} \sum_r \left(\frac{d-B(d)}{\sqrt{M(d)}}\right)^3.$

(9) Excess measure : $E(d) = \beta_2(d)-3 = \frac{1}{R} \sum_r \left(\frac{d-B(d)}{\sqrt{M(d)}}\right)^4 - 3.$

(10) PCV (Pseudo coefficient of variation) :

$$PCV(v) = \left[\frac{1}{R} \Sigma_r (\dot{v}-A)^2\right]^{1/2} / A.$$

The smaller the magnitudes of criteria 2 — 10, the better. The values are presented in the table below separately for 4 choices of Q_i as $\frac{1-\pi_i}{\pi_i x_i}$, $\frac{1}{\pi_i x_i}$ respectively suggested by Brewer (1979) and Hájek (1971), $1/x_i^2$ and $1/x_i$, denoted respectively by B, H, S and S' in table. Noting that both $\sqrt{\beta_1(d)}$ and E(d) often differ from 0, indicating deviation from normality of d, we also calculate CI as $CI' = t_g(r) \pm t_{0.95,(n-1)} \sqrt{\hat{v}}$ to calculate the ACP. Here $t_{0.95,(n-1)}$ denotes the 95% point on the right tail area of the Student's t distribution with (n-1) degrees of freedom which may approximate the distribution of d better than N(0,1). In the table below in Appendix-F, we give values for m_j', j=1,..,10 and those for m_j'', j=1,..,10, are similar but not shown. And, in the tables m_j''s are represented by m_j's.

6.4 CONCLUDING REMARKS.

Obviously all the procedures fare best in case III, i.e. when DR is available. Among the RR's, the case II and IV with smaller σ_a^2 fare better than others. In each case I - V, the variance estimators m_5 and m_6 fare best among the other competitors in our numerical illustrations.

Appendix F

The acronyms used in the tables are as explained on pp 92-93. To avoid monotony and save space we do not repeat them here. However, $B, H, S, S^{/}$ correspond respectively to choice of Q_i as $(1-\pi_i)/\pi_i x_i$, $1/\pi_i x_i$, $1/x_i^2$ and $1/x_i$.

TABLE

Relative performances of confidence intervals.

5 sections I-V represent respectively 5 choices I-V of A, B, given below. 4 sub-sections marked 1-4 in each section respectively correspond to 4 choices of Q_i as B, H, S and $S^{/}$. ACP for Student's t-table are given after a slash following those for N(0,1).

		10^5	10^4	10^3	10^2	10^2	10^4		10	
^ v	ACP	ACV	PCV	PB	PS	PL	B(d)	M(d)	$\sqrt{\beta_1}$	E(d)
Section I : Sub-section 1										
m_1	94.8/95.2	5575	5140	-74	48	94	37	1.12	.80	-.29
m_2	97.0/97.6	6219	5185	15	62	104	51	.89	.93	-.33
m_3	94.7/95.1	5575	5211	-73	49	94	41	1.12	.84	-.30
m_4	97.0/97.6	6215	5184	15	62	104	51	.89	.93	-.33
m_5	94.8/95.3	5569	5107	-77	48	93	33	1.12	.78	-.29
m_6	94.8/95.3	5569	5109	-77	48	93	33	1.12	.78	-.29
m_7	94.7/95.1	5571	5158	-75	48	94	37	1.12	.81	-.30
m_8	94.7/95.1	5571	5158	-75	48	94	37	1.12	.81	-.30
m_9	97.2/97.7	6324	5679	20	71	106	108	.87	1.14	-.26
m_{10}	97.2/97.8	6323	5749	20	72	106	112	.87	1.18	-.27
Section I : Sub-section 2										
m_1	94.8/95.2	5574	5138	-75	48	94	37	1.12	.80	-.29
m_2	97.0/97.6	6218	5181	15	62	104	51	.89	.93	-.33
m_3	94.7/95.1	5575	5209	-73	49	94	41	1.12	.84	-.30
m_4	97.0/97.6	6216	5181	15	61	104	51	.89	.93	-.33
m_5	94.8/95.3	5569	5107	-77	48	93	34	1.12	.78	-.29
m_6	94.7/95.3	5569	5109	-77	48	93	34	1.12	.78	-.29
m_7	94.7/95.1	5571	5158	-75	48	93	37	1.12	.81	-.30
m_8	94.7/95.1	5571	5158	-75	48	93	37	1.12	.81	-.30
m_9	97.2/97.7	6324	5679	20	71	106	108	.87	1.15	-.26
m_{10}	97.2/97.8	6323	5749	20	72	106	112	.87	1.18	-.27

Section I : Sub-section 3										
m_1	94.8/95.2	5576	5142	-74	48	94	36	1.11	.80	-.29
m_2	97.1/97.6	6220	5191	15	62	104	51	.89	.93	-.33
m_3	94.7/95.1	5575	5214	-73	49	94	42	1.12	.84	-.29
m_4	97.1/97.6	6215	5191	15	62	104	52	.89	.93	-.33
m_5	94.8/95.3	5569	5107	-77	48	93	33	1.12	.78	-.29
m_6	94.8/95.3	5569	5109	-77	48	93	33	1.12	.78	-.29
m_7	94.7/95.1	5571	5158	-75	48	94	37	1.12	.81	-.30
m_8	94.7/95.1	5571	5158	-75	48	94	37	1.12	.81	-.30
m_9	97.2/97.7	6324	5679	20	71	106	107	.87	1.14	-.26
m_{10}	97.2/97.8	6323	5749	20	72	106	112	.87	1.18	-.27

Section I : Sub-section 4										
m_1	94.7/95.2	5571	5135	-76	48	93	37	1.12	.80	-.29
m_2	97.0/97.6	6214	5167	15	61	104	52	.89	.93	-.33
m_3	94.7/95.1	5576	5203	-74	49	94	42	1.12	.83	-.30
m_4	97.0/97.6	6217	5168	15	61	104	52	.89	.93	-.33
m_5	94.7/95.3	5569	5107	-77	48	93	35	1.12	.79	-.29
m_6	94.7/95.2	5569	5109	-77	48	93	35	1.12	.79	-.29
m_7	94.6/95.1	5571	5158	-76	48	93	39	1.12	.81	-.30
m_8	94.6/95.1	5571	5158	-76	48	93	39	1.12	.82	-.30
m_9	97.1/97.8	6323	5679	20	71	106	109	.87	1.15	-.27
m_{10}	97.2/97.8	6323	5749	20	72	106	113	.88	1.18	-.27

Section II : Sub-section 1										
m_1	94.5/95.5	4708	3688	-56	35	96	-87	1.12	.14	-.06
m_2	94.7/95.7	4723	3821	-48	37	96	-81	1.12	.15	-.07
m_3	94.6/95.6	4708	3824	-54	37	96	-81	1.12	.15	-.07
m_4	94.7/95.7	4720	3821	-50	37	96	-81	1.12	.15	-.07
m_5	94.3/95.5	4701	3645	-60	35	96	-92	1.13	.14	-.06
m_6	94.3/95.5	4701	3648	-59	35	96	-92	1.13	.14	-.06
m_7	94.6/95.5	4703	3739	-57	36	96	-87	1.12	.15	-.07
m_8	94.6/95.5	4703	3740	-57	36	96	-87	1.12	.15	-.07
m_9	94.8/96.0	4820	4228	-4	42	98	17	1.09	.19	.02
m_{10}	94.8/95.8	4819	4333	-3	43	98	19	1.09	.20	.00

Section II : Sub-section 2										
m_1	94.5/95.6	4707	3687	-57	35	96	-87	1.12	.14	-.06
m_2	94.7/95.7	4722	3816	-49	37	96	-81	1.12	.15	-.07
m_3	94.5/95.6	4708	3821	-54	37	96	-81	1.12	.15	-.07
m_4	94.7/95.7	4720	3816	-50	37	96	-81	1.12	.15	-.07
m_5	94.3/95.5	4701	3645	-60	35	96	-91	1.13	.14	-.06
m_6	94.3/95.5	4701	3648	-60	35	96	-91	1.13	.14	-.06
m_7	94.6/95.5	4703	3739	-57	36	96	-87	1.13	.15	-.07
m_8	94.6/95.5	4703	3740	-57	36	96	-87	1.13	.15	-.07
m_9	94.8/96.0	4820	4228	-4	42	98	17	1.09	.19	.02
m_{10}	94.8/95.8	4819	4333	-2.6	43	98	19	1.09	.20	.00

Section II : Sub-section 3										
m_1	94.5/95.5	4710	3690	-55	35	96	-87	1.12	.14	-.06
m_2	94.8/95.7	4725	3830	-47	37	96	-80	1.12	.15	-.07
m_3	94.6/95.5	4708	3828	-54	37	96	-81	1.12	.15	-.07
m_4	94.6/95.7	4720	3830	-49	37	96	-81	1.12	.15	-.07
m_5	94.3/95.5	4701	3645	-59	35	96	-92	1.13	.14	-.06
m_6	94.3/95.5	4701	3648	-59	35	96	-92	1.12	.14	-.06
m_7	94.6/95.5	4703	3739	-57	36	96	-87	1.12	.14	-.07
m_8	94.6/95.5	4703	3740	-57	36	96	-87	1.12	.14	-.07
m_9	94.8/96.0	4820	4228	-3.8	42	98	16	1.09	.19	.02
m_{10}	94.9/95.8	4819	4333	-2.4	43	98	19	1.08	.20	.00

Section II : Sub-section 4										
m_1	94.5/95.5	4703	3685	-59	35	96	-86	1.13	.14	-.06
m_2	94.6/95.6	4717	3797	-51	36	96	-81	1.12	.15	-.07
m_3	94.5/95.6	4708	3813	-55	36	96	-81	1.13	.15	-.07
m_4	94.6/95.7	4720	3798	-50	36	96	-81	1.12	.15	-.06
m_5	94.3/95.5	4701	3645	-60	35	95	-90	1.13	.14	-.06
m_6	94.3/95.5	4701	3648	-60	35	95	-90	1.13	.14	-.07
m_7	94.5/95.5	4703	3739	-58	36	96	-85	1.13	.15	-.07
m_8	94.5/95.5	4703	3740	-58	36	96	-85	1.13	.15	-.07
m_9	94.8/95.9	4820	4228	-42	42	98	19	1.09	.19	.16
m_{10}	95.0/95.7	4819	4333	-28	43	98	21	1.09	.20	.49

Section III : Sub-section 1

m_1	94.0/95.2	4703	3638	-61	35	95	-28	1.13	.14	-.07
m_2	94.0/95.2	4707	3776	-58	36	95	-22	1.13	.14	-.07
m_3	94.0/95.2	4703	3773	-59	36	95	-22	1.13	.14	-.07
m_4	94.0/95.2	4703	3776	-59	36	95	-22	1.13	.14	-.07
m_5	93.9/95.1	4696	3596	-65	34	95	-32	1.13	.13	-.07
m_6	93.9/95.1	4696	3598	-65	34	95	-32	1.13	.13	-.07
m_7	94.1/95.2	4699	3688	-62	35	95	-28	1.13	.14	-.07
m_8	94.1/95.2	4699	3689	-62	35	95	-28	1.13	.14	-.07
m_9	94.6/95.8	4805	4197	-13	41	97	78	1.10	.19	.02
m_{10}	94.7/95.6	4804	4299	-12	42	97	79	1.10	.19	.01

Section III : Sub-section 2

m_1	94.0/95.2	4702	3637	-62	35	95	-28	1.13	.14	-.07
m_2	94.0/95.2	4705	3771	-59	36	95	-22	1.13	.14	-.07
m_3	94.0/95.2	4703	3770	-59	36	95	-22	1.13	.14	-.07
m_4	94.0/95.2	4703	3770	-59	36	95	-22	1.13	.14	-.07
m_5	93.9/95.1	4696	3596	-65	34	95	-32	1.13	.13	-.07
m_6	93.9/95.1	4696	3598	-65	34	95	-32	1.13	.13	-.07
m_7	94.1/95.2	4699	3688	-62	35	95	-28	1.13	.14	-.07
m_8	94.1/95.2	4699	3689	-62	35	95	-29	1.13	.14	-.07
m_9	94.6/95.8	4805	4197	-13	41	97	78	1.10	.19	.02
m_{10}	94.7/95.6	4804	4299	-12	42	97	80	1.10	.19	.01

Section III : Sub-section 3

m_1	94.0/95.2	4705	3640	-60	35	95	-28	1.13	.14	-.07
m_2	94.0/95.3	4709	3784	-57	36	96	-22	1.13	.14	-.07
m_3	94.0/95.2	4703	3777	-59	36	95	-22	1.13	.14	-.07
m_4	94.0/95.2	4703	3784	-59	36	95	-22	1.13	.14	-.07
m_5	93.9/95.1	4696	3596	-65	34	95	-33	1.13	.13	-.07
m_6	93.9/95.1	4696	3598	-64	34	95	-33	1.13	.13	-.07
m_7	94.1/95.3	4699	3688	-62	35	95	-29	1.13	.14	-.07
m_8	94.1/95.3	4699	3689	-62	35	95	-29	1.13	.14	-.07
m_9	94.6/95.8	4805	4197	-13	41	97	77	1.10	.19	.02
m_{10}	94.7/95.6	4804	4299	-12	43	97	79	1.10	.19	.01

Section III : Sub-section 4												
m_1	94.0/95.1	4698	3635	-63	35	95	-27	1.13	.14	-.06		
m_2	94.0/95.2	4701	3752	-61	36	95	-23	1.13	.14	-.07		
m_3	94.0/95.2	4703	3762	-60	36	95	-22	1.13	.14	-.07		
m_4	94.0/95.2	4703	3753	-60	36	95	-23	1.13	.14	-.07		
m_5	93.9/95.1	4696	3596	-65	34	95	-30	1.13	.13	-.07		
m_6	93.9/95.1	4696	3598	-65	34	95	-30	1.13	.13	-.07		
m_7	94.0/95.2	4699	3688	-63	35	95	-26	1.13	.14	-.07		
m_8	94.0/95.2	4699	3689	-62	35	95	-26	1.13	.14	-.07		
m_9	94.6/95.8	4804	4197	-13	41	97	80	1.10	.19	.02		
m_{10}	94.8/95.6	4803	4299	-12	42	97	81	1.10	.19	.01		

Section IV : Sub-section 1										
m_1	94.5/95.4	4740	3737	-52	36	96	4.63	1.12	.13	-.08
m_2	94.5/95.7	4784	3855	-32	37	97	10.07	1.10	.14	-.09
m_3	94.2/95.3	4739	3868	-50	37	96	9.81	1.12	.14	-.09
m_4	94.5/95.7	4780	3855	-34	37	97	10.12	1.10	.14	-.09
m_5	94.5/95.4	4733	3695	-55	35	96	-.06	1.12	.13	-.08
m_6	94.5/95.4	4733	3698	-55	35	96	.10	1.12	.13	-.08
m_7	94.3/95.4	4735	3786	-53	36	96	4.10	1.12	.13	-.09
m_8	94.3/95.4	4735	3786	-53	36	96	4.13	1.12	.13	-.09
m_9	95.0/96.1	4881	4272	12	43	99	106.24	1.07	.19	.00
m_{10}	95.0/96.0	4880	4376	14	44	99	108.24	1.07	.19	-.01

Section IV : Sub-section 2										
m_1	94.5/95.4	4739	3736	-52	36	96	4.58	1.12	.13	-.08
m_2	94.5/95.7	4782	3850	-33	37	97	9.85	1.10	.14	-.10
m_3	94.2/95.3	4740	3866	-50	37	96	9.63	1.12	.14	-.09
m_4	94.3/95.7	4780	3849	-34	37	97	9.87	1.10	.14	-.10
m_5	94.5/95.4	4733	3695	-55	35	96	.22	1.12	.13	-.08
m_6	94.5/95.4	4733	3698	-55	35	96	.37	1.12	.13	-.08
m_7	94.3/95.4	4735	3786	-53	36	96	4.34	1.12	.13	-.09
m_8	94.3/95.4	4735	3786	-53	36	96	4.36	1.12	.13	-.09
m_9	95.0/96.1	4881	4272	12	43	99	106.51	1.07	.19	.00
m_{10}	95.0/96.0	4880	4376	14	44	99	108.46	1.07	.19	-.01

Section IV : Sub-section 3

m_1	94.5/95.4	4741	3739	-51	36	96	4.87	1.12	.13	-.08
m_2	94.6/95.8	4786	3864	-31	38	97	10.57	1.10	.14	-.10
m_3	94.2/95.3	4739	3872	-50	37	96	10.26	1.12	.14	-.09
m_4	94.5/95.8	4780	3863	-33	37	97	10.67	1.10	.14	-.10
m_5	94.5/95.4	4733	3695	-55	35	96	-.32	1.12	.13	-.08
m_6	94.5/95.4	4733	3698	-55	35	96	-.16	1.12	.13	-.08
m_7	94.3/95.4	4735	3786	-52	36	96	3.90	1.12	.13	-.09
m_8	94.3/95.4	4735	3786	-52	36	96	3.92	1.12	.13	-.09
m_9	95.0/96.1	4881	4272	13	43	99	105.98	1.07	.19	.00
m_{10}	95.0/96.0	4880	4376	14	44	99	108.06	1.07	.19	-.01

Section IV : Sub-section 4

m_1	94.4/95.3	4735	3733	-54	36	96	4.72	1.12	.13	-.08
m_2	94.3/95.7	4778	3831	-35	37	97	9.26	1.10	.14	-.09
m_3	94.2/95.3	4740	3858	-50	37	96	9.35	1.12	.14	-.09
m_4	94.3/95.7	4780	3831	-34	37	97	9.22	1.10	.14	-.09
m_5	94.5/95.4	4733	3695	-55	35	96	1.44	1.12	.13	-.08
m_6	94.5/95.3	4733	3698	-55	35	96	1.59	1.12	.13	-.08
m_7	94.2/95.4	4735	3786	-53	36	96	5.42	1.12	.13	-.09
m_8	94.2/95.4	4735	3786	-53	36	96	5.44	1.12	.13	-.09
m_9	95.0/96.1	4881	4272	12	43	99	107.71	1.07	.19	.00
m_{10}	95.1/96.0	4880	4376	14	44	99	109.47	1.07	.19	-.01

Section V : Sub-section 1

m_1	94.1/95.3	4822	3966	-37	38	96	41	1.11	.14	.02
m_2	95.1/95.7	4949	4060	16	41	99	46	1.05	.15	-.01
m_3	94.2/95.2	4822	4077	-35	39	96	46	1.11	.15	.00
m_4	95.0/95.7	4946	4060	15	41	99	46	1.05	.15	-.01
m_5	94.0/95.1	4815	3928	-40	38	96	37	1.11	.14	.02
m_6	94.0/95.1	4815	3930	-40	38	96	37	1.11	.14	.02
m_7	94.2/95.2	4817	4005	-38	39	96	41	1.11	.14	.01
m_8	94.2/95.2	4818	4006	-38	39	96	41	1.11	.14	.01
m_9	95.2/95.7	5046	4500	62	48	101	134	1.02	.19	.07
m_{10}	95.2/95.8	5044	4582	63	49	101	136	1.02	.19	.04

				Section V : Sub-section 2						
m_1	94.1/95.3	4821	3965	-37	38	96	41	1.11	.14	.02
m_2	94.9/95.7	4948	4055	15	41	99	46	1.05	.15	-.01
m_3	94.2/95.2	4822	4075	-35	39	96	46	1.11	.15	.00
m_4	94.9/95.7	4946	4055	15	41	99	46	1.05	.15	-.01
m_5	94.0/95.1	4815	3928	-40	38	96	37	1.11	.14	.02
m_6	94.0/95.1	4815	3930	-40	38	96	37	1.11	.14	.02
m_7	94.2/95.2	4817	4005	-38	39	96	41	1.11	.14	.01
m_8	94.2/95.2	4817	4006	-38	39	96	41	1.11	.14	.01
m_9	95.2/95.7	5046	4500	62	48	101	134	1.02	.19	.07
m_{10}	95.2/95.8	5044	4582	63	49	101	136	1.02	.19	.04

				Section V : Sub-section 3						
m_1	94.2/95.3	4824	3967	-36	38	96	41	1.10	.14	.02
m_2	95.1/95.7	4951	4067	17	41	99	46	1.05	.15	-.01
m_3	94.2/95.2	4821	4081	-35	40	96	47	1.11	.15	.00
m_4	95.1/95.7	4945	4067	15	41	99	46	1.05	.15	-.01
m_5	94.0/95.1	4815	3928	-40	38	96	36	1.11	.13	.02
m_6	94.0/95.1	4815	3930	-40	38	96	37	1.11	.14	.02
m_7	94.2/95.2	4818	4005	-38	39	96	41	1.11	.14	.01
m_8	94.2/95.2	4818	4006	-38	39	96	41	1.11	.14	.01
m_9	95.2/95.7	5046	4500	62	48	101	133	1.02	.19	.07
m_{10}	95.2/95.8	5044	4582	63	49	101	135	1.02	.19	.04

				Section V : Sub-section 4						
m_1	94.1/95.2	4817	3962	-39	38	96	42	1.11	.14	.02
m_2	94.9/95.6	4944	4040	13	41	99	46	1.05	.15	.00
m_3	94.3/95.2	4822	4068	-36	39	96	47	1.11	.15	.01
m_4	94.9/95.6	4946	4040	14	41	99	46	1.05	.15	.00
m_5	94.0/95.2	4815	3928	-40	38	96	39	1.11	.14	.02
m_6	94.0/95.2	4815	3930	-40	38	96	39	1.11	.14	.02
m_7	94.2/95.2	4817	4005	-38	39	96	43	1.11	.14	.01
m_8	94.2/95.2	4817	4006	-38	39	96	43	1.11	.14	.01
m_9	95.2/95.7	5046	4500	61	48	101	14	1.02	.19	.07
m_{10}	95.2/95.8	5044	4582	62	49	101	14	1.02	.09	.04

Comments: Variation in Q_i hardly affects the results. ACP's are throughout good. With poorly designed RR technique, say, for $\underline{A_1}$, $\underline{B_1}$ and $\underline{A_2}$, $\underline{B_2}$ the methods perform badly compared to DR, but with better designs, say, with $\underline{A_4}$, $\underline{B_4}$ and $\underline{A_5}$, $\underline{B_5}$ one notices competitive results. Also, $m_j(j = 1, \ldots, 8)$ which are good for DR are also so for RR in contrast to m_9 and m_{10} which in both cases are poor.

CHAPTER SEVEN

ADJUSTMENTS FOR INCOMPLETE DATA BECAUSE OF PARTIAL NON-RESPONSE

7.0 SUMMARY.

The population U is supposed to consist of one part U_R of known individuals who will give required information on request and a complementary part U_C whose members will do so only with unknown and varying positive probabilities less than unity. Taking an initial sample from U_C first the probabilities of responses are estimated. A final sample is then taken from U to estimate the population total of a variable of interest. As in earlier chapters a super-population linear regression model is postulated connecting this variable of interest with another variable for which values are known. Generalized regression predictor is then amended to take account of this partial non-response. Variance estimators for it are then derived utilizing a model-cum-asymptotic-design-based approach as adopted throughout so far in this thesis. The main problem is then to construct appropriate confidence intervals with a right choice of a variance estimator for the point estimator of the total. To achieve this a simulation study is undertaken to compare the relative performances of the confidence intervals through numerical exercises.

7.1 INTRODUCTION.

Let U be dichotomized into U_R of size N_R and U_C of size N_C of known individuals. Let each unit of U_R be prepared to divulge facts demanded by the investigator but each member of U_C have an unknown probability q_i $(0<q_i<1,\ i\in U_C)$ of response. First a preliminary sample s_C of size n_C $(<N_C)$ is supposed to be drawn from U_C to estimate q_i's. A final sample s of size n is then drawn as in earlier chapters from U

with a probability p(s) admitting positive inclusion-probabilities π_i, π_{ij} for units in singles and in pairs. The model $\underset{\sim}{M}$ and its special cases are again postulated concerning the variable y of interest and the auxiliary variable x as in earlier chapters. The use of the generalized regression (greg) predictor for Y is again contemplated but adjustments on it are required because of partial non-response possibilities as indicated above.

To estimate q_i for $i \in U_C$ we consider two alternative methods, on choosing an SRSWOR s_C from U_C. In one, we make repeated attempts following the first failure and estimate q_i by q_{1i} which is the reciprocal of the number of attempts at which a response is procured. In the other, which is due to Politz and Simmons (1949, 1950), on the date of first success in response gathering, information is noted about number (say, j=0(1)6) of immediately preceding six days on which the respondent was available for response. Then q_i is estimated by $q_{2i} = (j+1)/7$, if i in s_C reported one of the numbers j above. We shall write q_i^* for either of q_{1i}, q_{2i}, $i \in s_C$. In order to estimate q_i for $i \in U_C$, we fit a logistic regression model.

Writing

$$\log_e \left(\frac{q_i^*}{1-q_i^*} \right) = a + b\, x_i + \eta_i,\ i \in U_C ,$$

with a, b as unknown constants and and η_i's as random errors, a, b are estimated by ordinary least squares methods as $\hat{a}$ and $\hat{b}$ using q_i^* for q_i, $i \in s_C$. Then, smoothed estimates of q_i for $i \in U_C$ are taken as

$$\hat{q}_i = \frac{\exp\ (\ \hat{a} + \hat{b}\ x_i)}{1 + \exp\ (\ \hat{a} + \hat{b}\ x_i)}\ ,\ i \in U_C\ .$$

For simplicity we shall write these $\hat{q}_i$'s also as true q_i's, $i \in U_C$ and take $\hat{q}_i = q_i$ equal to 1 for every i in U_R.

For analytical purpose we shall use the indicator

I_{ri} = 1 if i responds when i is selected in the final sample s,
= 0, else.

Denoting by E_q the operator of expectation with respect to this 'random' response-system, we have

$$E_q\ (I_{ri}) = q_i,\ i \in U .$$

Keeping these in mind, following Särndal and Hui (1981) we consider a modification of the greg predictor for Y as

$$t_g(R) = \sum' \frac{y_i}{\pi_i}\cdot\frac{I_{ri}}{q_i} + \hat{\beta}_q\left(X - \sum' \frac{x_i}{\pi_i}\cdot\frac{I_{ri}}{q_i} \right) ,$$

$$= \sum' \frac{y_i}{\pi_i}\cdot\frac{I_{ri}}{q_i}\, g_{si}(R),$$

where, $$\hat{\beta}_q = \frac{\Sigma' y_i x_i Q_i I_{ri}/q_i}{\Sigma' x_i^2 Q_i I_{ri}/q_i} ,$$

and, $$g_{si}(R) = 1 + \left(X - \sum' \frac{x_i}{\pi_i}\cdot\frac{I_{ri}}{q_i}\right) \frac{x_i Q_i \pi_i}{\Sigma' x_i^2 Q_i I_{ri}/q_i} .$$

Also, we shall write,

$$a_{sq} = \left(X - \sum' \frac{x_i}{\pi_i}\cdot\frac{I_{ri}}{q_i} \right) \Big/ \sum' x_i^2 Q_i \cdot\frac{I_{ri}}{q_i}$$

As in earlier chapters, here also we shall apply Brewer's (1979) asymptotic approach. In particular, in addition to the limiting expectation operator lim E_p we shall also use the limiting symbol lim E_q corresponding to E_q.

Let us write

$$\lim E_p \lim E_q E_m (t_g(R) - Y)^2 = M(R) \qquad (7.1.1)$$

and take it as a measure of error of $t_g(R)$ as an estimator for Y. For M(R) we seek an estimator m(R) which satisfies

$$\lim E_p \lim E_q E_m [\ m(R)\ ] = M(R) \qquad (7.1.2)$$

Next, as in earlier chapters we construct for Y confidence intervals using $t_g(R)$ and m(R).

Next, to compare their relative performances we again resort to simulation studies for numerical comparisons because analytic comparisons turn out dificult.

7.2 VARIANCE ESTIMATORS.

Using model $\underline{M}$ of earlier chapters we work out

$$E_m(t_g(R)-Y)^2 = a_{sq}^2 \ (\Sigma' x_i^2 \, Q_i^2 \, \frac{I_{ri}}{q_i} \, \sigma_i^2)$$

$$+ 2a_{sq} \ \Sigma' x_i^2 \, Q_i^2 \, \frac{I_{ri}}{q_i} \, (\frac{I_{ri}}{q_i} - 1) \, \sigma_i^2$$

$$+ \Sigma \, (\, \frac{I_{si} I_{ri}}{q_i} - 1 \,)^2 \, \sigma_i^2 .$$

Then, noting

$$\lim E_p \ (a_{sq}) = \left(X - \sum x_i \cdot \frac{I_{ri}}{q_i} \right) \Big/ \sum x_i^2 Q_i \pi_i \cdot \frac{I_{ri}}{q_i} = a_q, \text{ say,}$$

$$\lim E_p \ E_m \ (t_g(R)-Y)^2 = a_q^2 \sum x_i^2 Q_i^2 \, \frac{I_{ri}}{q_i} \, \sigma_i^2 \pi_i$$

$$+ 2a_q \sum x_i Q_i \pi_i \, \frac{I_{ri}}{q_i} \left(\frac{1}{\pi_i q_i} - 1 \right) \sigma_i^2$$

$$+ \sum \left(\frac{I_{ri}}{\pi_i q_i^2} - 2 \, \frac{I_{ri}}{q_i} + 1 \right) \sigma_i^2 ,$$

Then,

$$M(R) = \lim E_q \ \lim E_p \ E_m \ (t_g(R)-Y)^2 = \sum \left(\frac{1}{\pi_i q_i} - 1 \right) \sigma_i^2 ,$$

because,

$$\lim E_q \ (a_q) = E_q\left(X - \sum x_i \cdot \frac{I_{ri}}{q_i} \right) \Big/ E_q\left(\sum x_i^2 Q_i \pi_i \cdot \frac{I_{ri}}{q_i} \right) = 0.$$

To derive m(R) satisfying (1.2), let

$$t(\alpha) = \sum{}' \alpha_i I_{ri} \left(\frac{y_i}{x_i} - \sum{}' \frac{y_k}{x_k} \, \frac{I_{ri}}{q_i} \Big/ \sum{}' \frac{I_{ri}}{q_i} \right)^2 ,$$

with α_i's as assignable constants. Let,

$$t_1(\alpha) = E_m \ [\ t(\alpha) \ ]$$

$$= \sum{}' \alpha_i I_{ri} \frac{\sigma_i^2}{x_i^2} \left(1 - \frac{2}{q_i \sum{}' \frac{I_{ri}}{q_i}} \right)$$

$$+ \left(\sum{}' \alpha_i I_{ri} \right) \frac{\sum{}' \frac{\sigma_i^2}{x_i^2} \frac{I_{ri}}{q_i}}{\left(\sum{}' \frac{I_{ri}}{q_i} \right)^2} ,$$

$$t_2(\alpha) = \lim E_q [\, t_1(\alpha) \,]$$

$$= \sum{}' \left[\frac{\alpha_i q_i}{x_i^2} \left(1 - \frac{2}{nq_i} \right) + \frac{\Sigma' \alpha_i q_i}{n^2 x_i^2 q_i} \right] \sigma_i^2 ,$$

and,

$$t_3(\alpha) = \lim E_p [\, t_2(\alpha) \,]$$

$$= \sum \left[\frac{\alpha_i q_i}{x_i^2} \left(1 - \frac{2}{nq_i} \right) + \frac{\Sigma \, \alpha_i q_i \pi_i}{n^2 x_i^2 q_i} \right] \sigma_i^2 \pi_i .$$

Then, $m_1 = t(\alpha(1))$ is a choice of $m(R)$ subject to (1.2) if $t(\alpha(1))$ is $t(\alpha)$ with α_i equal to $\alpha_i(1)$, where the latter are the solutions for α_i's from the equations

$$\frac{\alpha_i q_i}{x_i^2} \left(1 - \frac{2}{nq_i} \right) + \frac{B}{n^2 x_i^2 q_i} = \frac{1}{\pi_i} \left(\frac{1}{\pi_i q_i} - 1 \right) , \quad i \in U,$$

writing $B = \Sigma \, \alpha_i q_i \pi_i$, which yields

$$\alpha_i(1) = \frac{1}{q_i(q_i - 2/n)} \left[\frac{x_i^2 q_i}{\pi_i} \left(\frac{1}{\pi_i q_i} - 1 \right) - \frac{B}{n^2} \right] ,$$

with B finally as

$$B = \sum \frac{x_i^2 q_i}{(q_i - 2/n)} \left(\frac{1}{\pi_i q_i} - 1 \right) \Big/ \left[1 + \frac{1}{n^2} \sum \frac{\pi_i}{(q_i - 2/n)} \right] .$$

Another choice of $m(R)$ subject to (1.2) is $m_2 = t(\alpha(2))$, where $t(\alpha(2))$ is $t(\alpha)$ with α_i equal to $\alpha_i(2)$ where $\alpha_i(2)$'s are solutions for α_i form the equations

$$\frac{\alpha_i q_i}{x_i^2}\left(1 - \frac{2}{nq_i}\right) + \frac{\Sigma' \alpha_i q_i}{n^2 x_i^2 q_i} = \frac{1}{\pi_i}\left(\frac{1}{\pi_i q_i} - 1\right), \quad i \in s.$$

Then, writing

$$A = \sum{}' \frac{x_i^2 q_i}{\pi_i (q_i - 2/n)}\left(\frac{1}{\pi_i q_i} - 1\right) \Big/ \left[1 + \frac{1}{n^2}\sum{}' \frac{1}{(q_i - 2/n)}\right],$$

we get $$\alpha_i(2) = \frac{1}{q_i(q_i - 2/n)}\left[\frac{x_i^2 q_i}{\pi_i}\left(\frac{1}{\pi_i q_i} - 1\right) - \frac{A}{n^2}\right].$$

For t_g based on the complete sample, variance estimators given by Särndal (1982) and further discussed by Särndal, Swensson and Wretman (SSW, 1992) are already available in the literature. Modifying them we get two variance estimators as follows for $t_g(R)$. Before finding them we consider a model-free estimator for Y based on incomplete sample as a modification of the well-known Horvitz - Thompson estimator HTE, namely,

$$\bar{t}\,(R) = \sum{}' \frac{y_i \, I_{ri}}{\pi_i \, q_i}.$$

Its variance is

$$E_p E_q(\bar{t}(R) - Y)^2 = \sum\sum \Delta_{ij}\pi_{ij}\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2 + \sum \frac{y_i^2}{\pi_i}\,\frac{1-q_i}{q_i}$$

$$= V, \text{ say}$$

where,

$$\Delta_{ij} = (\pi_i \pi_j - \pi_{ij}) / \pi_{ij}.$$

Two variance estimators for $\bar{t}(R)$ then readily emerge as

$$v_1(R) = \sum{}'\sum{}' \Delta_{ij}\left(\frac{y_i \, I_{ri}}{\pi_i \, q_i} - \frac{y_j \, I_{rj}}{\pi_j \, q_j}\right)^2 + \sum{}' \frac{y_i^2}{\pi_i}\,\frac{1-q_i}{q_i}\,\frac{I_{ri}}{q_i}$$

and, $$v_2(R) = \sum{}'\sum{}' \Delta_{ij}\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2 \frac{I_{ri} I_{rj}}{q_i \, q_j} + \sum{}' \frac{y_i^2}{\pi_i^2}\,\frac{1-q_i}{q_i}\,\frac{I_{ri}}{q_i}$$

uisng Yates and Grundy's (YG, 1953) variance estimator for original HTE, namely

$$\bar{t} = \sum\nolimits^{\checkmark} \frac{y_i}{\pi_i} .$$

Note that, $E_p\, E_q\, (v_1(R)) = V = E_p\, E_q\, (v_2(R))$.

Let us write

$$e_i = y_i - x_i\hat{\beta}, \quad e_{iq} = y_i - x_i\hat{\beta}_q , \quad E_i = y_i - x_i B_Q, \text{ and}$$

$$B_Q = \sum y_i x_i Q_i \pi_i \Big/ \sum x_i^2\, Q_i \pi_i .$$

Then, it follows with a little algebra that we suppress, from SSW (1992) that,

$$\lim E_p \lim E_q\, (t_g(R) - Y)^2$$

$$= \sum\sum \Delta_{ij}\pi_{ij}\left(\frac{E_i}{\pi_i} - \frac{E_j}{\pi_j}\right)^2 + \sum \frac{E_i^{\,2}}{\pi_i}\, \frac{1-q_i}{q_i} .$$

For t_g, Särndal's (1982) two variance estimators are

$$v_{s1} = \sum\nolimits^{\checkmark}\sum\nolimits^{\checkmark} \Delta_{ij}\left(\frac{e_i}{\pi_i} - \frac{e_j}{\pi_j} \right)^2$$

and,

$$v_{s2} = \sum\nolimits^{\checkmark}\sum\nolimits^{\checkmark} \Delta_{ij}\left(g_{si}\, \frac{e_i}{\pi_i} - g_{sj}\, \frac{e_j}{\pi_j} \right)^2 .$$

Noting v_{sj}, j=1,2, it follows that two reasonable variance estimators, following Yates and Grundy (YG, 1953), Särndal (1982) and SSW (1992), for $t_g(R)$ are

$$v_{sg}(1) = \sum\nolimits^{\checkmark}\sum\nolimits^{\checkmark} \Delta_{ij}\left(\frac{e_{iq}}{\pi_i}\, \frac{I_{ri}}{q_i} - \frac{e_{jq}}{\pi_j}\, \frac{I_{rj}}{q_j} \right)^2$$

$$+ \sum' \frac{e_{iq}^2}{\pi_i} \frac{1-q_i}{q_i} \frac{I_{ri}}{q_i}$$

and,
$$v_{sg}(2) = \sum\sum' \Delta_{ij}\left(g'_{si} \frac{e_{iq}}{\pi_i} \frac{I_{ri}}{q_i} - g'_{sj} \frac{e_{jq}}{\pi_j} \frac{I_{rj}}{q_j} \right)^2$$

$$+ \sum' (g'_{si})^2 \frac{e_{iq}^2}{\pi_i} \frac{1-q_i}{q_i} \frac{I_{ri}}{q_i} .$$

We consider CI's for Y based on $(\bar{t}(R), v_j(R))$, $j=1,2$ and $(t_g(R), v(R))$ with $v(R)$ as m_j and $v_{sg}(j)$, $j=1,2$. To assess their relative performances we consider simulation study.

7.3 SIMULATION STUDY.

We take N = 150, generate $\underline{X} = (x_1, \ldots, x_N)$ as a random sample from the exponential density

$$f_{\lambda, a_0}(x) = \frac{1}{\lambda} e^{-(x-a_0)/\lambda} , \quad x > a_0 = 7.0, \ \lambda = 8.5 ,$$

take ε_i's as random samples from N(0,1), take σ = 1.o, β = 2.0, h = 0.8, 1.4, θ = 0.0, 2.5 and generate four sets of $\underline{Y} = (y_1, \ldots, y_N)$ as

$$y_i = \theta + \beta x_i + \sigma x_i^{h/2} \varepsilon_i$$

where, the value θ = 0.0 represents the model (1.1) and the value θ = 2.5 is used to study the robustness of the various pairs (e,v) for a possible super-population intercept term in the model (1.1). Population I is generated with (θ=0.0, h=0.8), population II with (θ=0.0, h=1.4), population III with (θ=2.5, h=0.8) and population IV with (θ=2.5, h=1.4).

We first let U_R be the subset of U consisting of its last N_R = 50 units. The values of q_{1i}, q_{2i} for $i \in U_C$ and arbitrarily assigned to $i \in U_C$ are as follows :

$$q_{1i} = 1/2,\ 1/3,\ 1/4, \qquad \text{and,} \qquad q_{2i} = j/7,\ j=1(1)6.$$

We then take a simple random sample (SRS) s_C without replacement (WOR) of size $n_C = 20$. Writing q_i^* for q_{1i}, q_{2i} and using q_i^* for $i \in s_C$ we fit a logistic regression model

$$\log\left(\frac{q_i}{1-q_i}\right) = a + b\,x_i + \eta_i,\ i\in U_C.$$

Using this for $i\in s_C$, we fit it by least squares principle to obtain $\hat{a}$ and $\hat{b}$ — the estimates of a and b and derive q_i for $i\in U_C$ from

$$\frac{q_i}{1-q_i} = \exp\left[\hat{a} + \hat{b}\,x_i\right], \quad i\in U_C.$$

Of course we take $q_i = 1$ for i in U_R. Then we draw $\underline{W} = (w_1,\ldots,w_N)$ as a random sample from the density $f_{\lambda,a_0}(x)$ with $a_0 = 20.0$ and $\lambda = 15.0$, which we use as size-measures to draw a sample s of size n=32 from U. For this we apply the scheme given by Hartley and Rao (HR, 1962). The sampling is replicated R=1000 times. We write Σ_r as sum over these replicates and write

$$A = \frac{1}{R}\Sigma_r\, v, \quad \text{and,} \quad P = \frac{1}{R}\Sigma_r\,(e-Y)^2$$

for an estimator e for Y and an estimator v for MSE of e. For comparative study of the choices (e,v) we consider the following criteria :

1. ACP (Actual coverage percentage) : the percentage of samples for which CI's cover Y — the closer it is to $100(1-\alpha)$ the better the (e,v).

2. ACV (Average coefficient of variation) : the average, over the replicates of the values of $\sqrt{v}/e$ — this reflects the length of CI relative to e.

3. PCV (Pseudo coefficient of variation) :

$$PCV(v) = \frac{1}{A}\left[\frac{1}{R}\Sigma_r(v-A)^2\right]^{1/2}$$

4. AARE (Average absolute relative error) : $\frac{1}{R}\Sigma_r \left|\frac{e-Y}{Y}\right|$.

5. B(v) (Pseudo relative bias) : $\equiv (\frac{1}{R}\Sigma_r v - P)/P$.

6. S(v) (Pseudo relative stability) : $\equiv \frac{1}{P}\left[\frac{1}{R}\Sigma_r(v-P)^2\right]^{1/2}$

7. L(v) (Pseudo standardized length) : $\equiv \frac{1}{R}\Sigma_r \sqrt{v}/\sqrt{P}$.

8. B(d) (Bias of d) : $\equiv \frac{1}{R}\Sigma_r d$.

9. M(d) (Mean square error (MSE) of d) : $\equiv \frac{1}{R}\Sigma_r(d-B(d))^2$.

10. $\sqrt{\beta_1(d)}$ (Root beta one) : $\equiv \frac{1}{R}\sum_r\left(\frac{d-B(d)}{\sqrt{M(d)}}\right)^3$.

11. E(d) (Excess measure) : $\equiv \beta_2(d) - 3$ (beta two minus three)

$$= \frac{1}{R}\sum_r\left(\frac{d-B(d)}{\sqrt{M(d)}}\right)^4 - 3.$$

The smaller the numerical values of the measures 2 - 11, the better the pair (e, v).

The numerical findings based on simulations are summarized in the tables below in Appendix given at the end of this report and concluding remarks are summarized in the next section.

7.4 CONCLUDING REMARKS.

Unlike in the previous studies in DR (direct response) or RR (randomized response) set-up here our newly proposed variance estimators do not show impressive performances compared to Särndal's (1982). Moreover, almost with every variance estimator, with q_{1i}, ACP highly falls below the nominal confidence coefficient. Also, normality of the pivotal function used to construct the confidence intervals is often suspect. For q_{2i}, our newly proposed estimators yield highly promising results.

Appendix G

SUMMARY OF FINDINGS.

Values relating to only one choice of Q_i namely $Q_i = 1/(\pi_i x_i)$ (due to Hájek, 1971) are presented in the tables because other choices of $Q_i = (1-\pi_i)/(\pi_i x_i)$ (Brewer, 1979), $1/x_i^2$ and $1/x_i$ have been found on calculations to yield similar results.

The q_i^* stands for q_{1i} in tables G.1 and G.2, and for q_{2i} in tables G.3 and G.4, for $i \in U_C$.

The level of significance α is taken to be 5% and the ACP values are shown for τ only.

In the tables. we specify only the variance estimators used and not the predictors for Y, recalling that the variance estimators $v_1(R)$ and $v_2(R)$ are used if the predictor for Y is $\bar{t}(R)$ and the rest are used when the predictor is $t_g(R)$.

N.B. The acronyms used in Tables G.1-G.4 are as described on pp.109-110. AARE relates to error of e, ACP to coverage probabilities associated with confidence intervals (CI), ACV to coefficient of variation; PB, PS to bias and stability of v; PL to length of CI; B(.), M(.), $\sqrt{\beta_1(\cdot)}$ and E(.) to bias, MSE, skewness and excess of $d = (e - Y)/\sqrt{v}$. For q_{1i} and q_{2i} one may consult page 102.

Table G.1

Comparative statistics of different strategies in case of Population II ($\theta = 0.0$, $h = 1.4$) and $q_i^* = q_{1i}$

v	10^4 AARE	ACP	10^4 ACV	-10^3 PB	10^2 PS	10^2 PL	-10^2 B(d)	10^2 M(d)	10^2 $\sqrt{\beta_1(d)}$	E(d)
$v_1(R)$	2217	88.9	2586	5	93	93	-36	156	-93	1.31
$v_2(R)$	2217	87.1	2532	2	94	92	-44	192	-148	1.38
$v_{sg}(1)$	364	85.7	401	-136	67	87	-9	201	-8	1.75
$v_{sg}(2)$	364	89.5	411	-122	60	89	-6	151	-1	0.58
m_1	364	86.0	421	5	97	91	-12	204	-17	2.59
m_2	364	86.8	423	7	95	92	-12	199	-22	2.29

Comments: $v_j(R), j = 1, 2$ are too bad compared to others which seem adequate.

Table G.2

Comparative statistics of different strategies in case of Population IV ($\theta = 2.5$, $h = 1.4$) and $q_i^* = q_{1i}$

v	10^4 AARE	ACP	10^4 ACV	-10^3 PB	10^2 PS	10^2 PL	-10^2 B(d)	10^2 M(d)	10^2 $\sqrt{\beta_1(d)}$	E(d)
$v_1(R)$	2140	89.4	2523	2	86	94	-33	151	-90	1.27
$v_2(R)$	2140	87.2	2470	-2	87	93	-41	185	-148	3.88
$v_{sg}(1)$	330	86.5	362	-159	58	87	4	211	61	3.10
$v_{sg}(2)$	330	88.9	372	-132	55	89	-8	160	30	1.52
m_1	330	86.4	378	-78	66	91	-5	215	76	4.14
m_2	330	86.9	379	-70	65	91	-3	204	62	3.56

Comments: $v_j(R), j = 1,2$ are too bad compared to others which seem good enough.

Table G.3

Comparative statistics of different strategies in case of Population II ($\theta = 0.0$, $h = 1.4$) and $q_i^* = q_{2i}$

v	10^4 AARE	ACP	10^4 ACV	-10^3 PB	10^2 PS	10^2 PL	-10^2 B(d)	10^2 M(d)	10^2 $\sqrt{\beta_1(d)}$	E(d)
$v_1(R)$	4847	82.8	2310	-611	68	57	124	68	-89	2.65
$v_2(R)$	4847	90.4	2496	-539	64	64	111	54	-143	5.14
$v_{sg}(1)$	297	96.4	523	1137	169	140	-4	69	-30	1.04
$v_{sg}(2)$	297	91.8	364	16	56	97	0	121	-6	0.32
m_1	297	96.3	558	1524	236	150	-6	67	-39	1.26
m_2	297	96.7	560	1519	232	150	-5	65	-35	1.18

Comments: Again $v_j(R), j = 1,2$ turn out too poor compared to others which seem adequate.

Table G.4

Comparative statistics of different strategies in case of Population IV ($\theta = 2.5$, $h = 1.4$) and $q_i^* = q_{2i}$

v	10^4 AARE	ACP	10^4 ACV	-10^3 PB	10^2 PS	10^2 PL	-10^2 B(d)	10^2 M(d)	10^2 $\sqrt{\beta_1(d)}$	E(d)
$v_1(R)$	4748	81.3	2238	-627	68	58	128	69	-80	2.24
$v_2(R)$	4748	89.6	2421	-556	64	67	114	53	-134	4.48
$v_{sg}(1)$	267	97.5	470	1090	149	140	-9	66	26	1.36
$v_{sg}(2)$	267	92.7	328	19	50	98	-14	119	8	0.54
m_1	267	97.3	493	1315	177	147	-9	64	33	1.67
m_2	267	97.5	493	1310	174	147	-9	62	27	1.51

Comments: Again $v_j(R), j = 1,2$ turn out too bad compared to others which seem quite serviceable.

CHAPTER EIGHT

RAO-HARTLEY-COCHRAN STRATEGY -- CONFIDENCE INTERVALS BY TWO VARIANCE ESTIMATORS

8.0 SUMMARY.

In this final chapter we consider estimating the population total Y employing the Rao-Hartley-Cochran (RHC) (1962) strategy. RHC scheme of sampling consists in randomly splitting the population U into n groups ofsizes N_g ($g = 1,\ldots,n$, $\sum_{g=1}^{n} N_g = N$). For every unit there is a known positive normed size-measure, say, p_j ($0<p_j<1$, $\Sigma p_j=1$) , $j \in U$. From within each group thus formed, one unit is chosen with a probability proportional to its size-measure. This is repeated independently across each group. We shall write Q_g for the sum of the p_j-values say p_{gj} ($j=1,\ldots,N_g$, $g=1,\ldots,n$) of the units falling in the g-th group and Σ_n for the sum over the units chosen in the sample of size n so drawn. Writing y_{gj} as the y-value y_j for the unit falling in g-th group, The RHC estimator for Y is

$$t_R = \Sigma_n\, y_{gj}\, Q_g \,/\, p_{gj}.$$

For this, two variance estimators are available – one given by RHC themselves and the other by Ohlsson (1989). In his 1989 paper Ohlsson demonstrated by numerical calculations that his variance estimator has a smaller variance than the other in many situations and hence is to be preferred. Chaudhuri and Mitra (1991) however demonstrated contrary results in more realistic situations. A major portion of their work is reproduced below. Besides, further comparison is made here between these two variance estimators on examining their relative efficacies in producing confidence intervals for the population total. Analytic comparison being difficult, in this chapter also simulation studies are resorted to in attempting only numerical evaluations. For this purpose both live data from published results are utilized and observations are generated postulating linear regression models

[illegible] is used in implementing the RHC sampling scheme. These numerical findings also suggest an over-all balance of advantage in favour of RHC's variance estimator over Ohlsson's

8.1 INTRODUCTION.

We consider the problem of estimating the total Y of the values y_i for the units i of a population $U=(1,..,i,..,N)$ when the normed size measures p_i $(0<p_i<1)$ are available. The values y_i though unknown are supposed to be well-associated with the known p_i's. So, it is usual to utilize p_i's both in choosing a sample and in employing an efficient estimator. The simplest way to do so following Hansen and Hurwitz (1943) is to make n (<N) independent draws out of U assigning selection-probability p_i to i on each draw and estimate Y by the average of y_i/p_i values over the units drawn. Though this procedure is simple, selection with replacement is a shortcoming. This defect is overcome by the more efficient alternative procedure given by Rao, Hartley and Cochran (1962) briefly described below. First U is divided into n disjoint random groups of sizes N_g $(\Sigma_n N_g=N)$, $g=1,...n$. Let p_{gj} $(j=1,..,N_g,\ g=1,..,n)$ be the p_i values for the units falling in the g-th group and $Q_g = \sum_{j=1}^{N_g} p_{gj}$. A unit is chosen from the g-th group with a probability p_{gj}/Q_g and this is repeated independently over the n groups. Denoting by y_{gj} the value of y for the unit, say j, so chosen from the g-th group but suppressing the subscript j for simplicity from p_{gj} and y_{gj} , the unbiased estimator for Y based on this sampling scheme, as proposed by RHC, is

$$t_R = \Sigma_n (y_g/p_g) Q_g.$$

Its variance is

$$V(t_R) = V\ (\ \Sigma_n N_g^2 - N\ )\ /\ [\ N(N-1)\ ]$$

where,

$$V = \sum_{i=1}^{N} p_i (\ y_i/p_i - Y\ )^2$$

For simplicity we shall write $T = \Sigma_n N_g(N_g - 1)$. RHC gave the non-negative unbiased estimator for $V(t_R)$ as

$$v_1 = \frac{T}{N(N-1) - T} \frac{1}{2} \sum_{g \neq m} \sum \left(\frac{y_g}{p_g} - \frac{y_m}{p_m} \right)^2 Q_g Q_m$$

and Ohlsson (1989) recommended in preference to it the alternative non-negative unbiased estimator for $V(t_R)$ as

$$v_0 = \frac{T}{2n(n-1)} \sum_{g \neq m} \sum \left(\frac{y_g}{p_g} - \frac{y_m}{p_m} \right)^2 \frac{Q_g Q_m}{N_g N_m}$$

Both v_1 and v_0 are members of the general class of variance estimators given by Ohlsson namely

$$v_b = \frac{1}{2} \sum_{g \neq m} \sum \frac{b_{gm}}{N_g N_m} \left(\frac{y_g}{p_g} - \frac{y_m}{p_m} \right)^2 Q_g Q_m$$
$$= \sum_{g \neq m} \sum a_{gm} d_{gm}$$

where b_{gm}'s are non-negative constants such that $b_{gm} = b_{mg}$ for all $g \neq m$ and,

$$\sum_{g \neq m} \sum b_{gm} = T,$$

$$a_{gm} = b_{gm} / N_g N_m \ , \quad g \neq m$$

$$d_{gm} = \frac{1}{2} \left(\frac{y_g}{p_g} - \frac{y_m}{p_m} \right)^2 Q_g Q_m$$

If $\frac{N}{n}$ is an integer, then $V(t_R)$ is the minimum for the choice $N_g = \frac{N}{n}$, $g=1,\ldots,n$, and Ohlsson noted that in this case $v_1 = v_0$.

If $\frac{N}{n}$ is not an integer then we shall write $\Theta = [\frac{N}{n}]$, the largest positive integer less than $\frac{N}{n}$. In this case, to be called the "case A", $V(t_R)$ is the least for the choice

$$N_g = \Theta \quad \text{for } k \ (1<k<n) \text{ of the g's, and}$$
$$= \Theta+1 \quad \text{for the remaining } (n-k) \text{ of the g's,}$$

with k so chosen that $\Sigma_n N_g = N = k\Theta + (n-k)(\Theta+1)$.

In the above two cases RHC themselves showed that $V(t_R)$ is less than the variance equal to V/n of Hansen and Hurwitz (1943) estimator.

In "case A", T reduces to $(N-k)\theta$. Other possible but uninteresting choices of N_g are ruled out from our discussions to follow. In fact from now on we shall consider only the "case A" which is really of practical interest in choosing between v_0 and v_1.

Writing

$$B_1 = \sum_{1 \le g \ne m \le k}\sum d_{gl}\ , \quad B_2 = \sum_{k+1 \le g \ne m \le n}\sum d_{gl}\ , \quad B_3 = \sum_{1 \le g \ne m \le n}\sum d_{gl} - B_1 - B_2,$$

$$d_1 = \frac{T}{n(n-1)\theta^2}\ , \quad d_2 = \frac{T}{n(n+1)(\theta+1)^2}, \quad d_3 = \frac{T}{n(n+1)\theta(\theta+1)}, \quad c = \frac{T}{N(N+1)-T},$$

we have $v_0 = d_1B_1 + d_2B_2 + d_3B_3$ and $v_1 = c(B_1+B_2+B_3)$, with T as above. Observing that

$$\frac{d_1}{c} = \left(1 - \frac{n-k}{N}\right)^{-2} + O\left(\frac{1}{nN}\right) = 1 + O\left(\frac{n-k}{N}\right)$$

$$\frac{d_2}{c} = \left(1 + \frac{k}{N}\right)^{-2} + O\left(\frac{1}{nN}\right) = 1 + O\left(\frac{k}{N}\right)$$

$$\frac{d_3}{c} = \left[\left(1 - \frac{n-k}{N}\right)\left(1 + \frac{k}{N}\right)\right]^{-2} + O\left(\frac{1}{nN}\right) = 1 + O\left(\frac{n-2k}{N}\right)$$

we may expect v_0 $(V(v_0))$ not to differ much from v_1 $(V(v_1))$ for a given $\underline{Y} = (y_1, \ldots, y_i, \ldots, y_N)'$.

Recalling that the formula for $V(v_b)$ given by Ohlsson is quite complicated we find it difficult to compare the magnitudes of $V(v_1)$ and $V(v_0)$. But defining

$$G = 100\frac{V(v_1) - V(v_0)}{V(v_1)}$$

as the percent gain in efficiency of v_0 over v_1, we intuitively feel that

(i) the magnitude of G should be "quite small" in most of the cases of interest and that

(ii) the sign of G should be 'both positive and negative' over variations in $\underline{Y}$ for alternative choices of N and n.

So we carried out a numerical exercise in order to confirm or invalidate these hunches and some of our findings are reported in

Section 8.2 below.

8.2 SIMULATION STUDIES I.

Restricting to the "case A" we present in this section some numerical values of G.

(a) First we treat the natural population illustrated by Horvitz and Thompson (1952, p-682) also referred to by Ohlsson (1989). Here N=20 and for this population we have the following values :

N.B. In the tables I - III(B), G-values denote values of $G = 100\frac{V(v_1)-V(v_0)}{V(v_1)}$.

TABLE I

G-values for several n but N fixed at 20

n:	6	7	8	9	11	12	13
G:	0.1157	-0.0039	0.0240	0.0354	-0.4945	-0.6931	-0.7261

Comment: Since G can be both positive and negative neither v_1 nor v_0 is uniformly superior to the other.

(b) Next we consider another natural population occurring in Cochran (1977, p-152) also covered by Ohlsson (1989). Here N=49 and for this population we have the following values :

TABLE II

G-values for several n but N fixed at 49

n :	4	5	6	8	9	10	11
G :	0.0657	0.1241	0.1671	0.2874	0.9274	0.4846	1.5898
n :	12	13	14	15	16	17	18
G :	0.5850	1.9070	3.0920	2.7941	0.9722	2.5123	5.3826

Comment: Though G is throughout positive suggesting superiority of v_0 over v_1 its magnitude is quite small.

(c) It is well known, for example, from Chaudhuri and Arnab (1979) among many other sources, that if y_i-values are amenable to the following model then the RHC strategy is often appropriate.

The model :

$$y_i = \beta x_i + \varepsilon_i$$

with x_i (>0) as known size-measures, $X = \sum_{i=1}^{N} x_i$, $p_i = x_i/X$ and ε_i's are uncorrelated random variables with a common mean zero and variances $\sigma_i^2=\sigma^2 x_i^h$ (conditionally on x_i) with σ (>0) and h ($0 \le h \le 1$) as unknown constants.

As we note that for $\underline{Y}$ generated subject to this model, v_1 and v_0 are independent of β, we take $\beta=0$ and since the value of G, under this model, is free of σ, we take $\sigma=1$. In order to generate a $\underline{Y}$ subject to this model, for simplicity, we further assume that

(i) ε_i's are distributed as the variables $u_i x_i^{h/2}$, where u_i's are independently and identically distributed (i.i.d.) as

$$(\chi_1^2 - 1) / \sqrt{2}$$

where χ_1^2 is a chi-square variable with 1 degree of freedom and

(ii) x_i's are independently identically distributed (i.i.d.) with a common probability density

$$f(x) = \frac{1}{8.5} e^{-x/8.5}, \quad x > 0.$$

For the purpose of numerical illustration we take N=18 and n=4 and 5. With these stipulations, for N=18 we first generate $\underline{X} = (x_1,\ldots,x_i,\ldots,x_N)'$, then $\underline{u} = (u_1,\ldots,u_i,\ldots,u_N)'$, then separately for the choice of h as 0.4, 0.5, 0.6, 0.8, 0.9 and 1.0, generate $\underline{\varepsilon} = (\varepsilon_1,\ldots,\varepsilon_i,\ldots,\varepsilon_N)'$ and finally $\underline{Y} = (y_1,\ldots,y_i,\ldots,y_N)'$. Then considering samples of sizes n=4 and 5, applying RHC scheme in each case we obtain the values of G, calculating $V(v_0)$ and $V(v_1)$ using Ohlsson's (1989) formulae. Some of the values of G thus found are illustrated below in the tables III(A) and III(B).

From these we may conclude that in the realistic "case A", the new variance estimator v_0 may not appreciably beat the classical variance estimator v_1 and may even sometimes fare worse. So before opting for v_0 in preference to v_1 as is apparently recommended by Ohlsson (1989) further care seems necessary in view of what we numerically illustrate above. Also, it is not evident from Ohlsson's (1989) paper why one should have $V(v_1)$ greater than $V(v_0)$ in general excluding the case when v_1 equals v_0 if N_g=N/n for every g = 1,..n. Since in "case A" there is a

TABLE III(A)

G-values for several values of g when N=18 and n=4

g					
0.4	0.5	0.6	0.8	0.9	1.0
0.5569	0.4652	0.3594	0.1361	0.0378	-0.0426
0.9834	0.9778	0.9661	0.8820	0.7616	0.5622
0.5251	0.5190	0.5153	0.5126	0.5127	0.5134
0.9751	0.9710	0.9642	0.9350	0.9070	0.8675
0.9873	0.9870	0.9865	0.9838	0.9792	0.9663
0.7134	0.6864	0.6625	0.6278	0.6165	0.6083
0.8280	0.7329	0.5728	0.0923	-0.1166	-0.2515
0.5310	0.4499	0.3564	0.1480	0.0473	-0.0416
0.4340	0.2404	0.0567	-0.1828	-0.2418	-0.2767
0.9380	0.9202	0.8907	0.7655	0.6563	0.5250

Comments: The magnitude of G is quite small and it can be both positive and negative and hence the question of superiority of v_0 over v_1 or vice versa is inconclusive.

TABLE III(B)

G-values for several values of g when N=18 and n=5

g					
0.4	0.5	0.6	0.8	0.9	1.0
1.0234	0.8187	0.5940	0.1530	-0.0289	-0.1730
2.1163	2.1004	2.0675	1.8428	1.5406	1.0745
0.9545	0.9455	0.9418	0.9446	0.9489	0.9541
2.0967	2.0838	2.0630	1.9750	1.8927	1.7801
2.1327	2.1319	2.1307	2.1231	2.1101	2.0734
1.3702	1.3057	1.2508	1.1751	1.1524	1.1368
1.7039	1.4659	1.0909	0.0964	-0.2904	-0.5288
0.9560	0.7768	0.5774	0.1621	-0.0258	-0.1858
0.7869	0.3815	0.0270	-0.4065	-0.5090	-0.5691
1.9911	1.9386	1.8535	1.5132	1.2386	0.9309

Comments: Since both positive and negative but numerically small values emerge for G, neither of v_1 and v_0 can beat the other.

minimal variation among N_g's we were led to conjecture that v_0 should not deviate in this case substantially from v_1 so as to turn out more efficient than the latter for every realistic <u>Y</u>. Our conjecture seems sensible in the light of our numerical findings briefly reported above.

8.3 FURTHER STUDIES.

To pursue with the investigation of relative merits of v_0 and v_1

let us next consider the properties of confidence intervals for Y based on t_R respectively associated with v_0 and v_1. For this we follow the works relating to simulation-based comparative investigations of performances of several estimators of mean square errors (MSE) or variances of different estimators for Y as are reported by our predecessors, namely Royall and Cumberland (1981, 1985), Rao and Wu (1983) and Deng and Wu (1987) among others.

As is customary with inference making for finite populations, employing suitable estimators v for $V(t_R)$, we may regard as in the earlier chapters of this thesis, the standardized error (SZE, say)

$$e = \frac{t_R - Y}{\sqrt{v}}$$

as a variable distributed at least for moderately large n, over hypothetically repeated sampling by RHC method, as a Student's t-statistic with (n-1) degrees of freedom (df, in brief), which for still larger n may also be treated as a standardized normal deviate τ. Then if $t_{\alpha/2}$ and $\tau_{\alpha/2}$ are such that $\alpha = \Pr.[|t|>t_{\alpha/2}]$ and $\alpha = \Pr.[|\tau|>\tau_{\alpha/2}]$, for a pre-assigned α in the open interval (0,1), a $100(1-\alpha)\%$ confidence interval for Y based on t_R and v is provided by $(t_R \pm t_{\alpha/2}\sqrt{v})$ or by $(t_R \pm \tau_{\alpha/2}\sqrt{v})$ to be denoted respectively as t-interval and τ-interval. In order to investigate how well this confidence interval performs with v chosen as either v_0 or v_1, we proceed with a simulation-based study in a manner narrated in **Section 8.4** below.

8.4 SIMULATION STUDY II.

For reasons noted by Chaudhuri and Mitra (1991) we postulate a model under which we may write

$$y_i = \beta x_i + \varepsilon_i$$

where β is an unknown parameter and ε_i's are independent random variables ditributed, conditionally given $\underline{X} = (x_1,\ldots,x_i,\ldots,x_N)'$, with a common zero mean and variances $\sigma_i^2 = \sigma^2 X_i^{\,h}$, with σ (>0) and h ($0 \le h \le 1$) as unknown constants. For a $\underline{Y} = (y_1,\ldots,y_i,\ldots,y_N)'$ so modelled, t_R is well-known to be an appropriate estimator based on RHC scheme. So we consider it appropriate to generate several $\underline{X}$, $\underline{Y}$ vectors as modelled above, draw several samples s by RHC scheme of various sizes n, calculate $t_R(s)$, $v_0(s)$, $v_1(s)$, e(s) based on s and examine the performances of v_0 and v_1 from the undernoted considerations.

Choosing α = .01, .05 and .10, we consider the ACP (Actual Coverage Probability) values associated with v_0 and v_1 to see how close they are to the nominal confidence coefficients $100(1-\alpha)\%$. If they are closer for v_0 than for v_1 then v_0 is to be preferred to v_1 and vice versa. In order to sharpen our preference criterion conerning v_0 and v_1 we identify two ancillary statistics namely (1) $A_1 = \sum_n Q_g^2$ and (2) $A_2 = \sum_n Q_g/p_g$ whose variation may affect the variation in t_R, v_0 and v_1. So we consider 'conditional confidence intervals' fixing the magnitudes of (1) and (2). To do this we divide the realized samples into a few equal sized groups such that the samples with the lowest values of (1) go to the first group, the samples with a next higher set of values of (1) go to the second group and so on. We do likewise with values of (2). ACP values are then calculated group-wise.

It is easy to check that the expectations of A_1 and A_2 with respect to RHC sampling scheme, respectively are

$$E(A_1) = E(\sum_n Q_g^2) = \frac{T}{N(N-1)} + \frac{N - \sum_n N_g^2}{N(N-1)} \sum_{i=1}^{N} p_i^2 = C \text{ (say)}$$

and,
$$E(A_2) = E(\sum_n Q_g/p_g) = N$$

So if A_1 (A_2) differs appriciably from C (N) for realized samples, then following Deng and Wu (1987) one may consider employing alternative variance estimators, namely

$$v' = \left(\frac{A_1}{C}\right)^d v \quad \text{and,} \quad v'' = \left(\frac{A_2}{N}\right)^d v,$$

choosing appropriate values for d. These variance estimators will of course be design-biased and criterion for discriminating among them should be formulated in terms of their MSE's. In our numerical illustrations with conditional performances of v_0 and v_1 with sample variation of A_1, A_2 we do not notice much effects of the ancillaries. So we did not deal further with v' and v''.

In order to carry out our simulation we first draw a random sample of x_i-values, $i=1,\dots,N$, from the exponential distribution with a probability density function (pdf)

$$f(x) = \frac{1}{\lambda} e^{-x/\lambda}, \quad x > 0.$$

Then we draw random samples U_i, $i=1,\dots,N$ where,

(1) $U_i = \sqrt{12}\ (u(0,1)-0.5)$ where $u(0,1)$ is distributed uniformly over the interval $(0,1)$,

(2) U_i is distributed as $N(0,1)$, and ,

(3) $U_i = (\chi_1^2 - 1)/\sqrt{2}$ where χ_1^2 is a chi-sqare variable with one degree of freedom, and,

then we take $\varepsilon_i = U_i X_i^{h/2}$, choosing several values of h in $[0,1]$.

Noticing that v_0 and v_1 are free of β and e is free of σ, we take $\beta=0$ and $\sigma=1$. Also we take N=50 and n=11, for 4 populations and N=150 and n=32 for a 5th population described below, and draw 1000 samples. Consistently with standard conventions we consider the following measures of criteria for performance characteristics of v_0 and v_1 :

$V = \frac{1}{1000}\sum_s (t_R(s) - Y)^2$, the pseudo variance of t_R for the simulated samples,

$RB = \frac{1}{1000}\sum_s \frac{v(s)}{V} - 1$, a measure of relative bias of v as an estimator for $V(t_R)$,

$RS = \left[\frac{1}{1000}\sum\left(\frac{v(s)}{V} - 1\right)^2\right]^{1/2}$, a measure of relative stability of v as an estimator of $V(t_R)$,

$SL = \frac{1}{1000}\sum_s \sqrt{v(s)}/\sqrt{V}$, the standardized length of the confidence interval,

$PCV = \frac{1}{A}\left[\frac{1}{1000}\sum_s (v(s) - A)^2\right]^{1/2}$, pseudo coefficient of variation , where $A = \sum_s v(s) / 1000$.

Further we calculate the (a) mean, (b) variance, (c) measure of skewness, namely $\sqrt{\beta_1}$ coefficient and (d) measure of kurtosis, namely β_2-3 coefficient for the statistic e based on the 1000 samples to check the departure in the nature of the distribution of e from the two postulated ones. For the treatment of conditional confidence intervals, we form 10 groups of 100 samples each and calculate the following statistics; writing $\sum_{h_1}$ as the sum over the 100 samples s_{h_1} for the h_1-th group, $h_1=1,\ldots,10$, namely,

(i) $A_{jh_1} = \frac{1}{100}\sum_{h_1} A_j(s_{h_1})$, $j=1,2$

(ii) $V_{h_1} = \frac{1}{100}\sum_{h_1} (t_R(s_{h_1})-Y)^2$

(iii) $v_{h_1} = \frac{1}{100}\sum_{h_1} v(s_{h_1})$

As an over-all measure of efficiency of v we also calculate the statistic

$$d = \left[\frac{1}{10} \sum_{h_1=1}^{10} \left(\sqrt{v_{h_1}} - \sqrt{V_{h_1}} \right)^2 \right]^{1/2}$$

and use its magnitude as a criterion of comparison in the efficiencies of v_0 and v_1, the smaller the magnitude of D the better for v. The detailed findings are reported through the tables presented below. To put it in a nutshell, the important message conveyed by them is this that (i) there is little discernible qualitative differences in the merits of v_0 and v_1 from the point of view of their capabilities of yielding confidence statements, the ACP's corresponding to both being close to pronounced nominal confidence coefficients, but that (ii) in this respect the balance tilts in favour of v_1 even though it is quite slight though (iii) in terms of the criterion G as demonstrated by Chaudhuri and Mitra (1991), the preference might be attached to v_0 because the positive values of G far out-numbered the negative ones in their numerical illustrations and this is the main reason why this work reported here was undertaken.

We present numerical findings relating to unconditional performances for the five populations labelled $i = 1\ (1)\ 5$, described below :

(1) U_i distributed as $N(0,1)$, $\lambda = 2.5$, $g = 0.4$,
(2) U_i distributed as $N(0,1)$, $\lambda = 2.5$, $g = 0.9$,
(3) U_i distributed as $(\chi_1^2 - 1)/\sqrt{2}$, $\lambda = 13.59$, $g = 0.5$,
(4) U_i distributed as $(\chi_1^2 - 1)/\sqrt{2}$, $\lambda = 13.59$, $g = 0.6$,
(5) U_i distributed as $N(0,1)$, λ=8.5, but 10 is added to samples drawn from

$f_\lambda(x)$ to get x_i's, g=0.4.

In the tables, we show the ACP-values and in all the tables, the values relating to v_0 are shown underlined.

Conditional performances are numericlly shown for population 5 only.

Finally, we present below, in Section 8.5, the tables relevant to this section of the text.

8.5 NUMERICAL FINDINGS.

Acronyms used in Tables 8.1-8.3 are as given on pp.122-123. *V* denotes pseudo variance of t_R, $V(t_R)$ is true variance of t_R; RB, RS denote bias and stability of variance estimators v_0, v_1 and $\sqrt{\beta_1}$, $\beta_2 - 3$ give skewness and excess measures of standardized statistic or pivot. A_{1h} and A_{2h} are ancillaries, vide p-122 for *h*-th group, $h = 1, \ldots, 10$.

Table 8.1

Detailed performances of v_1 and v_0.

Pop. Id	$V(t_R)$	V	RB	RS	SL	Mean	Var	$\sqrt{\beta_1}$	β_2-3
1	71.11	68.82	0.00	0.71	0.94	-0.54	1.99	-1.54	4.80
			0.00	0.70	0.94	-0.53	1.98	-1.54	4.80
2	48.98	48.58	-0.02	0.59	0.95	-0.52	1.75	1.31	3.81
			-0.02	0.59	0.95	-0.51	1.74	1.30	3.83
3	357.82	386.92	-0.07	1.94	0.75	-0.76	2.33	-1.15	2.26
			-0.07	1.94	0.76	-0.76	2.30	-1.14	2.25
4	360.95	390.99	-0.07	1.81	0.77	-0.72	2.21	-1.11	2.12
			-0.07	1.81	0.77	-0.72	2.19	-1.10	2.11
5	7285.19	7339.11	-0.05	2.16	0.83	0.55	0.96	0.19	-0.28
			-0.06	2.10	0.83	0.54	0.96	0.18	-0.27

Comment: As *V* is close to $V(t_R)$ the simulation seems adequate. Both v_0 and v_1 seem keenly competitive.

Table 8.2

Coverage Probabilities of the τ- and t-intervals using v_1 and v_0

Pop. Id.	τ-interval 99%	95%	90%	t-interval 99%	95%	90%	PCV
1	92.8	86.1	80.7	96.0	90.0	83.8	0.7115
	92.8	85.9	81.0	95.9	90.0	84.4	0.7039
2	93.9	87.8	81.9	96.5	90.8	85.2	0.6070
	94.0	87.8	82.2	96.4	90.9	84.9	0.6004
3	87.9	80.9	75.3	92.2	84.5	78.6	2.0871
	87.8	80.9	75.0	92.4	85.1	78.6	2.0790
4	88.8	81.6	76.7	92.6	85.5	80.0	1.9503
	88.7	81.8	76.7	93.0	86.0	79.8	1.9416
5	99.2	96.3	91.7	99.6	96.7	92.8	2.2671
	99.3	96.2	91.8	99.6	96.8	93.0	2.2022

Comments: As $(\beta_2 - 3)$ from Table 8.1 is far from zero except for population 5, the ACP's fall short of the nominal confidence coefficient. But for population 5 with negligible $(\beta_2 - 3)$, the ACP's are adequate. But v_0 and v_1 are closely competitive.

Table 8.3

Conditional performances of v_1 and v_0 for the population 5, using ancillary A_1

Group (h)	A_{1h}	RB	RS	SL	Mean	Var	$\sqrt{\beta_1}$	β_2-3
1	0.03584	-0.07	1.84	0.80	-0.03	1.00	0.32	-0.13
		-0.22	2.06	0.81	-0.03	0.99	0.33	-0.11
2	0.03675	0.06	1.24	0.92	0.08	0.91	-0.19	-0.64
		0.06	1.23	0.92	0.08	0.91	-0.17	-0.64
3	0.03735	0.10	1.33	0.93	-0.02	0.75	0.21	0.02
		0.08	1.27	0.93	-0.02	0.76	0.19	0.07
4	0.03787	-0.07	0.77	0.90	0.01	0.99	0.37	-0.28
		-0.08	0.74	0.89	0.00	1.01	0.37	-0.24
5	0.03829	0.13	2.59	0.86	0.06	0.82	0.17	-0.21
		0.11	2.46	0.86	0.07	0.83	0.16	-0.23
6	0.03870	-0.29	2.70	0.66	-0.16	0.97	0.21	-0.66
		-0.29	2.60	0.66	-0.15	0.99	0.22	-0.62
7	0.03912	-0.20	1.09	0.78	0.02	1.04	0.29	-0.60
		-0.20	1.08	0.78	0.02	1.04	0.26	-0.58
8	0.03971	0.04	2.72	0.78	0.28	0.88	-0.15	-0.22
		0.02	2.55	0.79	0.27	0.87	-0.19	-0.23

Table 8.3 (continued)

Group (h)	A_{1h}	RB	RS	SL	Mean	Var	$\sqrt{\beta_1}$	β_2-3
9	0.04041	0.11	0.99	0.97	0.23	0.95	0.12	-0.40
		<u>0.10</u>	<u>0.97</u>	<u>0.97</u>	<u>0.24</u>	<u>0.95</u>	<u>0.12</u>	<u>-0.38</u>
10	0.04168	-0.03	1.45	0.86	0.06	1.10	0.47	0.38
		<u>-0.03</u>	<u>1.51</u>	<u>0.86</u>	<u>0.06</u>	<u>1.09</u>	<u>0.46</u>	<u>0.38</u>

Comments: With changing values of the ancillary A_1, the criteria measures change showing no discernible pattern. But within each group v_0 and v_1 perform almost similarly.

Table 8.4

Conditional coverage probabilities of the τ- and t-intervals using v_1 and v_0 and PCV for population 5, using ancillary A_1.

Group (h)	τ-interval 99%	95%	90%	t-interval 99%	95%	90%	PCV
1	99.0	98.0	80.0	99.0	99.0	90.0	1.9695
	<u>99.0</u>	<u>95.0</u>	<u>93.0</u>	<u>99.0</u>	<u>97.0</u>	<u>93.0</u>	<u>2.1109</u>
2	100.0	97.0	94.0	100.0	97.0	96.0	1.1731
	<u>100.0</u>	<u>97.0</u>	<u>94.0</u>	<u>100.0</u>	<u>97.0</u>	<u>96.0</u>	<u>1.1644</u>
3	99.0	98.0	97.0	100.0	98.0	98.0	1.2105
	<u>99.0</u>	<u>98.0</u>	<u>97.0</u>	<u>100.0</u>	<u>98.0</u>	<u>98.0</u>	<u>1.1651</u>
4	99.0	96.0	89.0	100.0	96.0	91.0	0.8243
	<u>99.0</u>	<u>96.0</u>	<u>89.0</u>	<u>100.0</u>	<u>96.0</u>	<u>90.0</u>	<u>0.8047</u>
5	99.0	98.0	95.0	99.0	98.0	95.0	2.2967
	<u>99.0</u>	<u>98.0</u>	<u>94.0</u>	<u>99.0</u>	<u>98.0</u>	<u>95.0</u>	<u>2.2122</u>
6	100.0	96.0	92.0	100.0	97.0	93.0	3.7645
	<u>100.0</u>	<u>96.0</u>	<u>92.0</u>	<u>100.0</u>	<u>96.0</u>	<u>93.0</u>	<u>3.6597</u>
7	100.0	96.0	86.0	100.0	96.0	87.0	1.3294
	<u>100.0</u>	<u>95.0</u>	<u>85.0</u>	<u>100.0</u>	<u>97.0</u>	<u>88.0</u>	<u>1.3162</u>
8	99.0	97.0	93.0	100.0	97.0	93.0	2.6275
	<u>100.0</u>	<u>96.0</u>	<u>93.0</u>	<u>100.0</u>	<u>97.0</u>	<u>94.0</u>	<u>2.5090</u>
9	99.0	96.0	90.0	99.0	97.0	93.0	0.8892
	<u>99.0</u>	<u>96.0</u>	<u>92.0</u>	<u>99.0</u>	<u>97.0</u>	<u>93.0</u>	<u>0.8729</u>
10	98.0	95.0	88.0	99.0	95.0	89.0	1.4819
	<u>98.0</u>	<u>95.0</u>	<u>89.0</u>	<u>99.0</u>	<u>95.0</u>	<u>90.0</u>	<u>1.5189</u>

Comments: For every group formed in terms of the ancillary A_1, both v_0 and v_1 perform closely.

Table 8.5

Conditional performances of v_1 and v_0 for the population 5, using ancillary A_2.

Group (h)	A_{2h}	RB	RS	SL	Mean	Var	$\sqrt{\beta_1}$	β_2-3
1	0.59393	0.33	0.94	1.11	0.23	0.89	-0.18	-0.05
		<u>0.33</u>	<u>0.95</u>	<u>1.11</u>	<u>0.23</u>	<u>0.89</u>	<u>-0.19</u>	<u>-0.03</u>
2	0.67812	0.29	1.13	1.06	0.22	0.80	0.22	-0.47
		<u>0.31</u>	<u>1.17</u>	<u>1.07</u>	<u>0.22</u>	<u>0.80</u>	<u>0.25</u>	<u>-0.45</u>
3	0.73719	0.18	1.05	1.01	0.10	0.77	0.33	-0.33
		<u>0.21</u>	<u>1.07</u>	<u>1.02</u>	<u>0.10</u>	<u>0.76</u>	<u>0.33</u>	<u>-0.33</u>
4	0.80221	0.51	1.22	1.17	0.14	0.57	0.06	-0.58
		<u>0.50</u>	<u>1.18</u>	<u>1.17</u>	<u>0.14</u>	<u>0.57</u>	<u>0.07</u>	<u>-0.61</u>
5	0.86822	0.09	0.94	0.97	0.03	1.03	0.28	-0.29
		<u>0.11</u>	<u>0.98</u>	<u>0.98</u>	<u>0.02</u>	<u>1.02</u>	<u>0.28</u>	<u>-0.29</u>
6	0.94703	0.15	1.11	0.99	0.22	1.04	0.60	0.13
		<u>0.15</u>	<u>1.14</u>	<u>0.99</u>	<u>0.22</u>	<u>1.03</u>	<u>0.58</u>	<u>0.09</u>
7	1.02617	-0.28	0.75	0.77	-0.10	1.13	0.27	-0.80
		<u>-0.28</u>	<u>0.75</u>	<u>0.77</u>	<u>-0.10</u>	<u>1.13</u>	<u>0.25</u>	<u>-0.79</u>
8	1.14125	0.12	1.10	0.98	-0.04	0.95	0.17	-0.19
		<u>0.14</u>	<u>1.07</u>	<u>0.99</u>	<u>-0.04</u>	<u>0.94</u>	<u>0.13</u>	<u>-0.17</u>
9	1.32708	0.08	1.31	0.93	0.01	1.05	0.36	-0.08
		<u>0.09</u>	<u>1.40</u>	<u>0.93</u>	<u>0.01</u>	<u>1.06</u>	<u>0.37</u>	<u>-0.05</u>
10	1.87295	-0.22	1.41	0.73	-0.26	1.11	0.17	-0.97
		<u>-0.23</u>	<u>1.35</u>	<u>0.72</u>	<u>-0.27</u>	<u>1.13</u>	<u>0.16</u>	<u>-0.98</u>

Comments: For the ancillary A_2 also the criteria measures vary with little discernible pattern. But v_0 and v_1 perform quite competitively.

Table 8.6

Conditional coverage probabilities of the τ- and t-intervals using v_1 and v_0 and PCV for population 5, using ancillary A_2.

Group (h)	τ-interval 99%	τ-interval 95%	τ-interval 90%	t-interval 99%	t-interval 95%	t-interval 90%	PCV
1	99.0	97.0	90.0	100.0	97.0	92.0	0.6657
	<u>99.0</u>	<u>96.0</u>	<u>92.0</u>	<u>100.0</u>	<u>98.0</u>	<u>92.0</u>	<u>0.6686</u>
2	100.0	94.0	92.0	100.0	95.0	92.0	0.8492
	<u>100.0</u>	<u>95.0</u>	<u>92.0</u>	<u>100.0</u>	<u>95.0</u>	<u>92.0</u>	<u>0.8605</u>

Table 8.6 (continued)

Group (h)	τ-interval 99%	95%	90%	t-interval 99%	95%	90%	PCV
3	100.0	97.0	94.0	100.0	98.0	94.0	0.8764
	100.0	97.0	94.0	100.0	97.0	94.0	0.8700
4	100.0	99.0	97.0	100.0	99.0	98.0	0.7288
	100.0	99.0	97.0	100.0	99.0	98.0	0.7129
5	98.0	96.0	92.0	99.0	96.0	93.0	0.8593
	98.0	96.0	92.0	99.0	97.0	94.0	0.8761
6	98.0	95.0	91.0	99.0	95.0	92.0	0.9582
	98.0	95.0	91.0	99.0	95.0	92.0	0.9756
7	100.0	96.0	87.0	100.0	97.0	90.0	0.9704
	100.0	96.0	88.0	100.0	97.0	92.0	0.9589
8	99.0	96.0	93.0	100.0	96.0	93.0	0.9760
	100.0	96.0	92.0	100.0	96.0	93.0	0.9328
9	98.0	95.0	92.0	98.0	96.0	93.0	1.2091
	98.0	95.0	92.0	98.0	96.0	93.0	1.2800
10	100.0	98.0	89.0	100.0	98.0	91.0	1.7798
	100.0	97.0	88.0	100.0	98.0	90.0	1.7387

Comments: Conditional performances in terms of the ancillary A_2 are also quite close for both v_0 and v_1.

Table 8.7

d-values for several populations using ancillary A_1.

	Description of Population: Distribution of U_1	λ	g	d-values ν_1	ν_0
(i)	N(0,1)	2.50	0.4	5.1803	5.3724
(ii)	N(0,1)	2.50	0.5	4.1939	4.3729
(iii)	N(0,1)	2.50	0.6	3.3013	3.4690
(iv)	N(0,1)	2.50	0.8	1.8317	1.9823
(v)	N(0,1)	2.50	0.9	1.3572	1.4941
(vi)	$(\chi_1^2 - 1)/\sqrt{2}$	13.59	0.4	6.7395	6.7278
(vii)	$(\chi_1^2 - 1)/\sqrt{2}$	13.59	0.5	7.2734	7.2560
(viii)	$(\chi_1^2 - 1)/\sqrt{2}$	13.59	0.9	11.1124	11.0695
(ix)*	N(0,1)	8.50	0.4	6.6736	6.6717

* x_1 is obtained by adding 10 to samples from $f_{8.5}(x)$

Comments: With varying populations v_1 beats v_0 and vice versa in terms of the d-criterion of p-124 in respect of A_1.

Table 8.8

d-values for several populations using ancillary A_2

	Description of Population: Distribution of U_i	λ	g	d-values: v_1	v_0
(x)	N(0,1)	2.50	0.5	7.2355	<u>7.3437</u>
(xi)	N(0,1)	2.50	0.6	6.2217	<u>6.3160</u>
(xii)	N(0,1)	2.50	0.8	4.4450	<u>4.5113</u>
(xiii)	N(0,1)	2.50	0.9	3.7094	<u>3.7606</u>
(xiv)	$(\chi_1^2 - 1)/\sqrt{2}$	13.59	0.4	8.3362	<u>8.3113</u>
(xv)	$(\chi_1^2 - 1)/\sqrt{2}$	13.59	0.5	8.6655	<u>8.6337</u>
(xvi)	$(\chi_1^2 - 1)/\sqrt{2}$	13.59	0.6	9.1356	<u>9.0962</u>
(xvii)	$(\chi_1^2 - 1)/\sqrt{2}$	13.59	0.8	10.6162	<u>10.5593</u>
(xviii)*	N(0,1)	8.50	0.4	9.5597	<u>10.0103</u>

* x_1 is obtained by adding 10 to samples from $f_{8.5}(x)$

Comments: With changing populations v_1 beats v_0 and vice versa in terms of the d-criterion of p-124 also in respect of A_2.

<u>Results for Cochran's data :</u>

N = 49, n = 11.

Table 8.9

Detailed performances of v_1 and v_0 for Cochran's data.

$V(t_R)$	V	RB	RS	SL	Mean	Var	$\sqrt{\beta_1}$	β_2-3
585392.5	606282.7	-0.07	9.21	0.52	-0.32	1.50	-0.73	0.24
		<u>-0.06</u>	<u>9.50</u>	<u>0.52</u>	<u>-0.32</u>	<u>1.51</u>	<u>-0.73</u>	<u>0.23</u>

Comments: As in Table 8.1, in this real population case also v_1 and v_0 perform quite closely.

Table 8.10

Coverage Probabilities of the τ- and t-intervals

using v_1 and v_0 and PCV for Cochran's data.

τ-interval 99%	τ-interval 95%	τ-interval 90%	t-interval 99%	t-interval 95%	t-interval 90%	PCV
95.4	88.6	81.5	97.8	92.0	85.5	9.9290
<u>95.3</u>	<u>88.4</u>	<u>81.7</u>	<u>97.7</u>	<u>91.8</u>	<u>85.7</u>	<u>10.1006</u>

Comments: As in Table 8.2, with this real population case also v_0 and v_1 compete quite closely.

BIBLIOGRAPHY

Bellhouse, D.R. (1988). A brief history of random sampling methods. in *Handbook of Statistics, Vol. 6, Ed. Krishnaiah, P.R. and Rao, C.R.*, 1-14.

Brewer, K.R.W. (1979). A class of robust sampling designs for large-scale surveys. *Jour. Amer. Stat. Assoc.* 74, 911-915.

——— (1990). Review of 'Unified theory and strategies of survey sampling', by Chaudhuri, A. and Vos, J, W.E. in *Jour. Off. Stat.* 6, 101-104.

Brewer, K.R.W., Foreman, E.K., Mellor, R.W. and Trewin, D.J. (1977). Use of experimental design and population modelling in survey sampling. *Bull. Int. Stat. Inst.* 47:3, 173-190.

Brewer, K.R.W. and Hanif, M. (1983). Sampling with unequal probabilities. *Springer-Verlag, N.Y.*

Cassel, C.M., Särndal, C.E. and Wretman, J.H. (1977). Foundations of inference in survey sampling. *John Wiley and Sons, N.Y.*

Chaudhuri, A. (1987). Randomized response surveys of finite populations : a unified approach with quantitative data. *Jour. Stat. Plan. Inf.* 15, 157-165.

Chaudhuri, A. and Arnab, R. (1979). On the relative efficiencies of sampling strategies under a super-population model, *Sankhyā C*, 41, 40-53.

Chaudhuri, A. and Mitra, J. (1992). A note on two variance estimators for Rao-Hartley-Cochran Estimator. *Comm. Stat. Theo. and Meth.*, 21, 3535-3543.

Chaudhuri, A. and Mukerjee, R. (1988). Randomized Response : Theory and Techniques. *Marcel Dekker Inc. N.Y.*

Chaudhuri, A. and Stenger, H. (1992). Survey Sampling : Theory and Methods. *Marcel Dekker Inc. N.Y.*

Cochran, W.G. (1977). Sampling techniques, 3rd. ed., *John Wiley, New York.*

Cumberland, W.G. and Royall, R.M. (1988). Does simple random sampling provide adequate balance? *Jour. Roy. Statist. Soc. B*, 50, 118-124.

Cramér, H. (1966). Mathematical methods of statistics. *Princeton Univ. Press.*

Deng, L.Y. and Wu, C.F.J. (1987). Estimation of variance of the regression estimator. *Jour. Amer. Stat. Assoc.*, 83, 568-576.

Fuller, W.A. and Isaki, C.T. (1981). Survey design under superpopulation modelling. In *Current topics in survey sampling, ed. Krewski, D., Platek, R. and Rao, J.N.K. Academic Press*, 199-226.

Godambe, V.P. (1989). Quasi-score *function*, quasi-observed Fisher information and conditioning in survey sampling. *Tech. Rep. Stat. 89-09, Univ. Waterloo.*

Godambe, V.P. and Joshi, V.M. (1965). Admissibility and Bayes estimation in sampling finite populations, I. *Ann. Math. Stat.* 36, 1707-1722.

Godambe, V.P. and Thompson, M.E. (1977). Robust near optimal estimation in survey practice. *Bull. Int. Stat. Inst.* 47:3, 129-146.

Hájek, J. (1971). *Comment on a paper by* Basu, D. *in Foundations of Statistical Inference. ed. Godambe, V.P.* and *Sprott, D.R. (1971)*, 203-242, *Holt, Rinehart, Winston, Toronto.*

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

Hartley, H.O. and Rao, J.N.K. (1962). Sampling with unequal probabilities and without replacement. *Ann. Math. Stat.* 33, 350-374.

Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe, *Jour. Amer. Statist. Assoc.*, 47, 663-685.

Isaki, C.T. and Fuller, W.A. (1982). Survey design under the regression superpopulation model. *Jour. Amer. Statist. Assoc.* 77, 89-96.

Kott, P.S. (1990 a). Recently proposed variance estimators for the simple regression estimator. *Jour. Off. Stat.* 6, 451-454.

——— (1990 b). Estimating the conditional variance of a design-consistent regression estimator. *Jour. Stat. Plan. Inf.* 24, 287-296.

Kumar, P., Gupta, V.K. and Agarwal, S.K. (1985). On variance estimation in unequal probability sampling. *Aust. Jour. Stat.* 27, 195-201.

Ohlsson, E. (1989). Variance estimation in Rao-Hartley-Cochran procedure, *Sankhyā B.*, 51, 348-361.

Politz, A. and Simmons, W. (1949). I An attempt to get the 'not at homes' into the sample without callbacks. II Further theoretical considerations regarding the plan for eliminating callbacks. *Jour. Amer. Statist. Assoc.* 44, 9-32.

——— (1950). Note on an attempt to get the 'not at homes' into the sample without callbacks. *Jour. Amer. Statist. Assoc.* 45, 136-137.

Rao, J.N.K., Hartley, H.O. and Cochran, W.G. (1962). On a simple procedure of unequal probability sampling without replacement, *Jour. Roy. Statist. Soc. B*, 24, 482-491.

Rao, J.N.K. and Wu, C.F.J. (1983). Methods for standard errors and confidence intervals from sample survey data : some recent work. *Invited Paper in 46-th Session of Int. Stat. Inst.*, 1-16.

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

Royall,R.M. and Cumberland,W.G. (1978a). Variance estimators in finite population sampling. *Jour. Amer. Statist. Assoc.* 73, 351-358.

——— (1978b). An empirical study of prediction theory in finite population sampling : *simple random sampling and the ratio* estimator. in *'Survey Sampling and Measurement', ed. Namboodiri, N.K.*, 293-309, *Academic Press, N.Y.*

——— (1981a). An empirical study of the ratio estimator and estimators of its variance. Jour. Amer. Statist. Assoc. 76, 66-77.

——— (1981b). The finite population linear regression estimator and estimator of its variance — an empirical study. *Jour. Amer. Stat. Assoc.*, 76, 924-930.

——— (1985). Conditional coverage properties of finite population *confidence* intervals. *Jour. Amer. Stat. Assoc.*, 80, 355-359.

Royall,R.M. and Eberhardt,K.R. (1975). Variance estimators for the ratio estimator. Sankhya C, 37, 43-52.

Särndal,C.E. (1980). On π-inverse weighting versus best linear weighting in probability sampling . *Biometrika*, 67, 639-650.

——— (1982). Implications of survey design for generalized regression estimation of linear *function*. *Jour. Stat. Plan. Inf.* 7, 155-170.

——— (1984). Design-consistent versus model-dependent estimation for small domains. Jour. Amer. Statist. Assoc. 79, 624-631.

Särndal,C.E. and Hui,T.K. (1981). Estimation for non-response situations : to what extent must we rely on models? in *Current topics in sampling, Ed. Krewski,D., Platek,R., and Rao,J.N.K. Academic Press, N.Y.* 227-246.

Särndal,C.E., Swensson,B.E. and Wretman,J.H. (1989). The weighted residual technique for estimating the variance of the general regression estimator of the finite population total. *Biometrika*, 76, 527-537.

——— (1992). Model asisted survey sampling. *Springer-Verlag, N.Y.*

Schaible,W.L. (1979). A composite estimator for small area statistics. In *'Synthetic estimates for small areas'. Steinberg,J.* 36-53.

——— (1992). *Use of* small area estimators in U.S. Federal Programs. *Invited Paper, Int. Sc. Conf. On Small area statistics and survey designs, Warsaw.*

Smith,H.F. (1938). An empirical law describing heterogeneity in the yields of agricultural crops. *Jour. Agri. Sci.* 28, 1-23.

Wolter,K.M. (1985). Introduction to variance estimation. *Springer Verlag, N. Y.*

Wu,C.F.J. (1982). Estimation of the variance of the ratio estimator. *Biometrika*, 69, 183-189.

Wu,C.F.J. and Deng,L.Y. (1983). Estimation of variance of the ratio estimator: an empirical study. in *Scientific Inference , Data Analysis and Robustness, ed Box,G.E.P. et al. Acad. Press,* 245-277.

Yates,F. and Grundy,P.M. (1953). Selection without replacement from within strata with probability proportional to size. *Jour. Roy. Stat. Soc. B,* 15, 253-261.