THE RATIO, MEAN-OF-THE-RATIOS AND
HORVITZ-THOMPSON ESTIMATORS UNDER
THE CONTINUOUS VARIABLE MODEL

## BY

## ANTHONY ALIFA CHAMWALI

B. SC., UNIVERSITY OF EAST AFRICA: THE UNIVERSITY COLLEGE, DAR-ES-SALAAM, 1970
A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE (BUS. ADMIN.)
IN THE FACULTY
OF
COMMERCE AND BUSINESS ADMINISTRATION
WE ACCEPI THIS THESIS AS CONFORMING TO THE REQUIRED STANDARD

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Commerce and Business Administration

The University of British Columbia Vancouver 8, Canada

Date $\qquad$ April., 1974

## ABSTRACT

This study investigates the performances of the ratio estimator, the mean-of-the-ratios estimator and the Horvitz-Thompson (HT) estimator under the continuous variable model of Cassel and Sarndal (1972a, 1972b, 1973). Under this model, the character, Y, which is of interest to the investigator is assumed to be related to an auxiliary variable, $X$, by

$$
\begin{array}{ll}
Y\left(X_{i}\right)=\theta\left(X_{i}+z\left(x_{i}\right)\right) & \text { where } \\
\varepsilon\left(z_{i} \mid x_{i}\right)=0 & \forall x_{i} \varepsilon(0, \infty) \\
\varepsilon\left(z_{i}^{2} \mid x_{i}\right)=\sigma^{2}\left(x_{i}\right) \quad=k^{2} x_{i}^{g} \\
\varepsilon\left(z_{i} z_{j} \mid x_{i} x_{j}\right)=0 & (i \neq j)
\end{array}
$$

It is assumed, in this paper, that $X$ is gamma distributed over $(0, \infty)$ with parameter r. The mean of $Y$ is to be estimated, under the additional assumptions that the design function, $P(X)$, is 1) polynominal 2) exponential, i.e.

$$
\begin{aligned}
& \text { 1) } P(X)=\frac{\Gamma(r) x^{m}}{\Gamma\left(m^{\dagger} r\right)} \\
& \text { 2) } P(X)=(1-c) r e^{c x}
\end{aligned}
$$

It is observed that for $g=0$ or 1 , the ratio estimator performs better than the other two. For $\mathrm{g}=0,1$ or 2 , and for a wider range of values of $m$ or $c$, the mean-of-theratios estimator performs better than the HT estimator.

When $P(X)$ is polynominal, the $H$ estinator is most efficient if the sampling design is approximately pps. The results compare wall with those of other researchers under similar assumptionso

## TABLE OF COITENTS

Page
ABSTPACT ..... i
ACKIOMLEDGEMENTS ..... vi
1 INTHODUCTION
$\frac{1}{1}$
Classical Sampling Theory ..... 4
The Sample Survey Theory
6
Search for Optimal Estimators and Sampling Designs.
The Ratio, Mean-of-the-ratios and HI Estimators ..... 10
2 THE PROBLEA. ..... 14
Statement of the Problem ..... 14
Method Used and Probleras of Evaluation. ..... 17
3 THE RESULTS.
21
Results for $p(x)$ Pojynominal $\dot{\operatorname{lon}} \dot{\mathrm{p}} \dot{\mathrm{k}} \dot{\mathrm{O}} . \dot{5}$ ..... 21
Results for $P(x)$ Polynominal and $k=1.0$
Results for $P(x)$ Polynominal and $k=1.0$ ..... 23 ..... 23
Results for $p(x)$ Exponential and $k=0.5$ ..... 37 ..... 45
Results for $P(x)$ Exponential and $k=1.0$
Results for $P(x)$ Exponential and $k=1.0$
Sumnary of Results and Some Conclusions ..... 52
4 DISCUSSION OF RESULTS ..... 55
Some Empirical Comparisons ..... 55
Comments and Some Implications ..... 60
Some Limitations
64
64
Some Recommendations ..... 65
REFERENCES ..... 66

## LIST OF TABLES

Table
Number
Page

I Variances of the Estimators when $X$ is Gamma Distributed, $P(x)$ is Polynominal, $k=0.5$, $g=0$ 24

II Variances of the Estimators when $x$ is Gamma Distributed $P(x)$ is Polynominal, $k=0.5$,
III Variances of $\hat{\mu} Y H T$ and $\hat{\mu} Y m r$ when $X$ is Gamma Distributed, $P(x)$ is Polynoninal, $k=0.5$,

IV Variances of the Estimators when $X$ is Gamma Distributed, $P(x)$ is Polynominal., $k=1$, $g=0$ 。
$V$ Variances of the Estimators when $X$ is Gamma Distributed, $P(x)$ is Polynominal, $k=1$, $g=1$
VI $\operatorname{Var}\left(\hat{M}_{Y H T}\right.$ and $\operatorname{Var}\left(\hat{M}_{Y m r}\right)$ when $x$ is Garma Distributed $P(x)$ is Polynominal, $k=1$, $g=2$.

VII Variances of the Estimators when $X$ is Gamma Distributed, $P(x)$ is Exponential, $k=0.5$, $g=0$

VIII Variances of the Estimators when $X$ is Gamma Distributed, $P(x)$ is Exponential, $k=0.5$,
 Distributed, $P(x)$ is Exponential, $k=0.5$, $g=2$
$X$ Variances of the Estimators when $X$ is Gamma Distributed, $P(x)$ is Exponential, $k=1, g=0$
XI Variances of the Estimators when $x$ is Ganma Distributed, $P(x)$ is Exponential, $k=1, g=1.43$
XII $\operatorname{Var}\left(\hat{\mu}_{Y H T}\right)$ and $\operatorname{Var}\left(\hat{H}_{Y m r}\right)$ when $X$ is Garama Distributed, $P(x)$ is Exponenijal, $k=1, g=250$

## LIST OF FIGURES

Figure
Number
1 Variances of the Estimators when $X$ is Gamma Distributed, $P(x)$ is Polynominal, $k=0.5$, $g=0$25

2 Variances of the Estimators when $X$ is Gamma Distributed, $P(x)$ js Polynominal, $k=0.5$, $\mathrm{g}=\mathrm{I}$
3 Variances of $\hat{\mu}_{Y H T}$ and $\hat{\mu}_{Y}$ mr when $x$ is Ganma Distributed, $P(x)$ is Polynominal, $k=0.5$, $g=2$

4 Variances of the Estimators when $X$ is Gamma Distributed, $P(x)$ is Polynominal, $k=1$, $g=0$

5 Variances of the Estimators when $X$ is Gamma Distributed, $P(x)$ is Polynominal, $k=1$, $g=1$
$6 \operatorname{Var}\left(\hat{M}_{Y H T}\right)$ and $\operatorname{Var}\left(\hat{M_{Y m r}}\right)$ when $x$ is Garnma Distributed, $P(x)$ is Polynominal, $k=1$, $g=2$35

7 Variances of the Estimators when $X$ is Garma
Distributed, $P(x)$ is Exponential, $k=0.5$,
$\mathrm{g}=0$

8 Variances of the Estimators when $X$ is Gamma Distributed, $P(x)$ is Exponential, $k-0.5$, $g=1$
$19 \operatorname{Var}(\hat{H} Y H T)$ and $\operatorname{Var}(\hat{M} Y m r)$ when $X$ is Gamma Distributed, $P(x)$ is Exponential, $k=0.5$, $g=2$

10 Variances of the Estimators when $X$ is Gamma Distributed, $P(x)$ is Exponential, $k=1$, $\mathrm{g}=0$. . . . . . . . . . . . . . . . . . 47

11 Variances of the Estimators when $X$ is Gamal Distributed, $P(x)$ is Exponential, $k=1$, $g=1$
$12 \operatorname{Var}\left(\hat{M}_{Y H T}\right)$ and $\operatorname{Var}\left(\hat{M}_{Y m r}\right)$ when $x$ is Gamma Distributed, $P(x)$ is Exponential, $k=1$,


## ACKNONLELGEMENTS

I would like to thank Dr. C. E. Sarndal
for his invaluable advice and help offered throughout this study. I would also like to extend my thanks to the Canadian Goverrment for sponsoring my education here in Canada. Lastly, my appreciation to the Tanzanian Government and the Principal of the Institute of Development Managenent, Tanzania, who, in one way or another, arranged for my study.

## CHAPTER I

## INTRODUCTION

Recently, much emphasis has been placed on the distinction between General Statistical Theory and Sample Survey Theory. It has been felt that the two theories should either be reconciled or be treated differently.

## Classical Sampling Theory

In General Statistical Theory, Sampling Theory involves making inferences from observed sample values to an infinite hypothetical population. The sample values are assumed to have been drawn randomly from a smooth density function defined on the units of the hypothetical population. Godambe (1969, 1970) suggests that the tendency to associate theoretical statistical sampling with an infinite hypothetical population that has a smooth density function may likely be due to the pqay General Statistical Theory developed. The early statisticians were mainly interested in biological and sociological phenomena like inheritance. They assumed that some chance mechanism operated behind these phenomena, and that the chance mechanisin uniquely determined the frequency function of the characteristic under study in the hypothetical population.

When discussing Sample Survey Theory, some optimality properties of estimators from General Statistical Theory are sonetimes encountered. The following are some of these:

Let $Y_{1}, Y_{2}, . . ., Y_{n}$ be a sample from a population with density function $f(Y, \theta)$. Let

$$
\hat{\theta}=d\left(Y_{1}, Y_{2}, \ldots ., Y_{n}^{\prime}\right)
$$

be the estimator of the parameter $\theta$. Since $\hat{\theta}$ is a function of the observed sample values, it is a random variable. Let

$$
\theta^{*}=d^{*}\left(Y_{1}, \ldots ., Y_{n}\right)
$$

by any other statistic which is not a function of $\hat{\theta}$. The estimator $\hat{\theta}$ is sufficient for $\theta$ if, for each $\theta^{*}$, the conditional density of $\theta^{*}$ given $\theta, P\left(\theta^{*} \mid \theta\right)$ does not contain $\theta$. A sufficient statistic, if it exists, contains all the information about the parameter to be estimated.

The estimator, $\hat{\theta}$, is an unbiased estimator of $\theta$ if its expected value, $E(\hat{\theta})$, equals $\theta$. $\hat{\theta}$ is a Uniformly Minimum Variance Unbiased Estimator (UNVUE) of $\theta$, if $\hat{\theta}$ is an unbiased estimator of $\theta$ and its variance is less than or equal to the variance of any other possible estimator, $\hat{\theta}_{p}$, of $\theta$ i.e. $\operatorname{Var}(\hat{\theta}) \leqslant \operatorname{Var}\left(\hat{\theta}_{\mathrm{p}}\right)$ for all $\hat{\theta}$.

Let

$$
\hat{\theta}_{n}=d n\left(Y_{1}, \ldots, Y_{n}\right)
$$

be an estimator of $\theta$ based on a sample of size $n$. The
sequence $\left\{\hat{\theta}_{n}\right\}$ is a consistent estimator of $\dot{\theta}$ if, for every $\varepsilon>0$

$$
\lim _{n \rightarrow \infty} p\left(\theta-\varepsilon<\hat{\theta}_{n}<\theta+\varepsilon\right)=1, \forall \theta
$$

i.e., $\theta_{n}$ approaches $\theta$ as $n$ gets large.

The sequence $\left\{\hat{\theta}_{n}\right\}$ is a Best Asymptotically
Normal (BAN) Estimator of $\theta$ if
(a) The distribution of $\sqrt{n}\left(\hat{\theta}_{n}-\theta\right)$ approaches the normal distribution with mean 0 and variance $\sigma^{2}(\theta)$ as $n$ approaches infinity
(b) For every $\varepsilon>0$.

$$
\lim _{n \rightarrow \infty} P\left\{\left|\hat{\theta}_{n}-\theta\right|>\varepsilon\right\}=0, \forall \theta
$$

(c) There is no other sequence of consistent estimators $\theta_{1}^{*}, \theta_{2}^{*}, \ldots . . \theta_{n}^{*}, \ldots$ for which the distribution of $\sqrt{n}\left(\theta_{n}^{*}-\theta\right)$ approaches the normal distribution with mean 0 and variance $\sigma * 2(\theta)$ and such that

$$
\frac{\sigma^{2}(\theta)}{\sigma * 2(\theta)}>1
$$

for all $\theta$ in some open interval.
If the unbiased $\hat{\theta}=d\left(Y_{1}, Y_{2}, \ldots, Y_{n}\right)$ is a linear function of $Y$ and if $\operatorname{Var}(\hat{\theta}) \leqslant \operatorname{Var}\left(\hat{\theta}_{p}\right)$ for all linear $\hat{\theta}$, then $\hat{\theta}$ is a Uniformly Best Linear Unbiased Estimator (UBLUE) of $\theta$.

It may be worthwhile noting that while general statistical theory aims at making inferences about some frequency function of a hypothetical population, it may
actually be inferring about the possible chance mechanism (discussed above) that produced the given set of observations which may have nothing to do with the hypothetical population at all.

## The Sample Survey Theory

In Sample Survey Theory, the populations considered are finite. They consist of elenents which are real and countable. They are not hypothetical like the ones considered in the General Statistical Theory. This is the main difference between the populations dealt with by the two sampling theories. In this sense the problems of inference in the two models of sampling theory may be taken to be different.

In survey sampling, since the investigator deals with real and finite populations, he has choice over the sampling design and the estimator he may want to use. Occasionally, circumstances may necessitate the use of nonprobability sampling altogether. Cochran (1963) gives the following situations that lead to nonprobability sampling:
(1) When the sample is restricted to a part of the population that is readily accessible.
(2) When the sample is selected haphazardly.
(3) With a small but heterogenous population, the sampler inspects the whole of it and selects a small sample of 'typical' units, i.e., units that are close to his impression of the average of the population. This method is sometimes called judgement or purposive selection.
(4) Lastly, when the sample consists essentially of volunteers.

He notes that these methods may give good results under the right conditions although they are not amenable to the development of a sampling theory owing to lack of irandom selection.

But whether one deals with Sample Survey Theory or General Statistical Theory, he is dealing with the same problem of inferring from the sample to the population usually with the help of Probability Theory and Statistics. In the General Statistical Theory, random sampling solved nearly all the sampling design problems. With survey populations, things are not that much easy. The nature of the population may make random sampling difficult; it may not be possible to identify all the units in the population. In most cases, the frequency function of the character under consideration is unknown. This makes it difficult to check some of the optimality conditions of estimators in the General Statistical Theory. And optimal estimators under the General Statistical Theory need not be so under the Sample Survey Theory. This raises the main problem of finding sampling designs and estimators that are optimal for the sample survey situation. Wathematically, in the sample survey model, we have a set $P$ of $N$ units that constitute the population,

$$
p=\left(U_{1}, u_{2}, \ldots, u_{N}\right)
$$

where $U_{i}$ stands for unit $i$. An unknown quantity $Y_{i}$ which is of interest to the surveyor is associated with $U_{i}$. The surveyor wants to know

$$
\theta=\theta\left(Y_{1}, Y_{2}, \ldots, Y_{i \stackrel{N}{\prime}}\right)
$$

He may also know the auxiliary variable $x$, i.e.

$$
a=\left(x_{1}, x_{2}, \ldots, x_{N}\right)
$$

associated with $P$. He selects a subset

$$
s=\left(u_{1}, u_{2}, \ldots, u_{n}\right)
$$

Of units of $P$ and observes the corresponding

$$
Y=\left(Y_{1}, Y_{2}, \ldots, Y_{n}\right)
$$

(response errors are ignored). The sampling plan generates $S$. He then calculates

$$
\hat{\theta}=\hat{\theta}\left(Y_{1}, Y_{2}, \ldots, Y_{n}\right)
$$

as an estimator of $\theta$. The problem is how to select $S$ (how to assign the probability that unit $i$ of $P$ will be included in S) and how to choose the random variable $\hat{\theta}$ to get an estimate which is as near $\theta$ as possible. Certainly, it is not just a matter of getting a 'representative' sample and calculating the mean, variance and Median of the sample values. Besides, what is a representative sample?

## Search for Optimal Estimators and Sampling Desions

In 1934 Neyman introduced the Gauss-itarkov theorem to obtain a linear unbiased mininum variance estimate for
the mean of a survey population. He established that, for simple random sampling, the sample mean was the minimum variance linear unbiased estimate of the survey population mean. This was an attempt to fit survey sampling into the hypothetical population model, and his findings stimulated most sample-survey statisticians to find, with the help of Gauss-inarkov theorem, efficient (i.e. minimum variance) unbiased estimators for a variety of more complex designs. Sampling procedures like, sampling with arbitrary probabilities, stratified sampling, cluster sampling, 2-stage sampling, multi-stage sampling, etc. were designed to reduce the variance of the estimators (see, for example, Goodman and Kish (1950), Durbin (1953), Hartley and Rao (1962), Cochran (1963). Horvitz and Thompson (1952) attempted to provide a general method for dealing with sampling without replacement from a finite population when variable probabilities of selection are used to the elements remaining prior to each draw. They, in particular, considered a general estimator of the population total of the form

$$
\hat{Y}=\sum_{i=1}^{n} \beta_{i} Y_{i}
$$

where $\beta_{i}(i=1, \ldots, N)$ is a constant to be used as a weight for the $i^{\text {th }}$ unit whenever it is selected for the
sample. Letting $P\left(X_{i}\right)$ to be probability that the $i^{\text {th }}$ element will be included in a sample of size $n$, they showed that

$$
\beta_{i}=\frac{1}{P\left(X_{i}\right)}
$$

makes $Y$ unbiased and of mininum variance ( $X_{i}$ is the value of the auxiliary variable associated with unit i). Godambe (1955, 1965) considered the general
estimator

$$
\hat{Y}=\sum_{i=1}^{n} b s_{i} Y_{i}
$$

where $b s_{i}$ is defined in advance for all the $N^{n}$ logically possible $S$ (samples), and for all in in This is a more general class of estimators than the ones considered by Horvit.z and Thompson. He proved the non-existence of a uniformly minimum variance (UN) unbiased estimator in this class of estimators for any design $P(s)$, excepting those in which no two $S$ with $P(S)>0$ have at least one common and one uncommon unit.

Because of the non-existence of UíV unbiased estimator, other criteria of goocness of an estimator were sought. Godambe and Joshi (1965) considered the criterion of admissibility and proved that the HorvitzThompson (HT) estimator is admissible in the class of all p-unbiased estimators of $Y$, the population total, for any design $P(S)$ such that

$$
\Pi_{j}=\sum_{s J_{u j}} p(s)>0, \forall u_{j}
$$

The admissibility criterion, however, was satisfied by many other estimators, and other new criteria like 'hyper-admissibility,' 'necessary bestness' (Basu, 1971) were introduced. A detailed examination of these optimality properties raise some questions regarding their relevance. For example, the Horvitz and Thompson estimator, $\hat{Y} H T, ~ i s ~ u n i q u e l y ~ ' h y p e r-a d m i s s i b l e ' ~ w h a t e v e r ~$ the character under investigation or the sample design. But the $\hat{Y} H T$ may lead to disastrous results (Sarndal, 1972; Basu, 1971).

Many variations of the regression type population model

$$
y_{i}=a x_{i}+z_{i} \quad i=(I, \ldots, N)
$$

have been used, where $y_{i}$ is the character of interest and $x_{i}$ is an auxiliary variable for unit $i$ in the population; $Z_{i}$ is an error component. Cochran (1946), Godambe (1955, 1965), Royall (1970, 1971) and many others used the superpopulation approach in conjunction with the regression model to compare efficiencies between sampling methods. Cassel and Sarndal (1972a, 1972b, 1973) use the continuous variable model of the form

$$
Y(X)=\theta(X+Z(X)) \quad \forall X \in(0, \infty)
$$

where $(z(x) \mid x)=0, \varepsilon(z(x) \mid x)^{2}=\sigma^{2}(x)$

$$
\varepsilon\left(z\left(x_{1}\right) z\left(x_{2}\right) \mid x_{1} x_{2}\right)=0 \quad\left(x_{1} \neq x_{2}\right)
$$

and the auxiliary variable $X$ is assumed to have a known distribution described by the distribution function $F(x)$ over ( $0, \infty$ ).

Different notions of unbiasedness have been used during this search for optimal estimators and designs: An estimator, $\hat{\bar{Y}}$ for $\bar{Y}$, is called design-unbiased or p-unbiased if

$$
E_{p}(\hat{\bar{Y}})=\sum_{s \varepsilon S} P(s) \hat{\bar{Y}}=\bar{Y}
$$

for all vectors $\left(Y_{1}, \ldots, Y_{N}\right)$, where $P(s)$ is a probability function defined on the set $S$ of subsets of $s$ of labels ( $1, \ldots, \ldots, N$ ).

An estimator is called model-unbiased or $\mathcal{E}$-unbiased if it is unbiased under the assumptions of the specified model.

Lastly, an estimator $\hat{\bar{Y}}$ is called Ep $\hat{\mathcal{G}}$-unbiased for $\bar{Y}$ if

$$
\operatorname{Ep} \varepsilon(\hat{\bar{Y}})=\sum_{\mathrm{s} \& S} P(S) \varepsilon(\hat{\bar{Y}})=\bar{Y}
$$

an estimator can be p-unbiased but not -unbiased and vice versa.

## The Ratio, Mean-of-the-Patios and the HT Estimators

It is a fact that using supplementary information in sampling designs and estimators greatly improves the accuracy of the estimates. Three estimators that have
received considerable attention in the literature of sample survey theory and which make use of supplementary information are the Ratio estimator, the Mean-of-the-ratios estimator and the Horvitz-Thompson (HI) estimator. If X is an auxiliary variable, the ratio estimator for the population mean, $\mathcal{M Y}$, is given by

$$
\begin{equation*}
\hat{\mu}_{Y R}=\frac{y}{x} \quad \bar{x}=\frac{\bar{y}}{x} \quad \bar{x} \tag{1.1}
\end{equation*}
$$

where $y=\sum_{i=1}^{n} Y_{i}, X=\sum_{i=1}^{n} x_{i}, \bar{X}=\frac{1}{N} \sum_{i=1}^{N} X_{i}, \bar{Y}=\frac{1}{n} \sum_{l=1}^{n} Y_{i}$, $\bar{x}=\frac{1}{n} \sum_{i=1}^{n} X_{i}, n$ is the sample size and $N$ is the population size.

The mean-of-the-ratios estimator is

$$
\begin{equation*}
\hat{\mu}_{y_{m r}}=\frac{\bar{x}}{n} \sum_{i=1}^{n} \frac{y_{i}}{x_{i}} \tag{1.2}
\end{equation*}
$$

The ratio estimators make use of the ratios $\frac{y}{X}$ and $\frac{Y_{i}}{X_{i}}$ in order to improve estimation of $\mathcal{H}$. With simple random sampling from an infinite population, the ratio estimator of My has been shown to be the best linear unbiased estimate if two conditions are satisfied (Cochran , 1963, p. 166, Theorem 6.4):
(1) The relation between $y_{i}$ and $X_{i}$ is a straight line through the origin
(2) The variance of $y_{i}$ about this line is proportional to $X_{i}$.

When the variance of $y_{i}$ about this line is proportional to $X_{i}^{2}$, using the mean-of-the-ratios estimator gives much better performance than the other estimator.

When the relation between $y_{i}$ and $X_{j}$ is linear but the line does not go through the origin, the regression estimator

$$
\hat{\bar{y}}_{l_{r}}=\bar{y}+b(\bar{x}-\bar{x})
$$

performs much better than the others. $\hat{\bar{Y}}_{1 n}$ reduces to $\bar{y}$ if $b=0$ and to $\hat{M}_{R Y}$ if $b=\frac{\bar{Y}}{\bar{X}}$. Both the ratio and regression estimators are consistent and, generally, slightly biased, but with sampling designs like stratified sampling, the bias of the ratio estimator may be considerable. This has led to searching unbiased or better ratio-type estimators. J.N.K. Rao (1969) gives some results of the search. The results indicate that under certain conditions, other ratio-type estimators perform better than the traditional ratio estimator.

The Horvitz-Thompson estimator is given by

$$
\begin{equation*}
\hat{\mu}_{Y H T}=\frac{1}{n i} \sum_{i=1}^{n} \frac{y(x i)}{P(x i)} \tag{1.3}
\end{equation*}
$$

where $P\left(x_{i}\right)$ is the probability that a unit with auxiliary variable $x_{i}$ will be included in the sample. The variance of the HT estimator may be negative, or it may not reduce to zero even when all the $Y$-values are the same and the variance should actually be zero.

In many studies the HT estimator has been shown to compete very well with the ratio estimators. Assuming a linear stochastic model of the form

$$
y_{i}=\alpha+\beta x_{i}+z_{i}, \quad \sigma_{i}^{2}=\sigma^{2} x_{i}^{2 \gamma},
$$

Foreman and Brewer (1971) showed that the HT estimator is
more efficient than the ratio estimator with equal selection probabilities if $\gamma>\frac{1}{2}$, but they caution that if the sampling fraction is large and $\alpha$ is appreciable, the ratio estimator may be made more efficient than the HT estimator. Rao (1967) gives some conditions under which the mean-of-the-ratios estimator is superior to both the ratio estimator and the HT estimator. Unfortunately, the mean-of-the-ratios estimator is not consistent like, say, the ratio estimator (Sukhatme and Sukhatme, 1970, p. 1 60).

## CHAPTER II

## THE PROBLEA

## Statement of the Proolem

The problem with simple random sampling is that it does not take into account the possible importance of the larger units in the population. Because of this, sampling designs like sampling with probability proportional to size (pps), and generally known as sampling with varying probabilities came to be used. These are more complex designs than simple random sampling. It was also realized that it makes a difference whether sampling is done with replacement or without replacement. However, the estimators considered were very complicated. To get simple estimators, sampling schemes like the Midzumo system of sampling, the Narain Method of sampling, Systematic Sampling with varying probabilities etc. were introduced. It is very clear, from such studies, that performances of estimators can be improved by, not only making use of supplementary information, but also by choosing the proper sampling design.

The ratio estimator, the mean-of-the-ratios
estimator and the Horvitz-Thompson estimator have performed quite efficiently in some of the research work in sample survey literature. But which of these three does better
estimation than the others depends on the sampling design used, the way the supplementary information has been used, the class of estimators used and the way 'bias' or 'unbiased' is defined. Joshi (1971), for example, states that
(i) The HT estimate is always admissible in the class of all unbiased estimates, linear and non-linear.
(ii) In the entire class of estimates, the HT estimator is admissible if the sampling design is of fixed sample size.
(iii) If the loss function, $V(t)$, where $t$ is the numerical difference between the estimated and true values, satisfies only
(a) $V(t)$ is non-decreasing in $[0, \infty]$,
(b) for every $K>0$
$\int_{0}^{\infty} v(t) \exp \left(-\frac{K t^{2}}{2}\right) d t<\infty$
then the sample mean, and more generally the ratio estimate is always, admissible as an estimate of the population mean. The use of these three estimators with the regression type models has shown to give very promising results. I would like to investigate the performances of these three estimators (the ratio estimator, mean-of-theratios estimator and the HT estinator) under the regression type continuous variable hodel of Cassel and Sarndal (1973).

I would like to consider the model

$$
\begin{aligned}
& \left.Y\left(x_{i}\right)=\theta\left(x_{i}\right)+z\left(x_{i}\right)\right) \\
& \varepsilon\left(z_{i} \mid x_{i}\right)=0 \quad \forall x_{i} \varepsilon(0,0) \\
& \varepsilon\left(z_{i}^{2} \mid x_{i}\right)=\sigma^{2}\left(x_{i}\right)=k^{2} x_{i}^{g} \\
& \varepsilon\left(z_{i} z_{j} \mid x_{i} x_{j}\right)=0 \quad(i \neq j)
\end{aligned}
$$

where $Y_{i}$ is the value of the character under investigation for unit $i, x_{i}$ is the value of the corresponding auxiliary variable. I further assume that the auxiliary variable $X$ is gamma distributed over $(0, \infty)$ ie.

$$
f(x)=\frac{1}{\Gamma(r)} x^{r-1} e^{-x} \quad \text { for } x \varepsilon(0, \infty)
$$

Hence its mean

$$
E(x)=\int_{0}^{\infty} x f(x) d x=r
$$

My aim is to estimate the population mean

$$
E_{p} \varepsilon(Y(x))=\theta \underline{r}=\mu_{Y}
$$

The three estimators of the population mean that I would like to consider are

$$
\begin{align*}
& \hat{\mu}_{Y H T}=\frac{1}{n} \sum_{i=1}^{n} \frac{y\left(x_{i}\right)}{p\left(x_{i}\right)}  \tag{2.1}\\
& \hat{\mu}_{Y \mathrm{Yr}}=\frac{r}{n} \sum_{i=1}^{n} \frac{y\left(x_{i}\right)}{x_{i}}  \tag{2.2}\\
& \hat{\mu}_{Y R}=r \frac{\sum_{i=1}^{n} y\left(x_{i}\right)}{\sum_{i=1}^{n} x_{i}} \tag{2.3}
\end{align*}
$$

Where the design function, $P\left(x_{i}\right)$ is the selection probability density of a unit with auxiliary variable $x_{i}, n$ is the number of units in the sample. Draws are made independently of each other according to the same distribution of inclusion probabilities. (2.1) is the Horvitz-Thompson estimator, (2.2) is the mean-of-theratios estimator and (2.3) is the classical ratio estimator.

The three estimators are Epfunbiased and their $M S E$ 's, then, equal their EpEvariances.

## Method used and Problems of Evaluations

The efficiencies of these estimators will be evaluated in terms of their EpEvariances, which are obtained from

$$
\operatorname{var}\left(\hat{\mu}_{Y}\right)=E_{p} \varepsilon\left(\hat{\mu}_{y}{ }^{2}\right)-\left(E_{p} \varepsilon\left(\hat{\mu}_{y}\right)\right)^{2}
$$

Generally, the expressions for the variances of these estimators are

$$
\begin{aligned}
& \operatorname{Var}\left(\hat{\mu}_{Y H T}\right)=\frac{\mu^{2} y}{n}\left(r^{-2} \int_{0}^{\infty} \frac{x^{2}+k^{2} x^{g}}{p(x)} f(x) d(x)-1\right) \\
& \operatorname{Var}\left(\hat{\mu}_{y m r}\right)=\frac{\mu_{y}^{2}}{n} \int_{0}^{\infty} k^{2} x^{g-2} p(x) f(x) d x \\
& \operatorname{Var}\left(\hat{\mu}_{Y R}\right)=M_{Y}^{2} \int_{0}^{\infty} \frac{\sum_{i=1}^{n} k^{2} x_{i}^{g}}{\left(\sum_{i-1}^{n} x_{i}\right)^{2}} p(\underline{x}) f(x) d(\underline{x})
\end{aligned}
$$

where $P(\underline{x})=\prod_{i=1}^{n} P\left(x_{i}\right), \quad f(\underline{x})=\prod_{i=1}^{n} f\left(x_{i}\right), \quad d(\underline{x})=\prod_{i=1}^{n} d x_{i}$
and the last integral is $n$-dimensional (over $0<x_{i}<\infty$, for $i=1, \ldots, n$ ). The variances depend essentially on three things, namely, the shape of the continuous distribution, $f(x)$, of the auxiliary variable, $x$; the variance of $\mathrm{Z}_{\mathrm{i}}$ and the selection probability density, $\mathrm{P}(x)$. I will compare the variances under 1) polynomial $p(x)$, 2) exponential $p(x)$.
(1) Let

$$
P(x)=\frac{\Gamma(r) x^{m}}{\Gamma\left(m^{+r}\right)}
$$

( $m=0$ represents simple random sampling, $m=1$ represents pps sampling scheme). In this case the expressions for the variances of these estimators are:
$\left.\operatorname{var} \hat{\mu}_{\text {YHT }}\right)=\frac{\mu_{y}^{2}}{n}\left\{\frac{\Gamma\left(\frac{m+r}{}\right)}{(\Gamma(r+1)}\left[\Gamma(2+r-m)+k^{2} \Gamma\left(g^{+r-m}\right)\right]-1\right\}(2.4)$
$\operatorname{var}\left(\hat{\mu}_{y m r}\right)=\frac{k^{2 \Gamma}\left(m^{+r}+g-2\right)}{\Gamma\left(m^{+r}\right)} \cdot \frac{\mu_{y}^{2}}{n}$
for $g=0$ :
$\operatorname{Var}\left(\hat{\mu}_{Y R}\right)=\frac{n k^{2}}{(n m+n r-1)(n m+n r-2)} \cdot M_{y}{ }^{2}$

$$
\begin{equation*}
\left.\bumpeq \quad \frac{\mu y^{2}}{n} \cdot \frac{k^{2}}{(m+r)^{2}} \quad \text { (if } n \text { is large }\right) \tag{2.6}
\end{equation*}
$$

for $g_{1}=1:$
$\operatorname{var}\left(\hat{M}_{Y R}\right)=\frac{k^{2}}{n m^{2} n r-1} \cdot \mu_{y}{ }^{2}$

$$
\begin{equation*}
\bumpeq \quad \frac{\mu_{y^{2}}}{n} \cdot \frac{k^{2}}{m+r} \quad(\text { for } n \text { large }) \tag{2.7}
\end{equation*}
$$

I evaluate these expressions numerically for $r=1,2,3$ and $k$ and $g$ fixed at 0.5 and 0 respectively, 0.5 and 1 respectively, 0.5 and 2 respectively, 1 and 0 respectively, etc., and as $m$ takes the values $-1(0.5) 2,3,4$. I will observe how the variances of these estimators behave.
(2) Let

$$
P(x)=(1-c)^{x} e^{c x}, \quad c \leqslant 1
$$

( $c=0$ represents simple random sampling). In this case the expressions for the variances are:
$\operatorname{Var}\left(\hat{H}_{\mathrm{YHT}}\right)=\frac{\mu^{2}}{n}\left\{\frac{1}{r^{2}(1-c) r(r)}\left[\frac{\Gamma(2+r)}{(c+1)^{2+r}}+\frac{k^{2} \Gamma(q+r)}{(c+1)^{+1}}\right]-1\right\}$ (2.8)
$\operatorname{Var}\left(\hat{M}_{y m r}\right)=\frac{\mu_{y}^{2}}{n} \cdot \frac{k^{2} P\left(g^{+r}-2\right)}{1(r)(1-c) g-2}$
for $\mathrm{g}=0$ :
$\operatorname{var}\left(\hat{\mathcal{H}}_{y R}\right)=\mu_{y}^{2} \cdot \frac{n k^{2}(1-c)^{2}}{(n r-1)(n r-2)}$

$$
\begin{equation*}
\bumpeq \quad \frac{\mu_{y}^{2}}{n} \cdot \frac{k^{2}(1-c)^{2}}{r^{2}} \text { (for } n \text { large) } \tag{2.10}
\end{equation*}
$$

for $g=1$ :
$\operatorname{var}\left(\hat{\mu}_{y R}\right)=\frac{\mu_{y}^{2} k^{2}(1-c)}{n r-1}$

$$
\begin{equation*}
\bumpeq \quad \frac{\mu_{y}^{2}}{n} \cdot \frac{k^{2}(1-c)}{r} \text { (for } n \text { large) } \tag{2.11}
\end{equation*}
$$

As with the polynomial case, these expressions are evaluated and compared as values of $r, c, g, k$ vary. It has not been possible to evaluate the variance
of the ratio estimator when $g=2$, only the variances of the HT estimator and mean-of-the-ratios estimator will be compared when $g=2$. With exponential $p(x)$, the variance of the mean-of-the-ratios estimator is finite only if $r+g>2$. Throughout, I have used the approximate expressions of $\operatorname{Var}\left(\hat{\mu}_{y R}\right)$ for my evaluations. The results are tabulated below and illustrated by graphs.

## CHAPTER 3

THE RESULTS

Results for $P(x)$ Polynominal and $k=0.5$
The results are presented in Tables I -. III and illustrated by figures $1-3$.

- When $g=0$ the ratio estimator has, generally, the smallest variance for fixed values of $r$ and $m$, followed by the mean-of-the-ratios estimator and then by the HT estimator. When $r=1$ not much comparison can be made between the HT estimator and the mean-of-the-ratios estimator as the variance of the $H T$ estimator is finite for $-1.0<m<1$ while the variance of the mean-of-the-ratios estimator is finite for $\mathrm{m}>\mathrm{l}$. In the range of values of m for which the variances of the ratio estimator and of either the HT estimator or the mean-of-the-ratios estimator are finite, the ratio estimator has smaller variance. When $r=2$ and $0<m<2$ where the variances of all the three estimators are finite, the ratio estimator has the smallest variance; the HT estimator initially has a smaller variance than the mean-of-the-ratios estimator. The variance of the HT estimator takes its minimum value at about $m=1.0$ and for $m>1.0$ the mean-of-the-ratios estimator's variance is smaller than that of the $H T$ estinator. The variance of the HP estimator starts at infinity, takes its minimun value at
$m=0.5$ when $x=1$, and $m=1.0$ when $r=2$ or 3 , and goes to infinity as makes higher values. When $r=3$ the ratio estimator is most efficient followed by the mean-of-the-ratios estimator and the. HI estimator is last. The variances of the ratio and mean-of-the-ratios estimators get smaller and smaller as m increases. The same is true with changes in the value of $r$. This can al. so be seen by observing formulas (2.5) and (2.6) for the variances of the two estimators.

When $g=1$ the ratio estimator still performs better than the other two. When $r=1$ and $m<l$, the HT estimator has smaller variance than the mean-of-the-ratios estimator. The reverse is true for m>l.

When $r=2$ or 3 , the mean-of-the-ratios estimator has smaller variance than the HT estimator except when the variance of the HT estimator takes its minimum value equal to the value of the variance of the mean-of-the-ratios estimator. When $r=1$ the $H T$ estimator beats the mean-of-the-ratios estimator for $1<\mathrm{m}$, the mean-of-the-ratios estimator beats the $H I$ estimator when $m>l$. The varience of the $H T$ estimator takes its minimum value at $m=1.0$, and the variances of the ratio and mean-of-the-ratios estimators become smaller and smaller as mets larger and larger. When $g=2$ the mean-of-the-ratios estinator's variance is a constant and smaller than that of the HT
estimator (except when the variance of the HT estimator takes its minimum value equal to the value of the variance of the mean-of-the-ratios estimator). The variance of the HT estimator is smallest at $m=1.0$.

Generally, as $m$ increases, the variances of the ratio and the mean-of-the-ratios estimators are reduced, but the extent to which these variances can be reduced by, say, a given big $m$ is lessened as $g$ takes higher values. At $m=4$, for example, the variances of the ratio and the mean-of-the-ratios estimators are approximately zero for $g=0,0.05 \mathrm{M}_{\mathrm{y}}^{2} / \mathrm{n}$ for $\mathrm{g}=\mathrm{J}$, and the mean=of-the-ratios estiniator has a constant variance of $0.25 \mathrm{k}_{\mathrm{y}}^{2} / \mathrm{n}$ when $\mathrm{g}=2$. For large values of $m$, the difference between the variance of the ratio estimator and that of the mean-of-the-ratios estimator is very small and not very much dependent on the value of $r$. The smallest possible value taken by the variance of the $H T$ estimator and the extent to which it is finite is greatly affected by the value r takes. As r takes bigger values, the variance of the HT estimator can thake smaller values and is finite for a wider range of the values of $m$.

Results for $P(x)$ Polynominal and $k=1$
The results are given in tables IV-VI and
illustrated by Figures 4-6.

$$
\text { When } g=0 \text { and } r=1 \text {, the variances of the hi }
$$

TABLE I
Variances of the Estimators when X is Gamma Distributed, $P(x)$ is Polynominal, $k=0.5, g=0$. (Each entry should be multiplied by $\mu_{y}^{2 / n}$ )

| m | $r=1$ |  |  | $r=2$ |  |  | $r=3$ |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\hat{\mu}_{\mathrm{YHT}}$ | $\hat{H}_{Y \mathrm{mr}}$ | $\hat{M}_{Y R}$ | $\hat{H}_{Y H T}$ | $\hat{\mu}_{Y \mathrm{mr}}$ | $\hat{\mu}_{Y R}$ | $\hat{\mu}_{Y H T}$ | $\hat{\mu}_{Y m r}$ | $\hat{\mu}_{Y R}$ |
| $-1.0$ | $\infty$ |  | $\infty$ | 5.13 |  | 0.25 | 2.38 | $\infty$ | 0.063 |
| -0.5 | 5.28 |  | 1.00 | 1.65 |  | 0.11 | 0.97 | 0.33 | 0.040 |
| 0.0 | 1.25 |  | 0.25 | 0.56 | $\infty$ | 0.063 | 0.36 | 0.13 | 0.028 |
| 0.5 | 0.57 |  | 0.11 | 0.22 | 0.33 | 0.040 | 0.11 | 0.067 | 0.020 |
| 1.0 | $\infty$ | $\infty$ | 0.063 | 0.13 | 0.13 | 0.028 | 0.042 | 0.042 | 0.016 |
| 1.5 | $\infty$ | 0.33 | 0.040 | 0.47 | 0.067 | 0.020 | 0.15 | 0.029 | 0.012 |
| 2.0 |  | 0.13 | 0.028 | $\infty$ | 0.042 | 0.016 | 0.50 | 0.021 | 0.010 |
| 3.0 |  | 0.042 | 0.016 |  | 0.021 | 0.010 | $\infty$ | 0.013 | 0.0069 |
| 4.0 |  | 0.02 | 0.01 |  | 0.013 | 0.0069 |  | 0.0083 | 0.0050 |


'TABLE II
Variances of the Estimators when $X$ is Gamma Distributed, $P(x)$ is Polynominal, $k=0.5, g=1$. (Each entry should be multiplied by $\mu_{Y}^{2 / n)}$.

| m | $\underline{r}=1$ |  |  | $r=2$ |  |  | $x=3$ |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\hat{\mu}_{Y H T}$ | $\hat{\mu}_{Y m r}$ | $\hat{\mu}_{\mathrm{YR}}$ | $\hat{\mu}_{\mathrm{YHT}}$ | $\hat{H}_{Y m r}$ | $\hat{H}_{\mathrm{YR}}$ | $\hat{\mu}_{\mathrm{YHI}}$ | $\hat{\mu}_{Y m r}$ | $\hat{H}_{Y R}$ |
| $-1.0$ | $\infty$ |  | $\infty$ | 5.37 | $\infty$ | 0.25 | 2.50 | 0.25 | 0.13 |
| -0.5 | 5.48 |  | 0.5 | 1.762 | 0.50 | 0.17 | 1.04 | 0.17 | 0.10 |
| 0.0 | 1.25 | $\infty$ | 0.25 | 0.63 | 0.25 | 0.13 | 0.42 | 0.13 | 0.083 |
| 0.5 | 0.37 | 0.5 | 0.17 | 0.21 | 0.17 | 0.10 | 0.14 | 0.10 | 0.073 |
| 1.0 | 0.25 | 0.25 | 0.13 | 0.13 | 0.13 | 0.083 | 0.08 | 0.083 | 0.063 |
| 1.5 | 2.52 | 0.17 | 0.10 | 0.29 | 0.10 | 0.071 | 0.18 | 0.071 | 0.055 |
| 2.0 | $\infty$ | 0.13 | 0.08 | 0.86 | 0.08 | 0.063 | 0.50 | 0.063 | 0.050 |
| 3.0 |  | 0.08 | 0.06 | $\infty$ | 0.063 | 0.050 | 3.17 | 0.050 | 0.043 |
| 4.0 |  | 0.06 | 0.05 |  | 0.05 | 0.042 | $\infty$ | 0.041 | 0.035 |



TABLE III
Variances of YHT and Ymr when $X$ is Gamma Distributed, $P(x)$ is Polynominal, $k=0.5, g=2$. (Each entry should be multiplied by $\mu_{y}^{2} / n$ )

| m | $\mathrm{r}=1$ |  | $r=2$ |  | $r=3$ |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\hat{\mu}_{Y H T}$ | $\hat{M}_{Y m r}$ | $\hat{\mu} \mathrm{YHT}$ | $\hat{H}_{Y m r}$ | $\hat{H}_{Y H T}$ | $\hat{\mu}_{Y m r}$ |
| - 1.0 | $\infty$ | 0.25 | 6.50 | 0.25 | 3.17 | 0.25 |
| -0.5 | 6.36 | 0.25 | 2.21 | 0.25 | 1.41 | 0.25 |
| 0.0 | 1.500 | 0.25 | 0.88 | 0.25 | 0.67 | 0.25 |
| 0.5 | 0.47 | 0.25 | 0.38 | 0.25 | 0.34 | 0.25 |
| 1.0 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 |
| 1.5 | 0.47 | 0.25 | 0.38 | 0.25 | 0.34 | 0.25 |
| 2.0 | 1.50 | 0.25 | 0.86 | 0.25 | 0.67 | 0.25 |
| 3.0 | $\theta$ | 0.25 | 6.25 | 0.25 | 3.17 | 0.25 |
| 4.0 | $\infty$ | 0.25 | $\infty$ | 0.25 | 24.00 | 0.25 |



Variances of the Estimators when $X$ is Gamma Distributed, $P(x)$ is Polynominal, $k=1, g=0$. (Each entry should be multiplied by $\mu_{y}^{2} / n$ )

| m | $r=1$ |  |  | $r=2$ |  |  | $r=3$ |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\hat{\mu} \mathrm{YHIT}$ | $\hat{M Y M r}^{\prime}$ | $\hat{\mu} Y \mathrm{R}$ | $\hat{M} Y \mathrm{HT}$ | $\hat{\mu}_{Y m r}$ | $\hat{\mu}_{Y R}$ | $\hat{\mu}_{Y H T}$ | $\hat{\mu}_{\mathrm{Ymr}}$ | $\hat{\mu}_{Y R}$ |
| -1.0. |  |  | $\infty$ | 5.50 |  | 1.00 | 2.50 | $\infty$ | 0.25 |
| -0.5 | 6.63 |  | 4.00 | 1.86 |  | 0.43 | 1.08 | 1.32 | 0.16 |
| 0.0 | 1.00 |  | 1.00 | 0.75 | $\infty$ | 0.25 | 0.44 | 0.52 | 0.11 |
| 0.5 | 1.75 |  | 0.44 | 0.40 | 1.33 | 0.16 | 0.19 | 0.27 | 0.08 |
| 1.0 | $\infty$ | $\infty$ | 0.25 | 0.50 | 0.50 | 0.11 | 0.17 | 0.17 | 0.06 |
| 1.5 | $\infty$ | 1.33 | 0.16 | 1.50 | 0.27 | 0.082 | 0.36 | 0.12 | 0.048 |
| 2.0 |  | 0.50 | 0.11 | $\infty$ | 0.17 | 0.063 | 1.00 | 0.084 | 0.04 |
| 3.00 |  | 0.17 | 0.063 | $\infty$ | 0.083 | 0.04 | $\infty$ | 0.052 | 0.028 |
| 4.00 |  | 0.083 | 0.040 |  | 0.050 | 0.028 |  | 0.033 | 0.020 |



Variances of YHT and Ymr when $X$ is Gamma Distributed, $P(x)$, is Polynominal, $k=0.5, g=2$. (Each entry should be multiplied by $\mu_{Y}^{2} / n$ )

| m | $r=1$ |  |  | $r=2$ |  |  | $r=3$ |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\hat{\mu_{Y H T}}$ | $\hat{H}_{Y r n r}$ | $\hat{\mu}_{Y R}$ | $\hat{H}_{Y H T}$ | $\hat{H}_{Y m \mathrm{~m}}$ | $\hat{H}_{Y R}$ | $\hat{H}_{Y H I}$ | $\hat{\mu}_{Y \mathrm{Ymr}}$ | $\hat{H}_{Y R}$ |
| $-1.0$ |  |  |  | 6.50 | $\infty$ | 1.00 | 3.00 | 1.00 | 0.50 |
| -0.5 | 7.44 |  | 2.00 | 2.32 | 2.00 | 0.67 | 1.36 | 0.67 | 0.40 |
| 0.0 | 2.00 | $\infty$ | 1.00 | 1.00 | 1.00 | 0.50 | 0.67 | 0.50 | 0.33 |
| 0.5 | 0.96 | 2.00 | 0.67 | 0.55 | 0.67 | 0.40 | 0.37 | 0.40 | 0.29 |
| 1.0 | 1.00 | 1.00 | 0.50 | 0.50 | 0.50 | 0.33 | 0.33 | 0.33 | 0.25 |
| 1.5 | 2.53 | 0.67 | 0.40 | 0.84 | 0.40 | 0.29 | 0.50 | 0.29 | 0.22 |
| 2.0 | $\infty$ | 0.50 | 0.33 | 2.00 | 0.33 | 0.25 | 1.00 | 0.25 | 0.20 |
| 3.0 |  | 0.33 | 0.25 | $\infty$ | 0.25 | 0.20 | 5.67 | 0.20 | 0.17 |
| 4.0 |  | 0.25 | 0.20 |  | 0.20 | 0.17 | $\infty$ | 0.17 | 0.14 |



Variances of $Y H T$ and $Y m r$ when $X$ is Gamma Distributed, $P(x)$ is Polynominal $k=1, g=2$. (Each entry should be multiplied by $\mu_{Y}^{2} / n$ )

| m | $r=1$ |  | $r=2$ |  | $r=3$ |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\hat{\mu Y H I}$ | $\hat{H}_{\mathrm{Mmr}}$ | $\hat{\mu} \mathrm{YHT}$ | $\hat{\mu} Y_{\operatorname{mar}}$ | $\hat{\mu}_{Y H I}$ | $\hat{\mu}_{Y m r}$ |
| $-1.0$ |  | 1 | 11.00 | 1 | 5.67 | 1 |
| -0.5 | 10.78 | 1 | 4.15 | 1 | 2.86 | 1 |
| 0.0 | 3.00 | 1 | 2.00 | 1 | 1.67 | 1 |
| 0.5 | 1.36 | 1 | 1.21 | 1 | 1.17 | 1 |
| 1.0 | 1.00 | 1 | 1.00 | 1 | 1.00 | 1 |
| 1.5 | 1.36 | 1 | 1.21 | 1 | 1.17 | 1 |
| 2.0 | 3.0 | 1 | 2.0 | 1 | 1.67 | 1 |
| 3.0 |  | 1 | 11.00 | 1 | 5.67 | 1 |
| 4.0 |  |  |  |  | 39.00 | * |
|  |  |  |  |  |  |  |


estimator and of the mean-of-the-ratios estimator are not comparable as the values of $m$ for which these variances are finite are not the same. In the range of values of $m$ for which either the variance of the HT estinator or the variance of the mean-of-the-ratios estimator and that of the ratio estimator are finite, the ratio estimator is more efficient. When $g=0, r=2$, and $0<m<1$ the ratio estimator has smallest variance followed. by the HT estimator and then by the mean-of-the-ratios estimator, but the mean-of-the-ratios estimator beats the HT estimator for $m>1$. When $g=0, r=3$ and $m<0.8$, the ratio estimator is best followed by the HT estimator and the mean-of-the-ratios estimator is last. The ratio and mean-of-therratios estimators become more efficient with an increase in the value of both $m$ and $r$. The $H i$ estimator's variance takes its minimum value for values of $m$ near 0.8 ; it also takes smaller values as $r$ increases.

When $\mathrm{g}=1$ the ratio estimator always has the smallest variance. When $r=1$ and $m<1$ the $H i$ estimator is better than the mean-of-the-ratios estimator. The reverse is true when m>l. When $r=2$ and for either m<0 or $m>l$, the $H T$ estimator has larger variance than the mean-of-the-ratios estimator, for $0<m<1$ the $H T$ estimator's variance is smaller than that of the mean-of-the-ratios. When $r=3$ the $H I$ estimator beats the mean-of-the-ratios for $0.3<m<l$ and the mean-of-the-ratios estinator beats the HI estimator for other values of m .

When $g=2$ the variance of the mean-of-the-ratios estimator is a constant at $K_{Y}^{2} / n$ and, except when the variance of the HT estimator takes its minimum value, it is snaller than that of the HT estimator.

As in the other case, an increase in the value of $g$ has a general effect of lessening the power of reducing the variances of the ratio and mean-of-the-ratios estimators by increasing $m$. No improvement can be made on the variance of the mean-of-the-ratios estimator when $g=2$. The variance of the HT estimator is smallest for values of $m$ near l.0.

The change in the value of $k$ from 0.5 to 1.0 has the effect of slightly increasing the values of the variances of the estimators for fixed values of $g, r$ and $m$.

## Results for $P(x)$ Exponential and $k=0.5$

The results are given in Tables VII-IX and illustrated by Figures 7-9.

When $\mathrm{g}=0$ or l , the ratio estimator is most efficient followed by the mean-of-the-ratios estimator and then by the HT estimator. The variances of the ratio and mean-of-the-ratios estimators are zero when $c=1$. The greater the value $r$ takes the more efficient the estimators become. The graph of the variance of the HT estimator is u-shaped; an increase in $r$ has the effect of pushing the minimum point of the variance of the HT estimator to the left.

When $g=2$, the variance of the mean-of-the-ratios estimator is constant at $0.25 \mathrm{M}_{\mathrm{Y}}^{2} / \mathrm{n}$ and is smaller than that of the HT estimator. The effect of increasing $r$ on the shape of the variance of the HT estimator's curve is simply to shift its minimum point to the left. The rate of improvement on the estimators by increasing c is reduced as $g$ changes from 0 to $l$, and the estimators become less efficient too.

## TABLE VII

Variances of the Estimators when $X$ is Gamma Distributed, $P(x)$ is Exponential, $k=0.5, g=0$. (Each entry should be multiplied by $\mu_{Y}^{2} / n$ )

| $c$ | $r=1$ |  | $r=2$ |  | $I=3$ |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\hat{\mu} \mathrm{YHT}$ | $\hat{\mu}_{Y R}$ | $\hat{H Y H T}$ | $\hat{H} Y R$ | $\hat{\mu}_{Y H T}$ | $\hat{\mu}_{Y \mathrm{Yr}}$ | $\hat{\mu}_{Y R}$ |
| -1.0 | $\infty$ | 1.00 | $\infty$ | 0.25 | $\infty$ | 0.50 | 0.11 |
| -0.8 | 138.58 | 0.81 | 288.50 | 0.20 | 713.50 | 0.40 | 0.09 |
| -0.6 | 18.90 | 0.64 | 21.25 | 0.16 | 29.84 | 0.32 | 0.07 |
| -0.4 | 6.00 | 0.49 | 4.96 | 0.12 | 4.70 | 0.24 | 0.05 |
| -0.2 | 2.50 | 0.36 | 1.61 | 0.09 | 1.38 | 0.18 | 0.04 |
| 0.0 | 1.25 | 0.25 | 0.56 | 0.06 | 0.36 | 0.12 | 0.03 |
| 0.2 | 0.70 | 0.16 | 0.20 | 0.04 | 0.10 | 0.08 | 0.02 |
| 0.4 | 0.52 | 0.09 | 0.21 | 0.02 | 0.24 | 0.04 | 0.01 |
| 0.6 | 0.60 | 0.04 | 0.58 | 0.01 | 1.21 | 0.02 | 0.004 |
| 0.8 | 1.44 | 0.01 | 2.53 | 0.003 | 12.68 | 0.005 | 0.001 |
| 1.0 | $\infty$ | 0.00 | $\infty$ | 0.00 | $\infty$ | 0.00 | 0.00 |



Variances of the Estimators when $X$ is Gamma Distributed, $P(x)$ is Exponential, $k=0.05, g=1$. (Each entry should be multiplied byk $/ \mathrm{n}$ )

| c | $r=1$ |  | $r=2$ |  |  | $x=3$ |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\hat{H}_{Y H T}$ | $\hat{M}_{Y R}$ | $\hat{H}_{Y H T}$ | $\hat{\mu}_{Y \mathrm{mr}}$ | $\hat{\mu}_{Y R}$ | $\hat{H}_{Y H T}$ | $\hat{M}_{Y m r}$ | $\hat{H}_{Y R}$ |
| -1.0 | $\infty$ | 0.50 | $\infty$ | 0.50 | 0.25 | $\infty$ | 0.25 | 0.17 |
| -0.8 | 141.30 | 0.45 | 290.30 | 0.45 | 0.23 | 713.50 | 0.23 | 0.15 |
| -0.6 | 19.51 | 0.40 | 22.20 | 0.40 | 0.20 | 30.93 | 0.20 | 0.13 |
| -0.4 | 6.11 | 0.35 | 5.08 | 0.35 | 0.18 | 5.33 | 0.18 | 0.12 |
| -0.2 | 2.58 | 0.30 | 1.60 | 0.30 | 0.15 | 1.47 | 0.15 | 0.10 |
| 0.0 | 1.25 | 0.25 | 0.63 | 0.25 | 0.13 | 0.42 | 0.13 | 0.083 |
| 0.2 | 0.66 | 0.20 | 0.27 | 0.20 | 0.10 | 0.16 | 0.10 | 0.067 |
| 0.4 | 0.43 | 0.15 | 0.25 | 0.15 | 0.075 | 0.39 | 0.08 | 0.050 |
| 0.6 | 0.47 | 0.10 | 0.73 | 0.10 | 0.050 | 1.50 | 0.05 | 0.033 |
| 0.5 | 1.10 | 0.05 | 3.54 | 0.05 | 0.025 | 11.15 | 0.03 | 0.017 |
| 1.0 | $\infty$ | 0.00 | $\infty$ | 0.00 | 0.00 | $\infty$ | 0.00 | 0.00 |

$\operatorname{Var}(\hat{M} Y m r)$ is not defined for $x=1$.


TABLE IX
Var $\left(\hat{H}(Y H T)\right.$ and $\operatorname{var}\left(\hat{\mu}_{Y m r}\right)$ when $X$ is Gamma Distributed, $P(x)$
is Exponential $k=0.5, g=2$. (Each entry should be
multiplied bykg/n)



Besults for $P(x)$ Exponential and $k=1$.
The results are given in Tables X-XII and illustrated by Figures 10-12.

The ratio estimator has the smallest variance throughout. When $g=0, r=3$ and $0.28<c$ or $c<-0.1$, the variance of the mean-of-the-ratios estimator is smaller than that of the HT estimator. The HT estimator beats the mean-of-the-ratios estimator for $0.1<c<0.28$.

When $g=1, r=2$ and $c<0.12$ or $c>0.3$, the mean-of-the-ratios estimator is more efficient than the HT estimator, which beats the mean-of-the-ratios estimator for $0.12<c<0.3$. When $r=3$ the mean-of-the-ratios estimator has smaller variance than the HT estimator.

When $g=2$ the variance of the mean-of-the-ratios estimator is constant at $\mu_{Y}^{2} / n$ and is smaller than that of the HT estimator. In this particular case the minimum point of the graph of the variance of HT estimator moves to the left and the value of the variance at its minimurn increases as $r$ increases. An increase in the values of $g$ has effects similar to those mentioned in the other cases above. As with the polynominal case, an increase in the value of $k$ from 0.5 to $l$ has the effect of making the estimators less efficient for fixed values of $c, r$ and $g$.

TABLE X
Variances of the Estimators when $X$ is Gamma Distributed, $P(x)$ is Exponential, $k=1, g=0$. (Each entry should be multiplied by $M_{Y}^{2 / n}$ )

| C | $\mathrm{r}=1$ |  | $r=2$ |  | $r=3$ |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\hat{\mu}_{Y H T}$ | $\hat{\mu}_{Y R}$ | $\hat{\mu}_{\mathrm{YHT}}$ | $\hat{\mu}_{Y R}$ | $\hat{\mu}_{Y H T}$ | $\hat{\mu}_{Y m r}$ | $\hat{M}_{Y R}$ |
| $-1.0$ | $\infty$ | 4.00 | $\infty$ | 1.00 | $\infty$ | 2.00 | 0.47 |
| $-0.8$ | 140.7 | 3.24 | 288.70 | 0.81 | 713.4 | 1.62 | 0.36 |
| -0.6 | 20.10 | 2.56 | 22.13 | 0.64 | 30.80 | 1.28 | 0.28 |
| -0.4 | 5.61 | 1.96 | 5.13 | 0.49 | 5.25 | 0.98 | 0.22 |
| -0.2 | 3.30 | 1.44 | 1.77 | 0.36 | 1.41 | 0.72 | 0.16 |
| 0.0 | 2.00 | 1.00 | 0.75 | 0.25 | 0.31 | 0.50 | 0.11 |
| 0.2 | 1.50 | 0.64 | 0.45 | 0.16 | 0.23 | 0.32 | 0.07 |
| 0.4 | 1.44 | 0.36 | 0.57 | 0.09 | 0.52 | 0.18 | 0.04 |
| 0.6 | 1.77 | 0.16 | 1.41 | 0.04 | 2.07 | 0.08 | 0.018 |
| 0.8 | 3.53 | 0.04 | 6.29 | 0.01 | 15.58 | 0.02 | 0.0044 |
| 1.0 | $\infty$ | 0.00 | $\infty$ | 0.00 | $\infty$ | 0.00 | 0.00 |
| $\operatorname{Var}(\hat{\mathcal{M}} \mathrm{Y} \mathrm{mr}$ ) is not defined for $\mathrm{k}=1, \mathrm{~g}=0, \mathrm{r}<2$. |  |  |  |  |  |  |  |



Variances of the Estimators when $X$ is Gamma Distributed, $P(x)$ is Exponential, $k=1, g=1$. (Each entry should be multiplied by $\mu_{Y}^{2} / n$ )

|  | $r=1$ |  | $r=2$ |  |  | $r=3$ |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $c$ | $\hat{\mu}_{Y H T}$ | $\hat{\mu}_{Y R}$ | $\hat{\mu}_{Y H T}$ | $\hat{\mu}_{Y m r}$ | $\hat{\mu}_{Y R}$ | $\hat{\mu}_{Y H T}$ | $\hat{\mu}_{Y m r}$ | $\hat{\mu}_{Y R}$ |
| -1.0 | $\infty$ | 2.00 | $\infty$ | 2.00 | 1.00 | $\infty$ | 1.00 | 0.67 |
| -0.8 | 151.8 | 1.80 | 292.20 | 1.80 | 0.90 | 714.80 | 0.90 | 0.60 |
| -0.6 | 22.44 | 1.60 | 59.55 | 1.60 | 0.80 | 31.30 | 0.80 | 0.53 |
| -0.4 | 7.60 | 1.40 | 5.61 | 1.40 | 0.70 | 5.59 | 0.70 | 0.47 |
| -0.2 | 3.56 | 1.20 | 2.25 | 1.20 | 0.60 | 1.70 | 0.60 | 0.40 |
| 0.0 | 2.00 | 1.00 | 2.00 | 1.00 | 0.50 | 0.67 | 0.50 | 0.33 |
| 0.2 | 1.31 | 0.80 | 0.67 | 0.80 | 0.40 | 0.50 | 0.40 | 0.27 |
| 0.4 | 1.07 | 0.60 | 0.78 | 0.60 | 0.30 | 0.98 | 0.30 | 0.20 |
| 0.6 | 1.20 | 0.40 | 1.64 | 0.40 | 0.20 | 3.03 | 0.20 | 0.13 |
| 0.8 | 2.26 | 0.20 | 6.43 | 0.20 | 0.10 | 20.79 | 0.10 | 0.07 |
| 1.0 | $\infty$ | 0.00 | $\infty$ | 0.00 | 0.00 | $\infty$ | 0.00 | 0.00 |



TABLE XII
$\operatorname{Var}\left(\hat{H}_{Y H T}\right)$ and $\operatorname{var}\left(\hat{H}_{Y m r}\right)$ when $X$ is Gamma Distributed, $P(x)$ is Exponential, $k=1, g=2$. (Each entry should be multiplied by $\left(\sum_{Y}^{2} / n\right)$

| c | $r=1$ |  | $r=2$ |  | $r=3$ |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\hat{\mu} \mathrm{YHT}$ | $\hat{\mu}_{\mathrm{Ymr}}$ | $\hat{\mu}_{\text {YHT }}$ | $\hat{\mu}_{Y} \mathrm{mr}$ | $\hat{\mu}_{Y H T}$ | $\hat{\mu} \mathrm{Ymr}$ |
| - 1.0 | $\infty$ | 1 | $\infty$ | 1 | $\infty$ | 1 |
| -0.8 | 276.78 | 1 | 346.30 | 1 | 742.0 | 1 |
| - 0.6 | 37.46 | 1 | 31.04 | 1 | 37.87 | 1 |
| - 0.4 | 12.16 | 1 | 8.95 | 1 | 7.50 | 1 |
| - 0.2 | 5.51 | 1 | 3.58 | 1 | 2.86 | 1 |
| 0.0 | 3.00 | 1 | 2.00 | 1 | 1.67 | 1 |
| 0.2 | 1.90 | 1 | 1.49 | 1 | 1.55 | 1 |
| 0.4 | 1.43 | 1 | 1.60 | 1 | 2.40 | 1 |
| 0.6 | 1.45 | 1 | 2.72 | 1 | 6.08 | 1 |
| 0.8 | 2.51 | 1 | 9.00 | 1 | 36.41 | 1 |
| 1.0 | ; $\infty$ | 1 | $\infty$ | 1 |  | 1 |



## Summary of Results and Some Conclusions

The following observations can be made from the results:
(1) Under the given assumptions about the distribution of the auxiliary variable, $X$, and the sampling designs, and for $g=0$ or 1 , the ratio estimator is better than both the mean-of-the-ratios and the HT estimators. Under a very wide range of the values of $m$, the mean-of-the-ratios estimator performs better than the HT estimator.
(2) The HT estimator is very sensitive to changes in values of the design function parameters $m$ and $c$. Within a certain (small) range of the values of these parameters the HT estimator has, sometimes, smaller variance than that of the mean-of-the-ratios estimator.
(3) The variances of these estimators are usually smaller when $k=0.5$ than when $k=1$.
(4) For fixed values of $k$ and $g$, the variances of the ratio and mean-of-the-ratios estimators usually increase with a decrease in the value of $r$ 。 With the HT estimator, it depends on the range of values of $m$ or considered and the design function used. With the polynominal design function, the efficiency of the HT estimator usually increases with $r$. With the exponential design function and for, say, c<0.3, the
efficiency of the HT estimator improves with an increase in the value of $r$; when $c>0.4$ the $H T$ estimator usually becomes less efficient as $r$ takes higher values.
(5) For the same values of $k$, and for the same ranges of the values of the design function parameters, $m$ or $c$, the effect of increasing $g$ is usually to make the ratio and the mean-of-the-ratios estimators less efficient.
(6) Other things fixed, the variances of the ratio and the mean-of-the-ratios estimators decrease with an increase in the values of $m$ or $c$.
(7) The ratio and mean-of-the-ratios estimators have smaller variances when selection is 'purposive' ( $m>0$ or $c>0$ ) than when simple randorn sampling ( $\mathrm{m}=\mathrm{c}=0$ ) is employed.
(8) The HT estimator is most efficient when the sampling scheme is approximately pps, i.e. when $m$ is approximately $=1$ (in the polynominal case).
(9) In most cases, the variances of these estimators can be equalized by simply choosing different combinations of the parameters. For example, the variance of the ratio estimator for $k=1, g=1$ and $r=1$ equals the variance of the mean-of-the-ratios estimator with $k=1, g=1$ and $r=2($ and $P(x)$ is polynominal).
(10) For a greater range of the values of $c$ or $m$, the estimators have smaller variances when $P(x)$ is polynominal than when it is exponential.
(11) With an exponential design function, the variances of the estimators are finite for a very small range of the values of $c$. The variances of the ratio and mean-of-the-ratios estimators are zero when $c=0$.

## CHAPTER 4

## DISCUSSION OF RESULTS

## Some Empirical Comparisons

P.S.R.S. Rao (1969) compared four ratio-type
estimators under the regression model:

$$
\begin{aligned}
& y=\alpha+\beta X+e \text { where } \\
& E\left(e_{i} \mid x_{i}\right)=0 \\
& E\left(e_{i} e_{j} \mid \dot{x}_{i} x_{j}\right)=0 \\
& \operatorname{Var}\left(e_{i} \mid x_{i}\right)=\delta x_{i}^{g}
\end{aligned}
$$

and X is gamma distributed with parameter $h$. He found out that when $\alpha=0$ and $g \geq 1$, the MSE's increase as $h$ increases. He considered values of $h>2$ and samples of size 2-10. These results are not quite compatible with my results for the same MSE's of the ratio and mean-of-the-ratios estimators (Rao's first estimator is the ratio estimator). But from (2.5) and (2.9), it is clear that, for $g>2$, the variance of the mean-of-the-ratios estimator increases with ro For g < 2, the variance decreases with r. Since I did not work out the variance of the ratio estimator for $g>1$, I do not have much to compare with. Certainly when $g=1$, the variance of the ratio estimator decreases as $r$ increases. Rao also found out that for fixed $n$ and $h$ (or $r$ ) the MSE's increase as $g$ increases. This agrees with my results and can be inferred from (2.5) (2.7). Lastly, his results show that for $\alpha=0, n>2$ and
$g=1$ or 2 , the ratio estimator has smaller HE than the other three. It is implied from his paper that a choice of $\alpha$ and g.combinations does affect the :ASE's of the estimators.

Wy results show that the variances of the ratio and mean-of-the-ratios estimators become smaller as the design function parameter $m$ or $c$ increases. (2.5) (2.7), (2.10) - (2.11) show similar results. But the larger the values of $m$ or $c$ the higher will be the inclusion probabilities for units with large values of $X$ compared with those with small values of $X$. This leads to the notion that if units with large values of $X$ are purposely included in the sample, the estimation procedure will be more efficient. Many researchers, notably, Royall (1970, 1971) have found similar results under similar conditions. Royall's results also show that the ratio estimator is better when combined with designs other than simple random sampling. For $g=1, k=1$, he shows that the ratio estimator is the best linear $\epsilon$ - unbiased estimator. When used with simple random sampling design, it remains optimal only if N is infinite. Since Godambe proved the non-existence of a uniformly minimum variance unbiased estinator among a class of all unbiased linear estimators for any sampling design, it may be proper to note that while discussing these optimality properties of the ratio estimator we are, most of the time, restricting
ourselves to a certain class of linear estimators only (Godambe's 1955 class (iii)). The same remarks apply to optimality conditions given by Cochran (1963). The HT estimator belongs to Godambe's (1955) sub-class (i) of estimators. It is the best and the only unbiased linear estimator in the sub-class.

Under super-population set up, Rao (1967) considered the problem of choosing a suitable strategy for the ratio, mean-of-the-ratios and the HT estimators. His results suggest that the mean-of-the-ratios estimator with MPs is better than the other two estimators used with the Midzumo-Sen sampling scheme. My results suggest that if the ratio and mean-of-the-ratios estimators are combined with simple random sampling, there always exists an $m$ or $c$ such that the variance of the mean-of-the-ratios estimator is less than the variances of the ratio and of the HT estimators. Similarly, the variance of the ratio estimator with varying probability of inclusion procedure can be made small compared with those of the HT and mean-of-the-ratios estimators with simple random sampling.

If in formula (2.2) for the mean-of-the-ratios estimator, we let

$$
\frac{x_{i}}{\mathrm{r}}=P\left(X_{i}\right)
$$

the HT estinator (2.1) is obtained. Hanurav (1967) was
interested in finding sampling designs under which this $P\left(X_{i}\right)=\frac{X_{i}}{x}$ and the variance of the mean-of-the-ratios estimator is unbiased (he was estimating the population total). For $n=2$, he gives two sequential sampling procedures that solve the problem. With my study, the conditions are easy to find:
when $P(X)$ is polynominal, we want

$$
P\left(X_{i}\right)=\frac{\Gamma(r) X_{i}^{m}}{\left(m^{m}\right)}=\frac{X_{i}}{r}
$$

when $P(X)$ is exponential we have

$$
P\left(X_{i}\right)=(1-c)^{r} e^{c x_{j}}=\frac{X_{i}}{r}
$$

It is easy to find an $r$ ( $c$ or $m$ ) that will give the required $P\left(X_{i}\right)$ for fixed $m$ or $c$ (or $r$ ) provided $c=m=0$. With polynominal $P(X)$, this study is really a special case of that considered by Cassel and Sarndal (1973). Nearly all of my findings are in agreement with their findings. They find out, for example, that:
(i) When $g=2$, the mean-of-the-ratios estimator has constant variance regardless of both the design and the distribution function, $f(x)$.
(ii) The ratio estimator combined with simple random sampling can be quite inefficient compared with it if used with other sampling designs.
(iii) Under certain conditions, one should use purposive selection of the units with largest $X$-values.
(iv) When $g \leq 1$ the ratio estimator is at least as efficient as the mean-of-the-ratios estimator for any design $\mathrm{P}(\mathrm{x})$. In the case that J. am considering, for $\mathrm{n}>\mathrm{l}$ and $\mathrm{g} \leq 1$, the expressions for the variances of the ratio and the mean-of-the-ratios estimators show that the ratio estimator has smaller variance than the mean-of-the-ratios estimator, and the difference between the two variances increases with $n$. For example, when $g=0$ and $P(x)$ is polynominal, we have

$$
\begin{aligned}
& \operatorname{Var}\left(\hat{\mu}_{\mathrm{Ymr}}\right)=\frac{n k^{2} \hat{\mu}_{Y}^{2}}{n^{2}\left(m^{+} r-1\right)(m+r-2)} \\
& \operatorname{Var}\left(\hat{H}_{\mathrm{YR}}\right)=\frac{n k^{2} \hat{\mu}^{2}}{\left(n m^{+} n r-1\right)\left(n m^{+} n r-2\right.}
\end{aligned}
$$

(v) Variances of the ratio and mean-of-the-ratios estimators increase with $m$ or $c$.
(vi) If $g=2$, regardless of the design function used, the mean-of-the-ratios estimator has variance that is smaller than, or equal to that of the ratio estimator (I did not work out the variance of the ratio estimator when $g=2$ ).
(vii) If $g>2$ the variance of the mean-of-the-ratios estimator becomes small if the design assigns the bulk of the selection probabilities to the units with smallest X-values.

## (viii) The HT estimator is generally highly sensitive to shifts in the design. Its variance is a minimum when selection probabilities are somewhat in the vicinity of pps procedure. The variance can become very large due to minor deviations from the point of minimum.

## Comments and Some Implications

From the results of this study, it would seem that instead of trying to look for estimators that are generally optimal like uniformly minimum variance estimators, more effort should be used in defining clearly and simply the conditions under which the popular estimators are efficient.

Under the usual regression models, the choice of $g$, it seems, determines the extent to which the best choice of sampling strategy can improve the estimation process. In most. cases studied, it has been found that the value of $g$ lies between 1 and 2 . This may be unfortunate as the three common estimators I have considered can be more efficient when $g=0$. On the other hand, $g=0$ implies that the variance of the error term, $Z_{i}$, is constant which is a very unlikely situation in practice. For $1<g<2$, the variances of the three estimators can be made to attain their minimur values by a proper choice of fi, $r$ and $g$. Perhaps the good thing with the continuous
variable model is that the expressions for the variances of the estimators are very, very simple. If one was interested, in getting the exact minimum values of the variances of these estimators, it should not be too difficult for him to do so. He will likely have to use the computer and some mathematical programming techniques.

The results also call for more attention to the choice of estimators and sampling designs when doing survey sampling. In particular, the results show again, that in most cases, simple random sampling is not an optimal sampling design. There are other better ones. And when estimating the population mean, the sample average need not give optimal results. There may be other better estimators.

In practice, the sampler, under this model, will partly be able to control the design function parameters. In such situations, studies like this may help the samplex make a proper choice of the design function parameter that will give best results. In this study, the ratio and rhean-of-the-ratios estimators, as in similar other studies, promise good results when selection is purposive under certain conditions. Most sample survey experts object to this method of selection because, as Hansen, Hurwitz and Madow (1953, p.9) put it:
(a) hethods of selecting samples based on the theory of probability are the only general methods known to us which can provide a measure of precision.

Only by using probability methods can objective numerical statements be made concerning the precision of the results of the survey;
(b) It is necessary to be sure that the conditions imposed by the use of probability methods are satisfied. It is not enough to hope or expect that they are. Steps must be taken to meet these conditions by selecting methods that are tested and are demonstrated to conform to the probability model.

They continue saying:
We assert that, with rare exceptions, the precision of estimates not based on known probabilities of selecting the samples cannot be predicted before the survey is made, nor can the probabilities or precision be estimated after the sample is obtained. If we know nothing of the precision, then we do not know whether to have much faith in our estimates, even though highly accurate measurements are made on the units in the sample.

Random sampling is usually supported for similar arguments, namely, it protects against failure of certain probabilistic assumptions, it averages out effects of unobserved or unknown random variables, it guards against unconscious bias on the part of the experimenter, it will usually produce a sample in which the $X^{\prime}$ s are spread throughout the range of $X$ values in the population and this enables the sampler to check the accuracy of assumptions concerning the relation of the $Y^{\prime}$ s to the $X ' s$ and, again, it enables the sampler to estimate, from the sample, the precision of his estimate.

But probability methods can do nothing more than give us expectations about, say, the possjole precision of
the results of the survey. The precision would usually be stated in terms of the probability that the estimate deviates from the real value, and as long as it is given in these probability terms, it does nothing more than give expectations, however high and refined the probabilities may be. On the other hand, if many studies point to the fact that non-probability methods, like purposive selection, lead, under certain conditions, to efficient estimations, the same probability theory may allow that under sjmilar sampling conditions, we can expect, with high probability, to obtain similar good results. As Royall (1970) argues, if the sampler believes it to be important that he obtain. a sample in which the $X$ values have a certain configuration, then he should choose a sample deliberately and not leave it to the choice of a certain chance mechanism.

In this study, the ratio estimator has once again shown it superiority over the mean-of-the-ratios and the HT estimators. The HT estimator has revealed its sensitivity to the choice of parameters of the model and of the sampling design. The mean-of-the-ratios estimator may not be very far off from the ratio estimator.

The results of the study are quite similar to similar studies under different regression type model.s. I would like to note that stratified sampling combined with simple random sampling in each stratum can be achieved, under this model,
by assigning the same $P(x)$ for all members of one straturn and different $P(x)$ for members of different strata.

## Sone Limitations

In putting the ideas contained in this study into practice, the order of events is (1) estimate $F(x)$, the distribution function of $x(2)$ approximate $g$ and $\theta$ and (3) investigate sone sampling designs and estimators and choose the ones that give best results. This study would help the sampler to approximately exarnine the behaviours of different designs and estimators he may be pondering to use.

The assumption concerning an infinite population that has approxinately a continuous frequency distribution, while it helps simplify the investigation of different designs, also makes the situation considered an idealization of the real situation. To estimate the frequency function of $x$, one could start by observing the histogram of the $x$ values and choose or fit an approximate continuous function, possibly by some mathernatical curve fitting techniques, that closely resembles the histograra; and the continuous function thus obtained has to be standardized to becone a cumulative distribution function. To approximate $\theta$ and $g$, we could use some likelihood methods like the ones suggested by Brewer (1963).

It is quite possible that some problems of evaluating the variances of the estimators we may want to investigate given the approximate distribution function of $x$ will be encountered. Should it, for example, turn out that the $x$ values are approximately normally distributed and the sampler wants to investigate the sampling designs and estimators $I$ have studied, it would not be easy to evaluate the variances. But with other distribution functions and sampling designs things should be easy going and studying the properties of such sampling strategies is easy. Cassel and Sarndal (1973) show that results obtained under this model are valid for values of $N$ as low as $N=10$.

## Some Recommendations

I think sample survey theorists should spend more time simplifying and unifying the results of their research. They should spend more time in refining and reducing the number of different sampling designs and estimators that they are considering. They should, somehow, formulate a simple unified theory of sampling that can easily be put into practice. In this connection, the estimators I have considered nay prove useful. Surely, some problems may crop up initially.

I also think that some ideas from General Statistical Theory can be useful in formulating a simple sample survey theory; there is no need of divorcing one of the sampling theories from the other.

## REFERENCES

Basu, D. (1971). An essay on the logical foundations of survey sampling, Part l. in Foundations of Statistical Inference, ed. by V. P. Godambe and D. A. Sprott, Toronto: Holt, Rinehart and Winston of Canada Ltd., 203-242.

Brewer, K.R.W. (1963). Ratio estimation and finite populations: some results deducible from the assumption of an underlying stochastic process. Aust. J. Statist. 5, 93-105.

Cassel, C.M. and Sarndal, C.E. (1972a). A model for studying robustness of estimators and informativeness of labels in sampling with varying probabilities. J. R. Statist. Soc. B., 35, 279-239.
, (l972b). An analytic framework offering some guidelines to the choice of sampling design and the choice of estimator for finite populations. Vorking paper No. 142, Faculty of Commerce, UBC.
, (1973). Evaluation of some sampling strategies for finite populations using a continuous variable framework. To appear in the April issue of Communications in Statistics.

Cochran, W.G. (1946). Relative accuracy of systematic and stratified random samples for a certain class of populations. Ann. Math. Statist., 17, 164-177.
_, (1963). Sampling Techniques. New York: Wiley.
Durbin, J., (1953). Some results in sampling theory when the units are selected with unequal probabilities. J. R. Statis. Soc. B., 15, 262-269.

Foreman, E.K., and Brewer, K.R.W. (1971). The efficient Use of Supplementary Information in Standard Sampling Procedures. J.R. Statist. Soc. B., 33, 391-400.

Godambe, V.P. (1955). A unified theory of sampling from finite populations. J.R. Statist. Soc. B., 17 , 269-278.
, (1965). A review of the contributions towards a unified theory of sampling from finite populations. Rev. Int, Statist. Inst., 32., 242-253.

Godambe, V.P. (1969). Some aspects of the theoretical developments in survey sampling. New Developments in Survey Sampling, ed. by N. L. Johnson and H. Smith, Jr., 27-58. New York: Wiley.
__ (1970). Foundations of Survey-sampling. The American Statistician, 24, 33-38.
, and Joshi, V.M. (1965). Admissibility and Bayes estimation in sampling finite populations,l. Ann. Math. Statist., 36, 1707-1722.
Goodman, R. and Kish, L. (1950). Controlled selection - A technique in probability sampling. J. Amer. Statist. Assoc.. 45, 350-372.
Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953). Sampling Survey Methods and Theory, Vol. 11. John Wiley \& Sons, New York.

Hanurav, T.V. (1967). Optimum utilization of Auxiliary Information: MPS sampling of two units from a stratum. J.R. Statist. Soc. B. 29, 374-391.

Hartley, H.O. and Rao, J.N.K. (1962). Sampling with unequal probabilities and without replacement. Ann. Math. Statist. 2, 77-86.

Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. J. Amer. Statist. Ass. 47, 663-685.

Joshi, V.M. (1969). Admissibility of Estimates of the mean of a finite population. New Developments in Survey Sampling, ed. by N. L. Johnson and H. Smith, Jr., 188-207, New York: Wiley.

Neyman, J. (1934). On the two different aspects of the representative method. The method of stratified sampling and the method of purposive selection. J. R. Statist. Soc. 97., 558-606.

Rao, J.N.K. (1969). Ratio and Regression Estimators, New Developments in Survey Sampling. Ed. by N.L. Johnson and H. Smith, Jr., 213-234, New York, Wiley.

Rao, P.S.R.S. (1969). Comparison of four ratio-type estimates under a model. J. Amer. Statist. Assoc. 64. 574-580.

Rao, T.J. (1967). On the choice of a strategy for the Ratio Method of estimation. J.R. Statist. Soc. B, 29, 392-397.

Royall, R.M. (1970). On finite population sampling theory under certain linear regression models. Biometrika, 57, 377-387.
___ (1971). Linear regression models in finite population sampling theory. In Foundations of Statistical Inference, ed. by V.P. Godambe and D. A. Sprott, Toronto: Holt, Rinehart and Winston of Canada Ltd., 259-279.

Sarndal, C.E. (1972). Sample Survey theory Vs. General Statistical theory: estimation of the population mean. Rev. Int. Statist. Inst., 40, 1-12.
Sukhatme, P.V. and Sukhatme, B.V. (1970). Sampling Theory of Surveys with Applications, Asia Publishing House, Bombay 1 .

