DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

METHODS USED IN REDUCTION OF ERRORS ARISING FROM NONRESPONSE

by
Müge BORAZAN ÇELİKBIÇAK

January, 2006
İZMİR

METHODS USED IN REDUCTION OF ERRORS ARISING FROM NONRESPONSE

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of
Dokuz Eylül University
In Partial Fulfillment of the Requirements for the
Degree of Master of Science in Statistics, Statistics Program

by
Müge BORAZAN ÇELİKBIÇAK

January, 2006
İZMİR

M.Sc THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"METHODS USED IN REDUCTION OF ERRORS ARISING FROM NONRESPONSE"** completed by **Müge BORAZAN ÇELİKBIÇAK** under supervision of **Assist. Prof. Dr. Özlem EGE ORUÇ** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Özlem EGE ORUÇ

Supervisor

Assist. Prof. Dr. Emel KURUOĞLU

(Jury Member)

Assist. Prof. Dr. Hamdi EMEÇ

(Jury Member)

Prof. Dr. Cahit HELVACI

Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENTS

I am very greatly obliged and indepted to **Prof.Dr. Hülya ÇINGI** and **Assist. Prof. Dr. Özlem EGE ORUÇ**, my supervisors, for their valuable guidances, professional advices, constrictive critisms and suggestions during my research.

Special Thanks to **Prof.Dr. Serdar KURT** for his kind help and suggestions throughout my M. Sc. Study.

Meanwhile I want to send my special thanks to my dear friend **Lügen ÜNAL** for her helps on C++ programming language.

Finally, I am very much indepted to my husband **Ömür ÇELİKBIÇAK**, and my parent for their support and encouragement during my M. Sc. Studies

Müge BORAZAN ÇELİKBIÇAK

METHODS USED IN REDUCTION OF ERRORS ARISING FROM NONRESPONSE

ABSTRACT

Various types of survey errors, especially nonresponse errors, may seriously deteriorate data quality. Nonresponse has more one reason to worry about its harmful effects on the survey estimates. So, nonresponse and methods dealing with nonresponse have increasingly become a standard part of survey sampling. A nonresponse occurs in a survey when, for any reason, a selected unit does not respond. The usual methods of estimation in the presence of nonresponse give biased results. Because of this special estimation techniques are required to deal with the problem. *Imputation* and *reweighting* are two standard methods provided by the literature for treating nonresponse. In the recent years, scientist became increased to concern with the calibration approach to reweigthing method in the presence of nonresponse. The calibration approach generates the final weights which are as close as possible to specified design weights, while respecting known auxiliary population totals or unbiased estimates of these totals. This calibration procedure requires the formulation of a suitable auxiliary vector, through a selection from a possible larger set of auxiliary variables.

In this study standard methods for the reduction of bias and errors arising from nonresponse are explained. The calibration approach is examined as theoretically and simulation is performed by macro generated in C++ programming language to study how alternative specifications of the auxiliary vector affect the quality of estimators derived by the calibration technique. In the application, the calibration estimators and quality measures such as relative bias, variance are computed and interpreted for the population total.

Keywords: Nonresponse, nonresponse errors, calibration, auxiliary information, nonresponse adjustment, nonresponse bias, weighting

YANITLAMAMADAN KAYNAKLANAN HATALARIN AZALTILMASINDA KULLANILAN YÖNTEMLER

ÖZ

Anket araştırmalarında karşılaşılan çeşitli problemlerden biri olan yanıtlamama hatası veri kalitesini ciddi şekilde bozmaktadır. Yanıtlamama, anket araştırmalarından elde edilecek tahminler üzerindeki tehlikeli etkilerinden dolayı kaygı duyulan bir durumdur. Bu nedenle yanıtlamama sorununu giderecek yöntemler örneklemenin standart bir bölümünü oluşturmaktadır. Yanıtlamamanın varlığı durumunda ideal koşullar için kullanılan tahmin yöntemleri yanlı sonuçlar vermektedir. Bu sorunun üstesinden gelmek için özel tahmin tekniklerinin kullanılması gerekmektedir. *Yerine Atama (ikame)* ve *Yeniden Ağırlıklandırma* yanıtlamama sorunu ile karşılaşıldığında kullanılan ve literatürde de belirtilen iki standart yöntemdir. Son yıllarda, bilim adamları, yanıtlamamanın varlığında, yeniden ağırlıklandırma yönteminde kalibrasyon yaklaşımını daha sık kullanmaktadırlar. Çünkü yardımcı bilginin etkin kullanımında kalibrasyon uyarlanabilir bir yaklaşımdır. Kalibrasyon yaklaşımı, yardımcı bilgi kitle toplamını veya yansız tahminin bilinmesini gerektirirken, tasarım ağırlıklarına olabildiğince yakın olması gereken son ağırlıkları yaratır. Kalibrasyon yöntemi, olası yardımcı değişken kümesinden seçim ile oluşturulan uygun yardımcı değişken vektörünün formulasyonunu gerektirir.

Bu çalışmada, yanıtlamamadan kaynaklanan yanlılığın ve hataların azaltılabilmesi için kullanılan yöntemler tanıtılmıştır. Bu yöntemlerden biri olan Kalibrasyon yaklaşımı teorik olarak incelenmiştir. Yardımcı değişken vektörünün değişik durumlarının kalibrasyon tekniği ile hesaplanan tahmin edicilerin kalitesini nasıl etkilediğini görebilmek amacıyla C++ programlama dili kullanılarak bir simülasyon çalışması yapılmıştır. Uygulamada kitle toplamı için kalibrasyon tahminleri, göreli yanlılıklar, varyanslar hesaplanarak yorumlanmıştır.

Anahtar Kelimeler : Yanıtlamama, yanıtlamama hatası, kalibrasyon, yardımcı bilgi, yanıtlamama düzeltmesi, yanıtlamama yanlılığı, ağırlıklandırma.

CONTENTS

Page

CHAPTER ONE

INTRODUCTION

The results of a sample survey are affected by many kinds of errors, one of the most important sources being nonresponse. Nonresponse has long been a matter of concern in survey sampling. Nonresponse means failure to obtain a measurement on one or more study variables for one or more elements *k* selected for the survey. It is present almost all surveys, but the extent and effect of the nonresponse can vary greatly from one type of survey to another. The main problem caused by nonresponse is that estimators of population characteristics must be assumed to be biased unless convincing evidence to the contrary is provided. If the nonrespondents are not a random subset of the sample, in other words, if their characteristics differ systematically then the achieved sample will produce biased population estimates. It is for this reason that survey researches are concerned about nonresponse. (Särndal, Swensson and Wretman, 1991)

In the best of surveys, nonresponse occurs, and special estimation techniques are required to deal with the problem. In the recent literature, the problem of nonresponse is viewed from two different but complementary angles: the prevention or avoidance of nonresponse before it has occured, and the special techniques required in estimation when nonresponse has occured. The principal methods for nonresponse adjustment are *reweigthing* and *imputation*. Reweigthing entails altering the weights of the respondents, compared to the weights that would have been used in the case of full response. Imputation entails replacing missing values by proxy values. In the recent years, there has been a great interest in the calibration approach to reweighting which has the favourable property of incorporating most "Standard" methods found in the different places in the literature. The calibration procedure generates final weights which are as close as possible to specified initial (design) weights , while respecting known auxiliary population totals or unbiased estimates of these totals. The calibration approach has only a single computational step and it requires no seperate modelling of a nonresponse mechanism. For these reasons, the calibration approach to reweighting, is better suited for a routine treatment of nonresponse in organization. (Lundström& Särndal, 2001)

This study contains seven chapters. In chapter two, general information about survey and survey errors are given. In chapter three, methods for the estimation under ideal conditions that the survey has without nonresponse are given. These estimators' features and variance estimators are mentioned. In chapter four, definition and sources of nonresponse are given. Common methods in the old literature dealing with nonresponse are investigated. The importance of auxiliary information used for adjusting nonresponse is described. Two standard approaches for the reduction of nonresponse bias and errors are given in chapter five. At first, the imputation method is introduced. In the second, reweighting method with two approach is explained. Especially, calibration approach which is the most powerful technique for the reduction of nonresponse bias is emphasized in details. Chapter six aims to emphasized how alternative specifications of the auxiliary vector $\mathbf{x}_\mathbf{k}$ affect the quality of the estimators derived by the calibration technique. Simulation study is done to measure this quality that relative bias and variance are computed for the estimators.

CHAPTER TWO

GENERAL INFORMATION ABOUT SURVEY ERRORS

The objective of survey is to provide information about unknown characteristics, called parameters, of a finite collection of elements called a population for example, a population of individuals, of households, or of enterprises. A typical survey involves many study variables and produces estimates of different types of parameters, such as the total or the mean of a study variable, or the ratio of totals of two study variables. Sometimes different kinds of elements are measured in the same survey, as when both individuals and households are observed. (Särndal& Lundström, 2001)

A government or some other users express a need for information about a social or economic issue and existing data sources are insufficient to meet this need is the origin of a survey. In the planning process to determine the survey objectives as clearly and unambiguously as possible is the first step. The next step, referred to as survey design, is to develop the methodology for the survey. (Särndal, Swensson and Wretman, 1991)

Survey design involves making decisions on a number of future survey operations. The data collection method must be decided upon, a questionnaire must be designed and pretested, procedures must be decided on, interviewers must be selected and trained, the techniques for handling nonresponse must be decided on, and procedures for tabulation and analysis be thought out.

There are three different types of survey and these are:

- ✓ Survey based on administrative registers,
- ✓ Census Survey,
- ✓ Sample survey.

Kish (1979) explained advantages and disadvantages of these three types of survey.

A survey will usually encounter various technical difficulties. No survey is perfect in all regards. The statistics that result from the survey are not error free. There are five types of error and these are:

- ✓ Sampling error,
- ✓ Nonresponse error,
- ✓ Coverage error,
- ✓ Measurement error,
- ✓ Coding error.

Errors in survey estimates are traditionally divided into two major categories: sampling error and nonsampling error.

The *sampling error* is the error caused by observing a sample instead of the whole population. When statisticians speak about sampling error they mean the error caused by the fact that values of a study variable are recorded only for a sample of elements, not for all elements of the population. The sampling error is subject to sample-to-sample variation. If the whole population were indeed observed, the sampling error would be zero. This situation is exceptional. There could be other errors, for example, measurement error and nonresponse error, but the sampling error would be zero. Statisticians usually measure "error" by a variance. Hence, the sampling error is measured by the variance of the estimator in use, assuming that there are no other errors. (Särndal & Lundström, 2001)

The *nonsampling errors* include all other errors. The two principal categories of nonsampling errors are:

- ✓ *Errors due to nonobservation.* Failure to obtain data from parts of the target population.

- ✓ *Errors in observations.* This kind of error occurs when an element is selected and observed, but the finally recorded value for that element, which the value that goes into the estimation and analysis phase, differs from the true value. Two major types are:

  (l) measurement error (error arising in the data collection phase)

  (2) processing error (error arising in the data processing phase).

There are two principal types of *nonobservation*. These are:

(1) undercoverage, that is, failure of the frame to give access to all elements that belong to the target population, such elements will obviously not be selected, much less observed, and they have zero inclusion probability.

(2) nonresponse, that is, some elements actually selected for the sample turn out to be non-observations because of refusal or incapacity to answer, not being at home, and so on. Nonobservation generally results in biased estimates.

Processing error comprises the errors arising from coding, transcription, imputation, editing, outlier treatment, and other types of preestimation data handling.

It becomes necessary at this point to distinguish target population from frame population. The *target population* is the set of elements about which information is wanted and parameter estimates are required. A survey aims at obtaining information about a target population. The delimitation of the target population must be clearly stated at the planning stage of the survey. The staticians's interest does not lie in publishing information about individual elements of the target population, but in providing descriptives measures (totals or functions of totals) for various domains, that is, for various aggregates of population elements.

The *frame population* is the set of elements that are either listed directly as units in the frame or can be identified through a more complex frame concept, such as frame for selection in several stages. Frame population is the set of all elements that could possibly be drawn that the sample is drawn from. The frame population and the target population are no always identical.

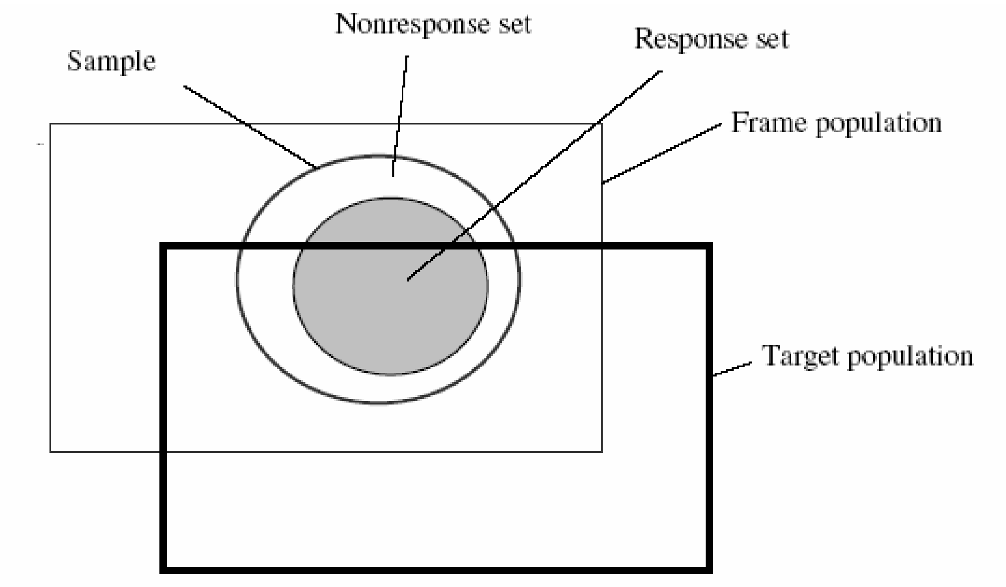

Figure 2.1. Source of errors (Särndal Lundström, 2001)

If a probability sample is selected from the frame population, valid statistical inference can be made about the frame population. If the frame population is different from the target population, valid inference about the target population may be possible, so goal of the survey may be missed. To construct a high quality frame for the target population is an important aspect of survey planning and adaequate resources must be set aside for this activity.

Target population is defined as the set of elements that the survey aims to encompass at the time when the questionnaire is filled in. This point in time is called the *reference time point for the target population*. The sampling frame is usually constructed at an earlier date, sometimes as much as twelve months earlier; this time point is referred to as the *reference time point for the frame population*. The lag between the two time points should be as short as possible, because the risk of coverage errors increases with the time lag. Three types of coverage error are commonly distinguished:

- ✓ undercoverage,
- ✓ overcoverage,
- ✓ duplicate listings.

Elements that are in the target population but not in the frame population constitute *undercoverage*. Especially in business surveys, a significant part of the undercoverage is made up of elements that are new to the target population but are not present in the frame population. These are commonly referred to as “births”. Undercoverage may, of course, also have other causes.

Elements that are in the frame population but not in the target population constitute *overcoverage*. Elements that have ceased to exist somewhere between the two reference time points can be a significant source of overcoverage. These elements are often referred to as “deaths”.

Duplicate listings refer to the type of errors occurring when a target population element is listed more than once in the frame.

It follows that undercoverage elements have zero probability of being selected for any sample drawn from the frame population. This is an undesirable feature, because if the study variable values differ systematically for undercoverage elements and other population elements, there is a risk of biased estimates. Bias from overcoverage can usually be avoided if it is possible to identify the sample elements that belong to the overcoverage. One procedure is to treat these elements as a special domain. However, it is usually impossible to correctly classify all sample elements as belonging either to the target population or to the overcoverage. The problem becomes particularly acute for nonresponding elements, and biased estimates can be the result.

Measurement error can be traced to four principal sources:

- ✓ The interviewer.
- ✓ The respondent.
- ✓ The questionnaire.
- ✓ The mode of the interview, that is, whether telephone, personal interview, self-administered questionnaire, or other medium is used.

A survey consists of a number of survey operations. Especially in a large survey, the operations may extend over a considerable period of time, from the planning stage to the ultimate publication of results. The operations affect the quality of survey estimates. Särndal, Swensson and Wretman (1991) distinguished five phases of survey operations, as follows:

- ✓ *Sample Selection*: Errors in estimates associated with this phase are:

  (1) frame errors, of which undercoverage is particularly serious, and
  (2) sampling error, which arises because a sample, not the whole population, is observed.

- ✓ *Data Collection*: Errors in estimates resulting from this phase include two type of errors:

  (1) measurement errors, for instance, the respondent gives incorrect answers, the interviewer understands or records incorrectly, the interviewer influences the responses, the questionnaire is misinterpreted,
  (2) error due to nonresponse, the survey is designed and conducted carefully, some of the desired data will be missing. The reason for this is that refuse to provide information or contact cannot be established with a selected element. At the result, nonresponding elements may be

systematically different than responding elements, there will be *nonresponse error*.

- ✓ *Data Processing*: Errors in estimates associated with this phase include transcription errors (keying errors), coding errors, errors in imputed values, errors introduced by or not corrected by edit.

- ✓ *Estimation and Analysis*: All errors from the phases (i) to (iii) above will affect the point estimates and they should ideally be accounted for in the calculation of the measures of precision.

- ✓ *Dissemination of Results and Postsurvey Evaluation*

If the operational definitions are clearly stated, the survey statistician can work toward the specification of a suitable survey design, including sampling frame, data collection method, staff required, sample selection, estimation method, and determination of the sample size required the obtain desired precision in the survey results. Särndal, Swensson and Wretman (1991) examined the some important aspects of *survey planning*.

Ideally, survey planning should lead to an optimal specification for the survey as a whole. The term *total survey design* has come to be used for planning processes that aim at overall optimization in a survey. The concept arose out of a desire for an overall control of all sources errors in a survey.

Several survey design is concerned with obtaining the best possible precision in the survey estimates while striking an overall economic balance between sampling and nonsampling errors.

Särndal, Swensson and Wretman (1991) followed Dalenius (1974) and they summarized the total survey design process in a diagram that Figure 2.1.

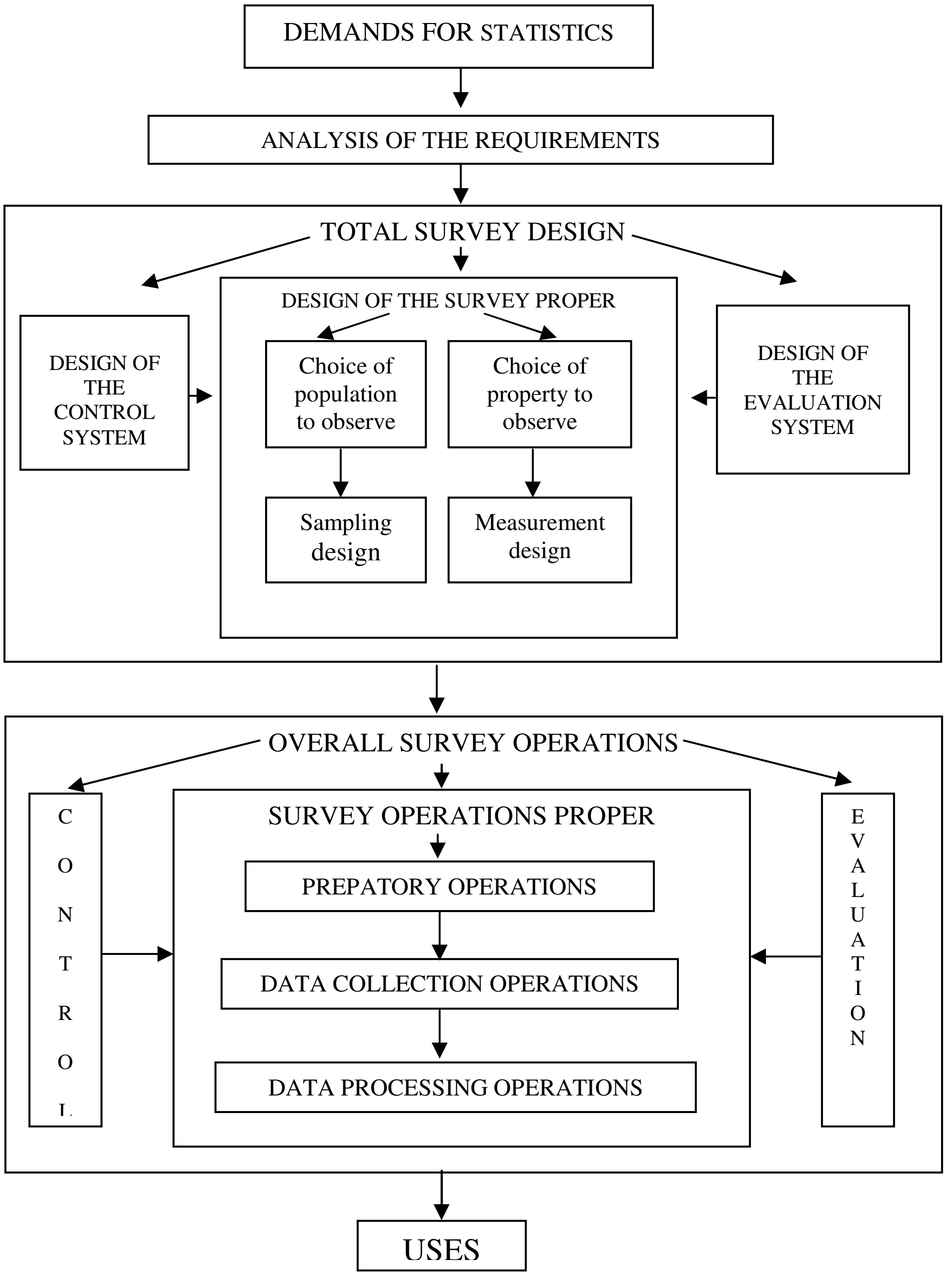

Figure2.1. Total Survey Design (Särndal, Swensson and Wretman, 1991, p.20)

A survey aims at obtaining information about a *target population*. The delimitation of the target population must be clearly stated at the planning stage of the survey. The statistician's interest does not lie in publishing information about individual elements of the target population, but in providing descriptive measures (totals or functions of totals) for various domains, that is, for various aggregates of population elements.

These unknown quantities are called *parameters or parameters of interests*. For example, three important objectives of a Labour Force Survey are to get information about the number of unemployed, the number of employed, the unemployment rate. These are examples of parameters. The first two parameters are *population totals*, the third is a *ratio of population totals*, namely, the number of unemploeds persons divided by the total number of persons divided by the total number of persons in the labour force. Examples of other population parameters are *population means*, for example, mean household income. And *regression coefficients*, for example regression coefficient of income (dependent variable) regressed on number of formal education (independent variable) , for a population individuals.

It can be estimated any of these parameters with the aid of data on the elements of a probability sample from the population. It is then assumed that all sampled elements are measured for the variables whose totals define the parameter of interest.

The computational load is increased by the fact that most surveys require estimation not only for the whole population but for a perhaps considerable number of subpopulations as well. They are called *domains of study* or *domains of interest* or simply *domains*. A domain of interest can be any subpopulation. Some domains may be very small in the sense that very few observed y-values fall into it. The precision of any estimate made for such a domain will be questionable. A special case arises when the domains form a set of mutually exclusive and exhaustive subpopulations. The domains are then said to form a *partition* of the population U.

The domains of interests are denoted by $U_1,...,U_d,...,U_D$. If it is wanted to estimate the total of the variable y for each domain seperately, the targets of estimation are then the D quantities $Y_1,...,Y_d,...,Y_D$. If the survey is required to give accurate information about many domains, a complete enumeration of these domains may become necessary, especially if they are small.

Sampling design is used as a generic term for the (usually probabilistic) rule that governs the sample selection. It is assumed that there is a function $p(.)$ such that $p(s)$ gives the probability of selecting s under the scheme in use. The function $p(.)$ will be called the sampling design. It plays central role because it determines the essential properties like sampling distribution, expected value, variance of random quantities calculated from a sample. Generally used sampling designs are:

- ✓ Simple random sampling (SRS),
- ✓ Stratified simple random sampling (STRS),
- ✓ Cluster sampling,
- ✓ Two-stage sampling,
- ✓ Poisson sampling.

With the possible exception of SRS, these designs require some planning before carried out. STSRS requires well-defined strata composition. Cluster sampling requires a decision on what clusters to use. Two-stage sampling requires that defined the first stage sampling unit (the psu's) and the second stage elements (the ssu's). A strategy is the combination of a sampling design and an estimator. For a given parameter, the general aim is to find the best possible strategy, that is, one that estimates the parameter as accurately as possible.

Two other important general concepts that are involved by every sampling design are:

- ✓ Inclusion probabilities,
- ✓ Design weights.

The inclusion probability of an element is the probability with which it is selected under the given sampling design.

An interesting feature of a finite population of N labeled elements is that the elements can be given different probabilities of inclusion in the sample. The sampling statistician often takes advantage of the identifiability of the elements by deliberately attaching different inclusion probabilities to the various elements. This is one way to obtain more accurate estimates.

It is supposed that a certain sampling design has been fixed. That is, $p(s)$, the probability of selecting s, has a given mathematical form. The inclusion of a given element k in a sample is a random event indicated by the random variable I_k, defined as

$$I_k = \begin{cases} 1 & if\ k \in S \\ 0 & if\ not \end{cases} \tag{2.1}$$

$I_k = I_k(S)$ is a function of the random variable S. I_k is called the *sample membership indicator* of element k. The basic properties of the statistics $I_k = I_k(S)$ for $k = 1,...,N$ is important. These are described as in the following.

For arbitrary sampling design $p(s)$, and for $k, l = 1,...,N$,

$$E(I_k) = \pi_k \tag{2.2}$$

$$V(I_k) = \pi_k(1 - \pi_k) \tag{2.3}$$

$$C(I_k, I_l) = \pi_{kl} - \pi_k \pi_l \tag{2.4}$$

for more details, see Särndal, Swensson and Wretman (1991).

The probability that element k will be included in a sample, denoted π_k, is obtained from the given design $p(.)$ as follows:

$$\pi_k = \Pr(k \in S) = \Pr(I_k = 1) = \sum_{k \in s} p(s) \tag{2.5}$$

Here, $k \in s$ denotes that the sum is over those samples *s* that contain the given *k*. The probability that both of the elements *k* and *l* will be included is denoted π_{kl} and is obtained from the given $p(.)$ as follows:

$$\pi_{kl} = \Pr(k \,\&\, l \in S) = \Pr(I_k I_l = 1) = \sum_{k\&l \in s} p(s) \tag{2.6}$$

$\pi_{kl} = \pi_{lk}$ for all *k*, *l*. If (2.6) applies also *k* = *l*, for in that case

$$\pi_{kk} = \Pr(I_k^2 = 1) = \Pr(I = 1) = \pi_k \tag{2.7}$$

With a given design $p(.)$ are associated the *N* quantities

$$\pi_1, \ldots, \pi_k, \ldots, \pi_N$$

They constitute the set of first-order inclusion probababilities. Moreover, with $p(.)$ are associated the $N(N-1)/2$ quantities

$$\pi_{12}, \pi_{13} \ldots, \pi_{kl}, \ldots, \pi_{N-1,N}$$

which are called the set of second-order inclusion probabilities. For computation of variance estimates it is needed to second-order inclusion probabilities. Inclusion probabilities of higher order can be defined and calculated for a given design $p(.)$.

For example, the inclusion probabilities of first and second orders can be calculated that considered the simple random sampling without replacement. There are exactly $\binom{N-1}{n-1}$ samples s that include the element *k*, and exactly $\binom{N-2}{n-2}$ samples *s* that include the element *k* and *l* $(k \neq l)$. All samples of size *n* have the same probability, $1\Big/\binom{N}{n}$, it is obtained from (2.5) and (2.6)

$$\pi_k = \sum_{k \in s} p(s) = \binom{N-1}{n-1} \Big/ \binom{N}{n} = \frac{n}{N} \qquad k = 1, ..., N \tag{2.8}$$

and

$$\pi_{kl} = \sum_{k \& l \in s} p(s) = \binom{N-2}{n-2} \Big/ \binom{N}{n} = \frac{n(n-1)}{N(N-1)} \qquad k \neq l = 1,...,N \tag{2.9}$$

Unless otherwise stated, it ia assumed that the sampling design is such that all first-order inclusion probabilities π_k are strictly positive, that is,

$$\pi_k > 0 \ , \quad \text{all} \quad k \in U \tag{2.10}$$

This requirement ensures that every element has a chance to appear in the sample. In order that a sampling design be called a *probability sampling design*, the must satisfy (2.10). A sample s realized by such a design is called a *probability sample*.

Another important property of a design occurs when the condition

$$\pi_{kl} > 0 \ , \quad \text{all} \quad k \neq l \in U \tag{2.11}$$

holds. A sampling design is said to be *measurable* if (2.10) and (2.11) are satisfied. A measurable design allows the calculations of valid variance estimates and valid confidence interval based on the observed survey data.

The *design weight* of an element is the inverse of this inclusion probability. And the design weight of element k is denoted as

$$d_k = \frac{1}{\pi_k} \tag{2.12}$$

Similarly the design weight of element *k&l* is denoted as

$$d_{kl} = \frac{1}{\pi_{kl}} \tag{2.13}$$

The design weights are very important for computing point estimators. The sampling design may generate different probabilities of selection for different elements. For example, in SRS and STRS with proportional allocation, all inclusion probabilities are equal, but this is not the case in general. The inclusion probability can never exceed one. Consequently, a design weight is greater than or equal to one. The inclusion probability and the design weight is equal to for an element that is selected with certainty.

CHAPTER THREE

ESTIMATION UNDER IDEAL CONDITIONS

Nonresponse is normal characteristic of any survey. But it is undesirable, because without it the quality of the statistics and the accuracy of the estimates would generally be better. This can be not completely treated at the *design stage* of the survey, so a procedure is needed for dealing with these annoyance factor at the *estimation stage*. It is easier to develop the principles for estimation under the assumption that the annoyance factor is absent. The following methods are used for this state. (Lundström and Särndal, 2001)

3.1 Horvitz-Thompson Estimator

At the beginning of explanation of Horvitz-Thompson estimator, at first, it must be examined the π estimator. Let us consider a population consisting of N elements labeled $k=1,...,N$, $\{u_1,...,u_k,...,u_N\}$. Thus the finite population is denoted as $U=\{1,...,k,...,N\}$. y denote a variable is called study variable, and y_k be the value of y for the *kth* population element. Thus, the population total that we want to estimate is

$$Y=\sum_U y_k \tag{3.1}$$

The π estimator for the population total is

$$\hat{Y}_\pi=\sum_s \frac{y_k}{\pi_k} \tag{3.2}$$

where π_k is the inclusion probability that the kth unit is in the sample. $\hat{Y}_\pi$ can be expressed as a linear function of indicators I_k ,

$$\hat{Y}_{\pi} = \sum_{U} I_k \frac{y_k}{\pi_k} \tag{3.3}$$

where I_k is called the sample membership indicator of element k is a random variable and is defined by (2.1).

The π expansion has the effect of increasing the importance of the elements in the sample. Because the sample contains fewer elements than the population, an expansion is required to reach the level of the whole population.

Horvitz and Thompson (1952) used the principle of π expansion to estimate the total $Y = \sum_{U} y_k$, and formula (3.2) is often called the Horvitz-Thompson estimator. Thus the Horvitz-Thompson estimator can be defined as

$$\hat{Y}_{HT} = \sum_{s} \frac{y_k}{\pi_k} \tag{3.4}$$

This estimator's distribution structure depends on unit numbers. So, it depends on I_k random variables.

Thus the design weight d_k is written in the estimate formula, the Horvitz-Thompson estimator for the population total is

$$\hat{Y}_{HT} = \sum_{s} d_k y_k \tag{3.5}$$

Under any sampling design satisfying $\pi_k > 0$ for all elements *k*, this estimator is unbiased for *Y*. The Horvitz-Thompson estimator is easily shown to be unbiased for *Y* as follows:

$$E\left(\hat{Y}_{HT}\right)=E\left(\sum_{k=1}^{N} I_k \frac{y_k}{\pi_k}\right)=\sum_{k=1}^{N} \pi_k \frac{y_k}{\pi_k}=\sum_{k=1}^{N} y_k = Y \tag{3.6}$$

where $E(I_k)$ is described by (2.2). (Lohr. 1999)

3.1.1 Variance Estimation of the Horvitz-Thompson Estimator

The variance of the Horvitz-Thompson estimator is

$$V\left(\hat{Y}_{HT}\right)=\sum\sum_{U}\left(\pi_{kl}-\pi_k\pi_l\right)\frac{y_k}{\pi_k}\frac{y_l}{\pi_l} \tag{3.7}$$

and the estimation of this variance is

$$\hat{V}\left(\hat{Y}_{HT}\right)=\sum\sum_{U}\frac{1}{\pi_{kl}}\left(\pi_{kl}-\pi_k\pi_l\right)\frac{y_k}{\pi_k}\frac{y_l}{\pi_l} \tag{3.8}$$

If the sampling design has been fixed, the variance and other statistical properties of $\hat{Y}_{HT}$ are also fixed. In other words, the variance of $\hat{Y}_{HT}$ cannot be changed after sampling and data collection because it is determined entirely by the choice of sampling design. As a consequence, then sampling design should be chosen so as to obtain a small variance for this estimator if the plan is to use the Horvitz-Thompson estimator. (Lundström and Särndal, 2001)

3.2 The Generalized Regression Estimator

The emphasis laid on the use of auxiliary information for improving the precision of estimates is characteristics of sampling theory. The regression estimator is one type of estimator that attempts to make efficient use of auxiliary information about the population. (Särndal, Swensson and Wretman, 1991) A wider and more efficient class of estimators is those that use auxiliary information explicitly at the estimation stage. Some information may already have been used at the design stage. Denote the auxiliary vector by **x**, and its value for element *k* by $\mathbf{x}_k = (x_{1k}, \ldots, x_{jk}, \ldots, x_{Jk})'$, a column vector with J components, where x_{jk} is the value, for element *k*, of the *jth* auxiliary variable. It is assumed that the population total $\sum_U \boldsymbol{x_k}$ is accurately known.

An estimator that uses this information is the *generalized regression estimator* (GREG estimator). This estimator is explained and illustrated by several examples in Sarndal, Swensson and Wretman (1991). The generalized regression estimator is given by

$$\hat{Y}_{GREG} = \hat{Y}_{HT} + \left(\sum_U x_k - \sum_s d_k x_k\right)' \hat{B} \tag{3.9}$$

where

$$\hat{B} = \left(\sum_s d_k q_k x_k x_k'\right)^{-1} \left(\sum_s d_k q_k x_k y_k\right) \tag{3.10}$$

is a vector of regression coefficients, obtained by fitting the regression of y on x, using the data $(y_k, \boldsymbol{x}_k)$ for the elements $k \in s$. The data are weighted by $d_k q_k$, where the factor q_k is specified by statistician. A simple choice is to take $q_k = 1$ for all *k*.

The GREG estimator is almost unbiased. The bias, although not exactly zero, tends to zero with increasing sample size and even for modest sample sizes it is normally so small that we do not need to consider it.

The term $\left(\sum_U x_k - \sum_s d_k x_k\right)' \hat{B}$ in the formula for $\hat{Y}_{GREG}$ can be viewed as a regression adjustment applied to the HT estimator, $\hat{Y}_{HT} = \sum_s d_k y_k$. The effect is an important reduction of the variance of $\hat{Y}_{HT}$, especially when there is a strong regression relationship between *y* and $\boldsymbol{x}$.

$\hat{Y}_{GREG}$ the GREG estimator is in reality a whole set of estimators, corresponding to the different specifications that Särndal and Lundström (2001) gave to the auxiliary vector x_k and to the factor q_k. If a number of auxiliary variables, or x-variables, each with a known population total, are available at the estimation stage, it may be included in x_k those x-variables that promise to be the most efficient ones for reducing the variance. That is, some or all of the available x-variables are selected for inclusion in the auxiliary vector x_k. Consequently, the vector x_k to be used in $\hat{Y}_{GREG}$ can take a variety of forms, given that a certain quantity of auxiliary information.

It can be waited until after sampling and data collection to specify which of the possible GREG estimators for use, because the decision on the x-variables to include in x_k need to be made until after these survey operations have been completed.

The estimator $\hat{Y}_{GREG}$ is expressed as a linearly weighted sum of the observed values y_k. When do this, $\hat{Y}_{GREG}$ is

$$\hat{Y}_{GREG} = \sum_{s} d_k g_k y_k \tag{3.11}$$

where the total weight given to the value y_k is the product of two weights, the design weight $d_k = 1/\pi_k$ and the weight g_k. The weight g_k depends both on the element k and on the whole sample s of which k is a member. It is given by

$$g_k = 1 + q_k \left(\sum_{U} x_k - \sum_{s} d_k x_k \right)' \left(\sum_{s} d_k q_k x_k x_k' \right)^{-1} x_k \tag{3.12}$$

The value of g_k is near unity for a majority of the elements $k \in s$, and the greater the size of the sample s, the stronger is the tendency for the g_k to hover close to unity. It is rare to find elements with a weight g_k that is greater than 4 or less than 0.

Negative weights are allowed; such weights do not invalidate the theory, but some users would like all weights to be positive.

The HT estimator is a special case of $\hat{Y}_{GREG}$, obtained when

i. $x_k = q_k = 1$ for all $k \in s$,

ii. The design satisfies $\sum_s d_k = N$.

the condition (ii) holds, for example, for the SRS and STSRS designs.

When the weight system $d_k q_k$ is applied to the auxiliary x_k , and sum over the elements, an estimate the population total of x_k is obtained. This estimate turns out to be exactly equal to the known value of that total, that is,

$$\sum_s d_k g_k x_k = \sum_U x_k \tag{3.13}$$

The weight system is called *calibrated* or, sometimes, *consistent*. More specifically, it is calibrated to the known population total $\sum_U x_k$.

3.2.1 The Variance Estimation of the Generalized Regression Estimator

An important survey objective is to estimate the variance $V(\hat{Y})$. The usual procedure is to start with the formula for the variance, and to transform it into an estimated variance. Once computed from the sample data, the estimated variance, denoted $V(\hat{Y})$, opens up the possibility of assessing the precision of $\hat{Y}$. For example $\sqrt{V(\hat{Y})}$ is used, and the point estimate $\hat{Y}$, to compute a confidence interval for the unknown parameter *Y*.

The expression is given for the variance of $\hat{Y}_{GREG}$, in order to close approximation, is

$$V\left(\hat{Y}_{GREG}\right) = \sum\sum_{U}\left(\frac{d_k d_l}{d_{kl}} - 1\right)E_k E_l \tag{3.14}$$

where the residuals are those arising from the "population regression fit". This fit, which cannot be carried out in practice, has the residuals

$$E_k = y_k - x_k' B \tag{3.15}$$

where

$$B = \left(\sum_{U} q_k x_k x_k'\right)^{-1}\left(\sum_{U} q_k x_k y_k\right) \tag{3.16}$$

$\hat{V}\left(\hat{Y}_{GREG}\right)$ is a function of the regression residuals arising from the regression of y_k on the auxiliary vector x_k, and is such that the smaller these residuals, the smaller the estimated variance of estimator $\hat{Y}_{GREG}$, which makes good intuitive sense. Then the variance estimator of the $\hat{Y}_{GREG}$ is

$$\hat{V}\left(\hat{Y}_{GREG}\right) = \sum\sum_{s}\left(d_k d_l - d_{kl}\right)\left(g_k e_k\right)\left(g_l e_l\right) \tag{3.17}$$

where

$$e_k = y_k - x_k' \hat{B} \tag{3.18}$$

with $\hat{B}$ determined by (3.10). This formula requires that all first and second order inclusion probabilities be strictly positive.

CHAPTER FOUR

NONRESPONSE

4.1 Definition of Nonresponse

By nonresponse is meant that the desired data are not obtained for the entire set of elements *s*, designated for observation. The objective of the survey is to observe the sampled elements with respect to q study variables, $y_1,..., y_j,..., y_q$. These may corresponds to q items on a questionnaire. y_{jk} denotes the value of the variable y_j for the element k and n_s denotes the size of *s*. (Särndal, Swensson and Deville, 1991)

By *full response* in the survey is meant that, after data collection and edit, the available data consist, for every $k \in s$, of a complete q-vector of observed values

$$y_k = (y_{1k},..., y_{jk},...y_{qk})$$

these values form a data matrix of dimension $n_s \times q$, with no value missing. In all other cases, there is *nonresponse*. That is after data collection and edit the $n_s \times q$ data matrix is *incomplete.* One or more of the $n_s \times q$ desired y_{jk} values are missing; there are some "blanks" instead of values in the data matrix.

Full response is seldom realized in a survey. There is variety of reasons for missing values y_{jk}. In a mail survey, the questionnaire may not be returned, or returned but not completely filled in. In a survey with personal interviewers, some individuals refuse to respond to some or all of the questions. Some individuals are not found at home, despite repeated calls. Illness or language problems may make it impossible to carry out the interview. A value supplied on a questionnaire or obtained at an interview may fail an edit check. Such a value may then be regarded as false or strongly suspect and recorded as a blank. (Särndal, Swensson and Deville, 1991)

4.1.1 Response Set

r_j denotes the *j*th response set, that is, the subset of sample *s* for which acceptable responses to the *j*th item are recorded. It can thus write

$$r_j = \{ k : k \in s \text{ and } y_{jk} \text{ is recorded } \}.$$

In the survey, there are *q* usually nonidentical response sets, $r_1,...,r_q$. In the case of a sample survey, the set *s* is the selected with a known sampling design. The response set r_j is a sub selection from *s*.

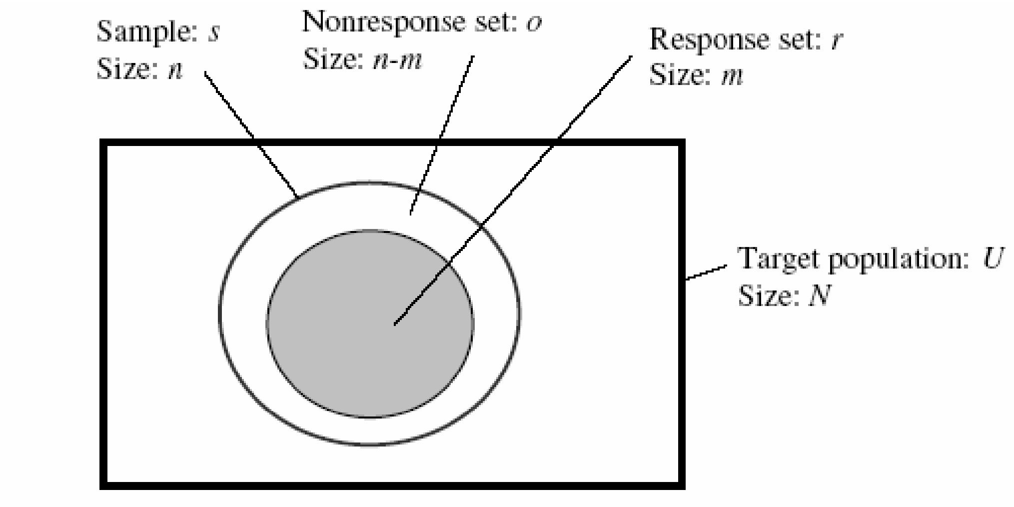

Figure 4.1 Illustration of nonresponse and response set and the sample. Lundström& Särndal, 2001, p.57)

In the Figure 4.1, it is assumed that response is obtained for the element in a set denoted r. Full response implies that *r* = *s*. Nonresponse implies that *r* is a proper set of *s*. The nonresponse set is denoted *o* = *s*-*r*.

4.1.2 Measuring Nonresponse

Nonresponse has more one reason to worry about its harmful effects on the survey estimates. The bias often increases with the rate of nonresponse. It is very difficult to get objective measures of the bias, but it is relatively simple to quantify the extent of the nonresponse. Different measures of the nonresponse are usually found in the quality declarations that statistical agencies and survey institutes often publish together with the survey results. The user can take this information into account when judging the credibility of the results.

A number of descriptive measures are used for the response, or its complement, the nonresponse. Two types of missing information for an element *k* can be distinguished from which response is solicited.

4.1.2.1 Item Nonresponse

The element *k* is an *item nonresponse* element if at least one, but not at all *q*, components of the vector $\mathbf{y_k} = (y_{1k} ,..., y_{jk} ,..., y_{qk})$ are missing. For example, the respondent returns a partially filled in questionnaire, or an interview results in responses to some but not all questions.

In the item nonresponse, the missing values are replaced by the proxy value from the response set, mean imputation, hot-deck imputation and regression estimation. (Son, Jung, 2004)

4.1.2.2 Unit Nonresponse

The element k is a *unit nonresponse* element if the entire vector of y -values, $\mathbf{y_k} = (y_{1k} ,..., y_{jk} ,..., y_{qk})$, is missing. An example is when the respondent fails to return the questionnaire, even after one or more reminders, or if he or she refuses to participate in a personal interview.

In the unit nonresponse it is adjusted the original design weight by new weight using auxiliary information which strong correlated with the variable of interested, such as the weighting adjustment, raking and calibration method. (Son and Jung, 2004)

In the following Table (4.1), Särndal and Lundström (2001) illustrated the results of a hypothetical data collection in a survey with 8 sampled elements. The symbol *x* indicates a presence of data, *nr* indicates that data are missing. In this table, for the two register variables all 8 sample element have data but elements 7 and 8 constitute the unit nonresponse, because neither of these has any response in the questionnaire part of the survey. Elements 2, 3 and 6 have the item nonresponse which values recorded for at least one questionnaire item Elements 1 to 6 forms the response set in this table.

Table 4.1 Unit nonresponse, item nonresponse and response set (Lundström& Särndal, 2001, p.25)

Identity	Register variables		Questionnaire variables		
	1	2	1	2	3
1	*x*	*x*	*x*	*x*	*x*
2	*x*	*x*	*x*	*x*	*nr*
3	*x*	*x*	*x*	*nr*	*x*
4	*x*	*x*	*x*	*x*	*x*
5	*x*	*x*	*x*	*x*	*x*
6	*x*	*x*	*nr*	*x*	*nr*
7	*x*	*x*	*nr*	*nr*	*nr*
8	*x*	*x*	*nr*	*nr*	*nr*

4.2 Sources of Nonresponse

Nonresponse refers to many sources of failure to obtain observations (responses, measurements) on some elements selected and designated for the sample. If accurate accounts are kept of all eligible elements that fall into the sample, the nonresponse rate can be measured. These are necessary for understanding the sources of nonresponse, for its control and reduction, for predicting it in future surveys, and for estimating its possible effects on the surveys. According to Kish (1969), when the classification was done with respect to response set, there was no need that definition of sampling unit or observation unit angle of sources of nonresponse. Furthermore, reporting the extent of nonresponse has become an accepted responsibility for better surveys. (Kish, 1969)

If the many possible sources of nonresponse are sorted into a few meaningful classes, these aims can be better served. A good classification of nonresponse depends on the survey situation. During classification of nonresponse sources, giving attention to the number of classes is very important. Because when the amount of classes is very high, nonresponse analysis is difficult or when the amount of classes is very low, some information can be lost in nonresponse analysis. Nonresponse can have many different causes; as a result, no single method can be recommended for every survey. (Teksoy, 1991)

In the literature, various classification types are seen about sources of nonresponse. For example, while Moser and Kalton (1971) evaluated that the states “not at homes” and “out of the city” in different classes, Kish (1969) classified two of them in the same group. Kish (1969) classified the sources of nonresponse as follows:

- ✓ Not at homes
- ✓ Refusals
- ✓ Incapacity or inability
- ✓ Not found
- ✓ Lost schedules

These categories refer to nonresponse involving the entire interview or questionnaire.

Platek (1977) classifies sources of nonresponse as related to

- ✓ Survey content,
- ✓ Methods of data collection,
- ✓ Respondent characteristic,

and illustrates various sources using the diagram in Figure 4.2 Groves (1989) and Dillman (1978) discuss additional sources of nonresponse. The following that are classified by Lohr (1999) are some factors that may influence response rate and data accuracy.

- ✓ *Survey content.* A survey on drug use or financial matters may have a large number of refusals. Sometimes the response rate can be increased for sensitive items by careful ordering of the questions or by using a randomized response technique.
- ✓ *Time of survey.* Some calling periods or seasons of the year may yield higher response rates than others. The vacation month of August, for example, would be a bad time to take a one-time household survey in Germany.
- ✓ *Interviewers.* Gower (1979) found a large variability in response rates achieved by different interviewers, with about 15% of interviewers reporting almost no nonresponse. Standard quality-improvement methods can be applied to increase the response rate and accuracy for interviewers. The same methods can be applied to the data-coding process.
- ✓ *Data-collection method.* Generally, telephone and mail surveys have a lower response rate than in-person surveys and they also have lower costs. Mail, fax and internet surveys often have low response rates. Possible reasons in a mail survey should be explored before the questionnaire is mailed: Is the survey sent to the wrong address? Do recipients discard the envelope as junk mail even before opening it? Will the survey reach the intended recipient? Will the recipient believe that filling out the survey is worth time?

- ✓ *Questionnaire design.* Question wording has a large effect on the responses received; it can also affect whether a person responds to an item on the questionnaire. In a mail survey, a well-designed form for the respondent may increase data accuracy.
- ✓ *Respondent burden.* Persons who respond to a survey are doing you an immense favor and the survey should be as nonintrusive as possible. A shorter questionnaire, requiring less detail, may reduce the burden to the respondent. Techniques such as stratified in reduce respondent burden because a smaller sample suffices to give the required precision.
- ✓ *Incentives and disincentives.* Incentives, financial or otherwise, may increase the response rate.
- ✓ *Follow-up.* The initial contact of the sample is usually less costly per unit than follow-ups of the initial nonrespondents. If the initial survey is by mail, a reminder may increase the response rate. Not everyone responds to follow-up calls, though, some persons will refuse to respond to the survey no matter how often they are contacted.

According to Lohr (1999), at least some information about nonrespondents is tried to obtain that can be used later to adjust for the nonresponse. There is no complete compensation for not having the data, but partial information may be better than none. Information about the race, sex, or age of a nonrespondent may be used later to adjust for nonresponse. Questions about income may well lead to refusals, but questions about cars, employment or education may be answered and can be used to predict income. If the pretests of the survey indicate a nonresponse problem that you do not know how to prevent, try to design the survey so that at least some information is collected for each observation unit.

4.3 Dealing with Nonresponse

Nonresponse is due to an effective reduction of sample size and bias estimates in survey results. The sample variance increases with the reduction of sample size so that the precision of estimates is reduced. For these important reasons nonresponse problem is very important. (Teksoy, 1991)

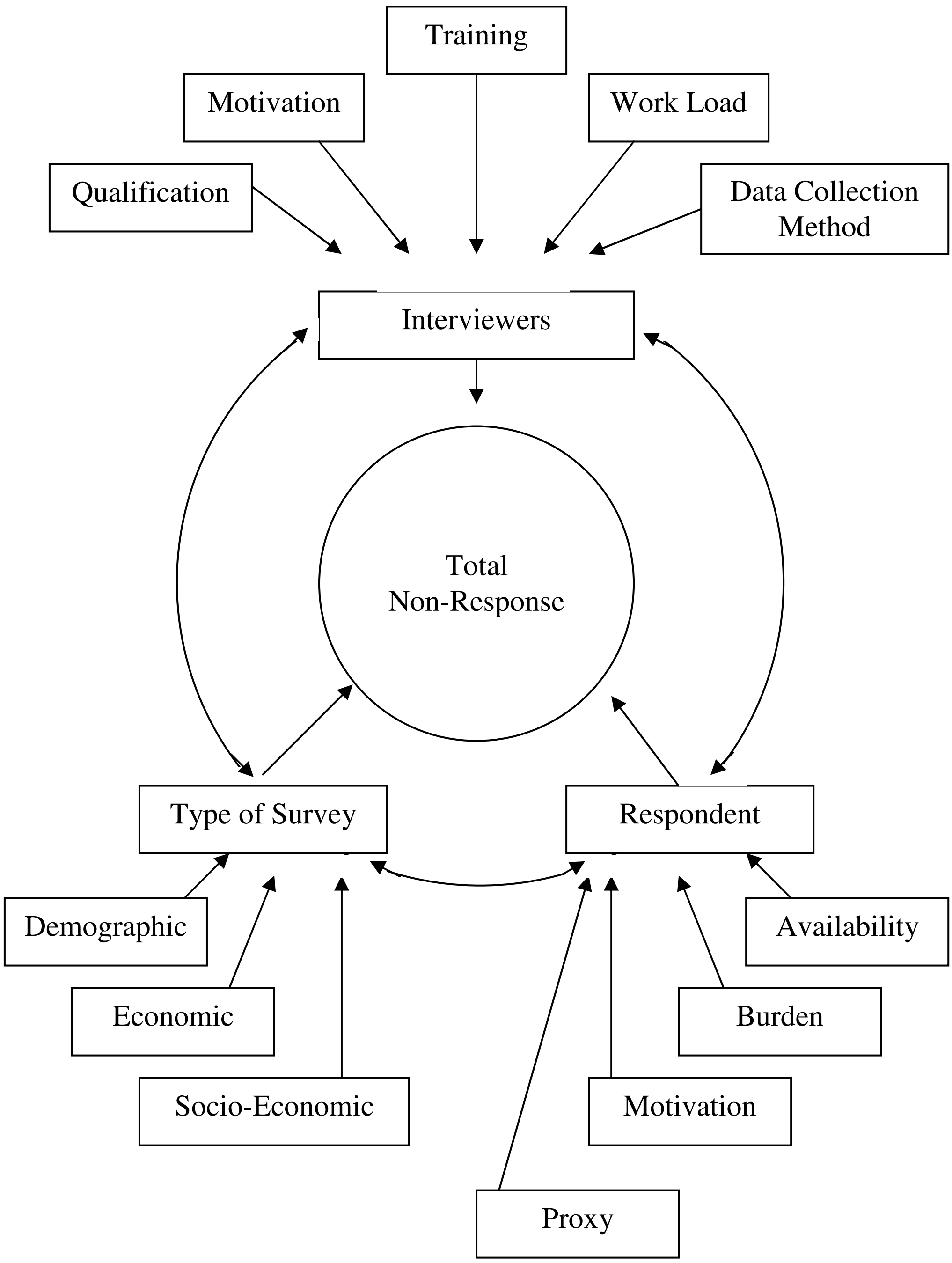

Figure 4.2 Some Factors Affecting Nonresponse. (Lohr, 1999, p.260)

The best way to deal with nonresponse is to prevent it. After nonresponse has occurred, it is sometimes possible to model the data, but predicting the missing observations is never as good as observing them in the first place. Nonrespondents often differ in critical ways from respondents. If the nonresponse rate is not negligible, inference based upon only the respondents may be seriously flawed. (Lohr, 1999)

According to Kish (1969), the following methods can be used in different situations to reduce either the percentages of nonresponse or its effects:

- ✓ Development of Data collection method
- ✓ Call-backs
- ✓ Subsampling Nonresponses
- ✓ Estimating the effect of nonresponse
- ✓ Substitutions for Nonresponse
- ✓ Kish and Hess Replacement Procedure
- ✓ Politz and Simmons Method
- ✓ Weighting Adjustment

Strategies for dealing nonresponse are classified by Särndal, Swensson and Wretman, (1991) as follows:

1. Effective measures are taken to reduce the nonresponse to insignificant levels before and during data collection, so that any remaining nonresponse causes little to the validity of the inferences. There are two approaches under this strategy and these are:

Planning of the Survey: Special efforts must be made at the planning stage to foresee how alternative survey operations may influence the response. The choice of data collection method, such as personal interview, telephone interview, mail inquiry or other, is important because of the length and the content of the questionnaire or schedule.

Callbacks and Follow-Ups: In surveys with personal interviews, if the first attempted contact results in too many unsuccessful interviews, it is common to make one or more callbacks at more convenient times. Callbacks can also give valuable information about the selective effects of nonresponse. Follow-ups is generally used in mail surveys.

2. Special, perhaps costly techniques for data collection and estimation are used that permit unbiased estimation. Two techniques in this strategy are:

Subsampling of Nonrespondents: A long series of callbacks or follow-ups may prove costly and time consuming. An alternative approach is to take a subsample of the nonrespondents, and make every possible effort to obtain responses from all elements in the subsample. There are several schemes for subsampling the nonrespondents.

Randomized Response: In some surveys, many respondents either refuse to participate or give false or evasive responses. A solution provided with methods is that protect anonymity. Such protection is built into the randomized response technique, introduced by Warner (1965). The assumption is that randomized response will ensure the cooperation of all selected individuals, as well as truthful responses. The method is used for surveys where data collection is by personal interview.

3. Model assumptions about the response mechanism and about relations between variables are used to construct estimators that "adjust" for a nonresponse that cannot be considered harmless. Two techniques in this strategy are:

Weighting Adjustment: To compensate for the values that are lost of nonresponse, weighting adjustment implies that higher than normal weights are applied in the estimation to the *y* values of the respondents. Weighting adjustment based on response distribution modeling and estimation of response probabilities.

Imputation: Imputation implies the substitution of "good" artificial values for the missing values. There are two principal uses of imputation. At the first, imputation for the item nonresponse only. Then Imputed values are provided for missing values corresponding to elements *k* in the item nonresponse set. Weighting adjustment is then applied to compensate for the unit nonresponse. At the second, imputation for the item nonresponse as well as for the unit nonresponse. Then imputed values are provided for missing values. No weighting adjustment is applied and estimates are computed using original as well as imputed values.

The nonresponse is affected by a number of the operations that define the survey. The ideal survey has no nonresponse. For satisfies this ideal, it requires careful planning and often considerable expense.

4.4 Nonresponse Adjustment

Nonresponse adjustment is a term for the various attempts made by statisticians to deal with nonresponse. By the word "adjustment" is meant that changes are made to an original or "ideal" estimation procedure, namely, the one intended for use in the ideal case of full response. The principal methods for nonresponse adjustment are *reweigthing* and *imputation*. (Lundström and Särndal, 2001)

4.4.1 The Importance of Auxiliary Information

Recent years have seen theoretical developments and increased use of methods that take account of substantial amounts of auxiliary. The key to successful nonresponse adjustment lies in the use of “strong” auxiliary information. Such use will reduce both nonresponse bias and the variance.

Register variables play an important role in many of surveys. They are used in creating an appropriate sampling design and/or in the computation of the survey estimates. In both uses, the register variables can be called auxiliary variable, because they assist and improve the procedures. Most often, the term “auxiliary variables” refers to a variable used at the estimation stage to create better alternatives to the simplest estimators.

Auxiliary information can be used both at the design stage in constructing the sampling design and at the estimation stage in constructing the estimators. In the Reweigthing procedure for nonresponse second type of usage is valid.

Terms frequently used in the following are auxiliary variable, auxiliary vector, auxiliary information and auxiliary population total or totals. It can be explain that use of these terms. The minimum requirement to qualify as an auxiliary variable is that the values of the variable are available for every sampled element that is, for both responding and nonresponding elements. For many surveys, such variable values can be found in available registers, and are then usually known not only for the sampled elements but, more extensively, for all elements in the population.

An auxiliary vector is made up of one or more auxiliary variables. There are two important steps in the process leading to the form of the auxiliary vector that will be ultimately used in the estimation. These are:

a. Making an inventory of potential auxiliary variables;
b. Selecting and preparing the most suitable of these for entry into the auxiliary vector.

The auxiliary variables deemed potentially useful for the estimation may come from several registers allowing the possibility of linking of elements. A rather long list of potential variables may result from this searching look. The next important step is the procedure by which is arrived at the final form of the auxiliary vector to be used in the estimation. This process requires considerable reflection and study. The decisions to be taken include the selection of variables from the available larger set, the setting appropriate group boundaries for converting a quantitative variable into a categorical variable, and fixing rules for collapsing very small groups into larger groups.

The estimator scheduled for use in the survey will usually require a known population total for each variable in the auxiliary vector. The term “auxiliary information” is used with reference both to the auxiliary vector itself and to the known totals for the variables in the vector. When register variables are used in the construction of the sampling design, their values must be known for every element in the population, as when strata are constructed for a stratified design. When auxiliary variables are used at the estimation stage, such detailed information may not be necessary. It may suffice to know the population total for each auxiliary variable, which knowledge of individual variable values may be limited to the sampled elements only.

Lundström and Särndal (2001) illustrates with an example that how nonresponse bias can be reduced by incorporating relevant auxiliary information in the estimation procedure. They also show that the calibration estimator provides an exact estimate of Y where a perfect linear relationship exists between the study variable and the auxiliary vector. Because of that if powerful auxiliary information can be identified and used then both the sampling error and the nonresponse bias will be small.

To reduce the nonresponse bias and variance of the estimator, one should select an auxiliary vector that satisfies as far as possible one or both of the following principles:

a. auxiliary vector explains the variation of the response probabilities.
b. the auxiliary vector explains the variation of the main study variables.
c. the auxiliary vector should identify the most important domains.

When principle (a) is fulfilled the nonresponse bias is reduced in the estimates for all study variables. However, if only principle (b) is fulfilled the nonresponse bias is reduced only in the estimates for the main study variables. Then the variance of these estimates will also be reduced. When principle (c) is fulfilled the effect is mainly a reduction of the variance for the domain estimates. (Lundström and Särndal, 2001)

There exists a large literature on the selection of auxiliary information and on the resulting specification of the auxiliary vector.

CHAPTER FIVE

METHODS FOR THE REDUCTION OF NONRESPONSE BIAS AND ERRORS

5.1 Imputation

5.1.1 Introduction

Imputation entails replacing missing values by proxy values. Imputation is the procedure whereby missing values for one or more study variables are 'filled in' with substitutes. These substitutes can be constructed according to some rule, or they can be observed values but for elements other than the nonrespondents. Thus imputed values are artificial and they contain error. Imputation error is similar to measurement error in that the true value is not recorded. (Särndal and Lundström, 2001)

Imputed values can be classified into three major categories:

(1) values constructed with the aid of a statitical prediction rule;
(2) values observed not for the nonresponding elements themselves, but for responding elements
(3) values constructed by expert opinion or 'best possible judgment'.

First and second categories use a statistical technique to produce a reasonably close substitute value, so they can be termed statistical rules. First category is often based on regression prediction. Second category methods can also be described as donor-based, in that another value of another observed element is imputed. Third category methods are more subjective and often rely heavily expert skill.

Some of the more commonly used statistical rules are:

(1) ratio imputation;
(2) (multiple) regression imputation;
(3) nearest neighbour imputation;
(4) hot deck imputation;
(5) respondent mean imputation.

Imputation is regarded by both statisticians and subject matter specialists with some suspicion. This is because it goes against common statistical sense to use values known at the outset to be more or less wrong. (Särndal and Lundström, 2001)

The imputed values must come as close as possible to the true unobserved values for which they are substitutes. Because of this, the construction of imputed values should be carried outr with professional care. Imputation may be disallowed for legal reasons. Some countries prohibit imputation, at least for some categories of observed elements.

There are two frequently used approaches for imputation, both leading to rectangular data matrices, namely the ITIMP-approach and the UNIMP-approach.

ITIMP-approach: Imputation is used to treat the item nonresponse only. In this procedure, values are imputed for the m elements for which at least one but not all y values are missing. The resulting rectangular data matrix has the dimensions m by J where J is the number of y variables. Reweighting is then applied to compensate for the unit nonresponse.

UNIMP-approach: Imputation is used for both item nonresponse and unit nonresponse. In this procedure , values are imputed for all elements having at least one y value missing. The resulting completed rectangular data matrix has the dimensions n by *J*, where *n* is the sample size. There is no nonresponse weight adjustment.

If the element k is a nonrespondent and imputation is used for this element, the imputed value is denoted as $\hat{y}_k$. More than one imputation method may be used in the same survey,so not all $\hat{y}_k$ may result from the same method.

The completed data set is defined as the set of values $\{y_{.k} = k \in s\}$, where

$$y_{.k} = \begin{cases} y_k & k \in r \\ \hat{y}_k & k \in o \end{cases} \tag{5.1}$$

That is, the value $y_{.k}$ equals the observed value y_k when *k* is a respondent, or the imputed value $\hat{y}_k$ when *k* is a nonrespondent. Traditional descriptive statistics can be computed from the complete data set.

5.1.2 Point Estimation when Imputation is Used

The population total is wanted to estimate used for the variable, *y,* $Y = \sum_U y_k$ when imputation is used. It is supposed that imputed values are considered are as "good" as true observations. Such a belief is a justification for using exactly the same estimation method as in the ideal case of full response. It will be employed the "standard estimator formula" and simply apply it to the completed data set. Consequently, when an estimate is computed, element k will receive the same weight whether its recorded y-value is a true observation, y_k, or an imputed value, $\hat{y}_k$. This is worth pointing out, because some would argue that imputed values should be weighted according to some principle other than that used for truly observed values. Current practice for point estimation is in fact that survey statisticians treat imputed data as real, observed data. That is, the procedure is: determine an estimator suitable for full response, then, after imputation, compute this estimator for the completed data set. (Särndal and Lundström, 2001) The estimator intended for use in the case of full response will be called the *full response estimator*. Here we assume that this estimator is the GREG estimator discussed in Chapter 3.2. of the form $\hat{Y}_{GREG} = \sum_{s} d_k g_k y_k$ as described by formulas (3.11) and (3.12). A special case is the Horwitz-Thompson estimator, $\hat{Y}_{HT} = \sum_s d_k y_k$, discussed in Chapter 3.1.

Now suppose there is nonresponse treated by imputation. Then have a completed data set, given by (5.1). It replaces the desired but not realised data set composed of full real observations. We apply the weighting, $d_k g_k$, of the full response estimator (3.11). This gives the imputed GREG estimator

$$\hat{Y}_I = \sum_{s} d_k g_k y_{.k} \tag{5.2}$$

which can also be written as

$$\hat{Y}_I = \sum_{r} d_k g_k y_k + \sum_{o} d_k g_k \hat{y}_k \tag{5.3}$$

In current practice, point estimation in the presence of imputation is thus very simple, since the weights are not changed. By contrast, variance estimation becomes a complex issue. The imputed HT estimator is

$$\hat{Y}_I = \sum_{s} d_k y_{.k} \tag{5.4}$$

5.2 Reweighting

> *Weights are commonly assigned to respondent records in a survey data file in order to make the weighted records represent the population of inference as closely as possible. The weights are usually developed in a series of stages to compensate for unequal selection probabilities, nonresponse, noncoverage, and sampling fluctuations from known population values. The first stage of weighting for unequal selection probabilities is generally straightforward. Each sampled element whether respondent or nonrespondent is assigned a base weight that is either the inverse of the element's selection probability or proportional to that inverse. The second stage of weight development is usually to attempt to compensate for unit, or total, nonresponse. The base weights of responding elements are adjusted to compensate for the nonresponding elements. The general strategy is to identify respondents who are similar to the nonrespondents in terms of auixiliary information that is available for both respondents and nonrespondents, and then to increase the base weights of respondents so that they represent similar nonrespondents. The third stage of weight development involves a further adjustment to the weights to make the resultant weighted estimates from the sample conform to known population values for some key variables.* (Kalton and Cervantes, 2003)

Reweigthing entails altering the weights of the respondents, compared to the weights that would have been used in the case of full response. Since observations are lost by nonresponse, reweigthing will imply increased weights for all, almost all, of the responding elements. (Särndal and Lundström, 2001)

5.2.1 Two-Phase Approach

For a number of years, survey statisticians favored with the *deterministic model* of survey response for nonresponse. The population was assumed to consist of two non-overlapping parts, a response stratum and a nonresponse stratum. Every element in the former was assumed to respond with certainty if selected for the sample, and every element in the latter stratum had probability zero to respond. An obvious criticism that could be levied against that model is that it is simplistic and unrealistic. Moreover the sizes of the two strata could usually not be assumed to be known. An approach that was sometimes used the total for the response stratum, and then to add a term to compensate for the nonresponse stratum. Several proposed techniques for nonresponse treatment are based on this model. And this model inspired much of the work on nonresponse in central statistical offices in the 1960s and 1970s.

In the 1980s, a more satisfactory *two-phase approach to reweighting for nonresponse* became popular. The name refers to a view of the selection process as one in which a desired sample *s* is first selected from the population *U*, whereupon a set of respondents, *r*, is realized as a subset of s. The approach is more realistic than the deterministic one in that it allows every element *k* to have its own individual response probability $\boldsymbol{\theta}_k$ where $\mathbf{0} \leq \boldsymbol{\theta}_k \leq \mathbf{1}$ for all *k*. This generality is not without a price: the reponse probabilities $\boldsymbol{\theta}_k$ are usually unknown and progress with this approach requires that the $\boldsymbol{\theta}_k$ be replaced with estimates, constructed with the aid of auxiliary information. (Särndal and Lundström, 2001)

The two-phase approach is discussed in the literature, for example, Särndal, Swensson and Wretman(1991). And also, in their book, they develops the theory of two-phase sampling in the presence of auxiliary information. In the traditional formulation of two-phase sampling, a first sample is selected from *U*, certain variable are observed, then a smaller subsample is realized from the first sample, and the study variable(s) are observed for the elements of the subsample. All inclusion probabilities are known by design, those for the first phase as well as those for the second phase.

Särndal, Swensson and Wretman (1991) adapts the two-phase theory to the case where sampling is followed by nonresponse. It is assumed for a moment that the response distribution $q\langle r|s\rangle$ is known. In practice this is not the case. This implies that the first and second order response probabilites,

$$\mathbf{Pr}\langle k \in r \mid s \rangle = \boldsymbol{\theta}_k \tag{5.5}$$

and

$$\mathbf{Pr}\langle k \;\&\; l \in r \mid s \rangle = \boldsymbol{\theta}_{kl} \tag{5.6}$$

are known. $\mathbf{x}_k$ denotes the auxiliary vector to be used in the estimator. Under these conditions, the two-phase GREG estimator of the population total $Y = \sum_U y_k$, as obtained from Särndal, Swensson and Wretman (1991), is given by

$$\hat{Y}_{SSW} = \sum_r d_k g_{k\boldsymbol{\theta}} y_k / \boldsymbol{\theta}_k \tag{5.7}$$

and

$$g_{k\boldsymbol{\theta}} = \mathbf{1} + c_k \left(\sum_U x_k - \sum_r d_k x_k / \boldsymbol{\theta}_k\right)' \left(\sum_r d_k c_k x_k x_k' / \boldsymbol{\theta}_k\right)^{-1} x_k \tag{5.8}$$

The transformation of this estimator into one that is useful for a sample survey with nonresponse requires replacing the unknown $\boldsymbol{\theta}_k$ by estimates $\hat{\boldsymbol{\theta}}_k$. This step entails :

(a) the formulation of a realistic model for the response mechanism with the response probabilities $\boldsymbol{\theta}_k$ as unknown parameters,

(b) the estimation of these response probabilities, using any relevant auxiliary variables and the fact that some sample elements were observed to respond whereas the others did not. (Särndal, Swensson, Wretman, 1991)

An often used model states that the population consists of nonoverlapping groups with the property that all elements within one and the same group respond with the same grup respond with the same probability, and in an independent manner. Such groups are known as response homogeneity groups (RHGs). In a survey of individuals, the groups may be based on age by sex categories, for example. The auxiliary information required is that we can uniquely classify every sampled element, respondent or nonrespondent, into one of the groups. The point estimator obtained from (5.7) when the unknown $\boldsymbol{\theta}_k$ are replaced by the estimates $\hat{\boldsymbol{\theta}}_k$ flowing from this RHG model is discussed in detail in Särndal, Swensson and Wretman (1992)'s book. These authors also give an appropriate variance estimator, composed as a sum of two components, one measuring the sampling variance, the other the nonresponse variance. (Särndal, Swensson, Wretman, 1991)

The point estimator is essentially unbiased if the assumed RHG model is a true representation of the response pattern in the survey; it is diffucult in practice , because, that it in essence impossible to foresee the true response pattern. Other attempts at modelling the response mechanism have been made, including logistic regression modelling, as in Ekholm and Laaksonen(1991) , Alavi and Beaumont (2003).

The two-phase approach to reweighting has the following characterisitics:

- ✓ The modelling of the response mechanism constitutes separate step.
- ✓ If a set of auxiliary variables is available, one subset of these varaibles is used in the estimation of the response mechanism. Another subset is used to formulate the auxiliary vector $\boldsymbol{x}_k$ required for the estimator (5.7) of Y. In this formula θ_k is replaced by $\hat{\theta}_k$.

In practice, the two-phase approach to reweighting requires analysis and decision making. The statician must decide on the best use of the total set of available auxiliary variables. If nothing else, these selection rasks will take time. Therefore, the calipration approach to reweighting is proposed that a simpler alternative to two-phase approach. (Särndal and Lundström, 2001)

5.2.2 The Calibration Approach

5.2.2.1 Introduction

Reweighting is treated by Särndal and Lundström (2001) with the calibration approach, which has the favourable property of incorporating most “Standard” methods found in the different places in the literature.

In a survey, statistician usually control the sample selection process, and thus the design weights can be calculated. However the fact that some of the desired data will be missing. If the response probabilities were known, an unbiased two-phase estimator could be constructed. But, in practice response probabilities never known. In conventional techniques one determines proxies of the response probabilities by modelling the response distribution. There will always be some difference between this model and the true response distribution, su the estimator will suffer from nonresponse bias. (Lundström, 1997)

Särndal and Lundström (1999) can be practically certain that the nonresponse is not the result of a simple random selection mechanism so they try to adjust for the selection bias at the estimation stage. They suggests a simple and a unified approach is called calibration to the use of auxiliary information both the sampling error and the nonresponse bias in a survey. When population totals are used, the resulting point estimators are consistent in the sense that the final weights give perfect estimates when applied to each variable. This approach requires neither a response model nor a regression model but which nevertheless guarantees effective use of auxiliary information.

The calibration estimators derived by Deville and Särndal (1992) are a family of estimators appealing a common base of auxiliary information. A calibration estimator uses calibrated weights, which are as close as possible, according to the given distance measure, to the original sampling design weights, d_k ,while also respecting a set of constraints, the calibration equation. They discusss the merits of different metrics for the distance between w_k and d_k. For every distance measure there is corresponding set of calibrated weights and a calibration estimator.

The calibration procedure generates final weights which are as close as possible to specified initial (design) weights , while respecting known auxiliary population totals or unbiased estimates of these totals. Calibration is used by Särndal and Lundström (2001) as a main tool for nonresponse. This calibration approach requires the formulation of a suitable auxiliary vector, through a selection from a possible larger set of available auxiliary variables. This step follows a few basic and simple principles. The next step is computational.

The calibration approach leads to a calibration estimator of *Y*, denoted $\hat{Y}_W$, and a corresponding variance estimator denoted $\hat{V}(\hat{Y}_W)$. The term "weighting" is denoted by the index W. The calibration approach meets the objective of reducing both the sampling error and the nonresponse error. The approach is general in that it can be applied for most of the common sampling designs and with any number of variables present in the auxiliary vector.

The calibration approach has only a single computational step, in which the calibrated weights are produced. It is thereby more direct than the two-phase approach because it requires no seperate modelling of a nonresponse mechanism. For these reasons, the calibration approach, is better suited for a routine treatment of nonresponse in organization.

There are many people studied about calibration in the literature. Some of them are as follows:

Théberge (1999) extended the calibration technique to estimate population parameters other than totals and means. And he extended the technique to the case where there is no solution to the calibration equation.

Estevao and Särndal (2000) defined and developed an alternative to distance minimization approach , that is the functional form approach. In this approach the calibrated weights are given a simple mathematical form that depends on two parameters.

Wu and Sitter (2001) proposed a model-calibration approach in order to use complete auxiliary information from survey data.

Wu and Luan (2003) proposed optimal calibration estimators for the population mean under the two-phase sampling. The proposed optimal calibration estimators under two-phase sampling are applicable estimation problems under measurement errors or nonresponse.

Wu (2003) showed that the model-calibration estimator for the finite population mean, which was proposed by Wu and Sitter (2001), is optimal among a class of calibration estimator. He also presented optimal calibration estimators for the finite population distribution function , the population variance, the variance of a linear estimator and other quadratic finite population functions.

5.2.2.2 Point Estimation Under Calibration Approach

It is first considered that the finite population of N elements $U = \{1,...,k,...,N\}$. The objective is to estimate the population total $Y = \sum_U y_k$, where y_k is the value of a study variable, *y*, for the *k*th element. The sample of size *n*, s, drawn from *U* with the probability *p(s)*. When nonresponse occurs, the response set *r* of size *m* is obtained, where $r \subseteq s$ and $m \leq n$.

If the survey that interests us had full response ($r = s$), the GREG estimator with a specified vector $\boldsymbol{x}_k$ would be chosen. The population total of $\boldsymbol{x}_k$ - vector, $\sum_U \boldsymbol{x}_k$, is a required input. This estimator would be a good choice , because of the reasons that written in below,

a) it is unbiased,

b) its variance is small when $\boldsymbol{x_k}$ is a good explanatory vector for the study variable y_k,

c) it is consistent in the sense that the weights satisfy the calibration equation $\sum_s d_k g_k x_k = \sum_U x_k$.

However, we are concerned here with surveys with nonresponse, so y_k values are available only for the elements *k* in the response set *r*, a subset of the sample. Then, whatever the estimation technique, there will be some bias. Desirable properties of the chosen estimator are now:

a) a small nonresponse bias,

b) a small total variance,

c) agreement with the GREG estimator when $r = s$.

The total variance is the sum of the sampling variance and the nonresponse variance. Property (a) is particularly important.

The calibration estimator is, like the GREG estimator, formed as a linearly weighted sum of the observed y_k values. It is defined by

$$\hat{Y}_w = \sum_r w_k y_k \tag{5.9}$$

where $w_k = d_k v_k$ and

$$v_k = 1 + q_k \left(\sum_U x_k - \sum_r d_k x_k \right)' \left(\sum_r d_k c_k x_k x_k' \right)^{-1} x_k \quad , k \in r . \tag{5.10}$$

The principle behind the derivation that leads to calibtared weights w_k is to minimize a function measuring the distance between the old weights , d_k, and the new weights , w_k, subject to the calibration equation

$$\sum_r d_k v_k x_k = \sum_U x_k \ . \tag{5.11}$$

The calibrated weights are “as close as possible” with respect to the given distance measure to the design weights d_k, and they ensure consistency with the known auxiliary variable totals.

$E_p(.)$ denotes expectation with respect to the sampling design *p(s),* a measure of average distance reminiscent of the chi-square statistic is $E_p\left\{\sum_s (w_k - d_k)^2 / d_k\right\}$. For more generality in this expression, let the *k*th term have an individual known positive weight $1/q_k$, unrelated to d_k, which gives the average distance

$$E_p\left\{\sum_s (w_k - d_k)^2 / d_k q_k\right\}. \tag{5.12}$$

The uniform weighting $1/q_k = 1$ is likely to dominate in applications, but unequal weights $1/q_k$ are sometimes motivated. (Deville and Särndal, 1992)

The objective is to derive new weights that modify as little as possible the original sampling weights $d_k = 1/\pi_k$, which have the desirable property of yielding unbiased estimates. The survey statistician wants to stay close to these weights. Deville and Särndal (1992) thus seek the minimum of calibration equation subject to (5.12). Acoording to them, this is equivalent to minimizing, for any particular *s*, the quantity $\sum_s (w_k - d_k)^2 / d_k q_k$, subject to constraint, calibration equation. And this minimization leads to calibrated weight.

For a successful reduction of both the sampling error and the response bias, strong auxiliary information is a prerequisite. It is assumed that there exists an auxiliary vector, $\boldsymbol{x}$, containing such information. This vector’s value for the kth element is denoted $\boldsymbol{x}_k$. Särndal and Lundström (1999) define the two "information levels" called Info-S and Info-U and they classified two information level in the following :

a) Info-S : $\boldsymbol{x}_k$ is known for all $k \in s$,

b) Info-U: $\sum_U \boldsymbol{x}_k$ is known and moreover $\boldsymbol{x}_k$ is known for all $k \in s$.

In first case (a) the information goes up to the sample level but in second case (b) the information goes up to the population level. Because of this, in case (b) the information is more extensive than in case (a).

After having specified the auxiliary information, calibrated weights, denoted w_k , are computed and the estimator $\hat{Y}_w = \sum_r w_k y_k$ of Y is constructed. This estimator is called w-estimator or calibration estimator. The weights w_k are "as close as possible" to the d_k , and they also satisfy a calibration equation given for Info-S by

$$\sum_r w_k x_k = \sum_s d_k x_k \tag{5.13}$$

and for Info-U by

$$\sum_r w_k x_k = \sum_U x_k \ . \tag{5.14}$$

Särndal and Lundström (1999) use the calibration technique for y_k values are observed the response set only, rather than for the full sample. They seek new weights w_k that satisfy the calibration equation (5.13) or (5.14). The distance function to be minimized is

$$\sum_r (w_k - d_k)^2 / d_k q_k \tag{5.15}$$

where the q_k are specified positive factors. In the case of full response $(r = s)$, this distance function leads to the generalised regression estimator in Deville and Särndal (1992).

5.2.2.3 Derivation of the Generalised Regression Estimator by Calibration

Calibration estimators in the full response case are described in Deville and Särndal (1992). They seek an estimator of the form $\hat{Y}_{DS} = \sum_s w_k y_k$ with weights w_k as close as possible to the desin weights d_k while respecting the calibration equation $\sum_s w_k x_k = \sum_U x_k$. They minimize the distance measure , $\sum_s (w_k - d_k)^2 / d_k q_k$,by the Lagrange multiplier method subject to single constraint, calibration equation, $\sum_s w_k x_k = \sum_U x_k$.

Lagrange function for the distance function under the single calibration equation can be written like this,

$$L = \sum_s (w_k - d_k)^2 / d_k q_k - \lambda(\sum_s w_k x_k' - \sum_U x_k') . \tag{5.16}$$

$$\frac{\partial L}{\partial w_k} = \frac{2(w_k - d_k)}{d_k q_k} - \lambda(x_k') = 0 \qquad k = 1,...,n$$

$$\frac{\partial L}{\partial w_k} = 2w_k - 2d_k = \lambda x_k' d_k q_k \qquad k = 1,...,n$$

This minimization leads to the calibrated weight,

$$w_k = d_k + \lambda x_k' d_k q_k \qquad k = 1,...,n$$

$$w_k = d_k (1 + \lambda x_k' q_k) \qquad k = 1,...,n . \tag{5.17}$$

In this formulation, λ is the vector of Lagrange multipliers. To obtain this vector, determination of Lagrange function is taken with respect to λ .

$$\frac{\partial L}{\partial \lambda} = -\left(\sum_{s} w_k x_k - \sum_{U} x_k \right) = 0$$

$$\frac{\partial L}{\partial \lambda} = \sum_{U} x_k = \sum_{s} w_k x_k$$

$$\sum_{U} x_k = \sum_{s} \left(d_k + \lambda x_k^{'} d_k q_k \right) x_k$$

$$\sum_{U} x_k = \sum_{s} d_k x_k + \sum_{s} \lambda x_k x_k^{'} d_k q_k$$

$$\sum_{U} x_k - \sum_{s} d_k x_k = \lambda \sum_{s} x_k x_k^{'} d_k q_k$$

$$X - \hat{X}_{HT} = \lambda \sum_{s} x_k x_k^{'} d_k q_k$$

$$\lambda = \frac{\left(X - \hat{X}_{HT} \right)}{\sum_{s} x_k x_k^{'} d_k q_k} \tag{5.18}$$

When we write λ obtained at (5.17) in the calibrated weight (5.18), final form of the calibrated weight is

$$w_k = d_k \left(1 + \frac{\left(X - \hat{X}_{HT} \right)}{\sum_{s} x_k x_k^{'} d_k q_k} x_k^{'} q_k \right) \tag{5.19}$$

Deville and Särndal (1992)'s aim was to obtained the GREG estimator by calibration. Calibration estimator form of the population total was like that

$\hat{Y} = \sum_s w_k y_k$. If we write calibrated weight w_k obtained (5.19) in the calibration estimator formula, this estimator formula leads to GREG estimator.

$$\hat{Y} = \sum_s w_k y_k$$

$$\hat{Y} = \sum_s \left(d_k + \frac{\left(X - \hat{X}_{HT}\right)}{\sum_s x_k x_k' d_k q_k} x_k' q_k d_k \right) y_k$$

$$\hat{Y} = \sum_s d_k y_k + \left(X - \hat{X}_{HT}\right) \frac{\sum_s x_k d_k q_k y_k}{\sum_s x_k x_k' d_k q_k}$$

$$\hat{Y} = \hat{Y}_{HT} + (X - \hat{X}_{HT})\hat{B}_s = \hat{Y}_{GREG} \tag{5.20}$$

where $\hat{Y}_{HT} = \sum_s d_k y_k$ denotes the Horvitz-Thompson estimator for the $\boldsymbol{x}$ vector , $\hat{X}_{HT} = \sum_s d_k x_k$ denotes the Horvitz-Thompson estimator for the y study variable and

$$\hat{B}_s = \frac{\sum_s d_k q_k x_k y_k}{\sum d_k q_k x_k x_k'} \tag{5.21}$$

is a weighted estimator of the multiple regression coefficient. It seen from that, we obtain GREG estimator by calibration approach.

5.2.2.4 Deriving the Calibrated Weights when $\sum_U x_k$ is known

For Info-U, minimization of (5.15) under the constraint (5.14) leads to *w* estimator. Lagrange function is

$$L = \sum_r (w_k - d_k)^2 \Big/ d_k q_k - \lambda(\sum_r w_k x_k' - \sum_U x_k') \qquad (5.22)$$

$$\frac{\partial L}{\partial w_k} = \frac{2(w_k - d_k)}{d_k q_k} - \lambda(x_k') = 0 \qquad k = 1,\ldots,m$$

$$\frac{\partial L}{\partial w_k} = 2w_k - 2d_k = \lambda x_k' d_k q_k \qquad k = 1,\ldots,m$$

This minimization leads to the calibrated weight for nonresponse case,

$$w_k = d_k(1 + \lambda x_k' q_k) \qquad k = 1,\ldots,m \qquad (5.23)$$

$$\frac{\partial L}{\partial \lambda} = -\left(\sum_r w_k x_k - \sum_U x_k\right) = 0$$

$$\frac{\partial L}{\partial \lambda} = \sum_U x_k = \sum_r w_k x_k$$

$$\sum_U x_k = \sum_r (d_k + \lambda x_k' d_k q_k) x_k$$

$$\sum_U x_k = \sum_r d_k x_k + \sum_r \lambda x_k x_k' d_k q_k$$

$$\sum_U x_k - \sum_r d_k x_k = \lambda \sum_r x_k x_k' d_k q_k$$

$$\lambda = \frac{\sum_U x_k - \sum_r d_k x_k}{\sum_r x_k x_k' d_k q_k}$$

$$w_k = d_k + \frac{\sum_U x_k - \sum_r d_k x_k}{\sum_s x_k x_k^{'} d_k q_k} x_k^{'} d_k q_k$$

$$w_{kU} = d_k \left[1 + \left(\sum_U x_k - \sum_r d_k x_k \right) \left(\sum_s x_k x_k^{'} d_k q_k \right)^{-1} x_k^{'} q_k \right] \quad (5.24)$$

The *w* estimator form is $\hat{Y}_w = \sum_r w_k y_k$. If we know information at the populution level , we calculate the calibrated weight , w_k , like above. When we write this weight in the *w* estimator , we obtain calibration estimator at the population level $\sum_U x_k$.

And this estimator is

$$\hat{Y}_{wU} = \sum_r w_{kU} y_k \quad (5.25)$$

where $w_{kU} = d_k \mathrm{v}_{Uk}$

$$\nu_{Uk} = 1 + \left(\sum_U x_k - \sum_r d_k x_k \right) \left(\sum_s x_k x_k^{'} d_k q_k \right)^{-1} x_k^{'} q_k . \quad (5.26)$$

If the auxiliary total $\sum_U x_k$ is unknown, we can instead calibrate on the unbiased estimate $\sum_s d_k x_k$.

5.2.2.5 Deriving the Calibrated Weights when $\sum_U x_k$ is unknown

It can be also produced a calibration estimator for a survey in which the auxiliary vector values $\boldsymbol{x_k}$ are known up to the level of the sample s . We still know enough to form the sample-based HT estimator of the total, namely, $\sum_s d_k x_k$. For Info-S, minimization of the distance (5.15) under the constraint (5.13) leads to *w* estimator. Lagrange function is

$$L = \sum_r (w_k - d_k)^2 \Big/ d_k q_k - \lambda(\sum_r w_k x_k' - \sum_s d_k x_k') \tag{5.27}$$

$$\frac{\partial L}{\partial w_k} = \frac{2(w_k - d_k)}{d_k q_k} - \lambda(x_k') = 0 \qquad k = 1,\ldots,m$$

$$\frac{\partial L}{\partial w_k} = 2w_k - 2d_k = \lambda x_k' d_k q_k \qquad k = 1,\ldots,m$$

$$w_k = d_k (1 + \lambda x_k' q_k) \qquad k = 1,\ldots,m$$

$$\frac{\partial L}{\partial \lambda} = -\left(\sum_r w_k x_k - \sum_s d_k x_k \right) = 0$$

$$\frac{\partial L}{\partial \lambda} = \sum_s d_k x_k = \sum_r w_k x_k$$

$$\sum_s d_k x_k = \sum_r (d_k + \lambda x_k' d_k q_k) x_k$$

$$\sum_s d_k x_k = \sum_r d_k x_k + \sum_r \lambda x_k x_k' d_k q_k$$

$$\sum_s d_k x_k - \sum_r d_k x_k = \lambda \sum_r x_k x_k' d_k q_k$$

$$\lambda = \frac{\sum_s d_k x_k - \sum_r d_k x_k}{\sum_r x_k x_k' d_k q_k}$$

$$w_k = d_k + \frac{\sum_U x_k - \sum_r d_k x_k}{\sum_s x_k x_k' d_k q_k} x_k' d_k q_k$$

$$w_{ks} = d_k \left[1 + \left(\sum_s d_k x_k - \sum_r d_k x_k \right) \left(\sum_s x_k x_k' d_k q_k \right)^{-1} x_k' q_k \right] \tag{5.28}$$

And calibration estimator for information at the sample level , $\sum_s d_k x_k$,is

$$\hat{Y}_{ws} = \sum_r w_{ks} y_k \tag{5.29}$$

where $w_{ks} = d_k v_{sk}$

$$v_{sk} = 1 + \left(\sum_s d_k x_k - \sum_r d_k x_k \right) \left(\sum_s x_k x_k' d_k q_k \right)^{-1} x_k' q_k \,. \tag{5.30}$$

5.2.2.6 Calibration Estimators for Domains

When the survey has nonresponse, and reweighting is carried out by the calibration approach, then the estimation of the domain Y_d proceeds as follows:

A set of calibrated weights are given by (5.26), if the auxiliary information consists of the known vector total $\sum_U x_k$. They were used in (5.25) to produce an estimator of the whole population total Y . For the domain total Y_d , the same weights are kept and changed only the study variable from y into y_d .

The resulting calibration estimator of the domain total Y_d is therefore

$$\hat{Y}_{dw} = \sum_r w_k y_{dk} \tag{5.31}$$

where $w_k = d_k v_k$ and v_k is given by (5.26).

The domains of interest $U_1,...,U_d,...,U_D$ form a partition of U, as when the domains are regions making up a country in the some applicatios. The D domains estimates $\hat{Y}_{1w},...,\hat{Y}_{dw},...,\hat{Y}_{Dw}$ then have the appealing property that they add up to the calibration estimate made for the whole population, that is, $\hat{Y}_w$ given by (5.9). This property follows from

$$\sum_{d=1}^{D}\hat{Y}_{dw}=\sum_{d=1}^{D}\sum_{r}w_k y_{dk}=\sum_{r}w_k\sum_{d=1}^{D}y_{dk}=\sum_{r}w_k y_k=\hat{Y}_w .$$

Similarly, it can be adapted the calibration estimator (5.29), which has auxiliary information up to the sample level. The calibrated weights are as in (5.30). To obtain an estimator of the domain total , the weights are again preserved and substitute y_k for y_{dk} . The result is

$$\hat{Y}_{dws}=\sum_{r}d_k v_{sk} y_{dk} . \tag{5.32}$$

5.2.2.7 Variance Estimation Under The Calibration Approach

It is needed to estimate the variance of the calibration estimators for statements of precision and confidence intervals. Särndal and Lundström (1999) started by examining the mean squared error (MSE) of calibration estimator.

Jointly under the sampling design $p(s)$ and the response distribution $q(r|s)$, the mean squared error of $\hat{Y}_w$ is

$$MSE_{pq}\left(\hat{Y}_w\right)=V_{SAM}+V_{NR}+2Cov_p\left(\hat{Y}_s,B_{NR|s}\right)+E_p\left(B^2_{NR|s}\right) \tag{5.33}$$

where $V_{SAM}=V_p\left(\hat{Y}_s\right)$ is the sampling variance, $V_{NR}=E_pV_q\left(\hat{Y}_w|s\right)$ is the nonresponse error variance, $B_{NR|s}=E_q\left(\hat{Y}_w-\hat{Y}_w|s\right)$ is the nonresponse bias (conditionally on *s*), and $Cov_p\left(\hat{Y}_s,B_{NR|s}\right)$ is the covariance of $\hat{Y}_s$ and $B_{NR|s}$ under the sampling design.

Särndal and Lundström (1999) showed that if the condition

$$B_{NR|s}=0 \qquad \text{for all } s \tag{5.34}$$

is verified, then (5.33) becomes

$$V_{pq}\left(\hat{Y}_w\right) = V_{pq}^{0}\left(\hat{Y}_w\right)$$

where

$$V_{pq}^{0}\left(\hat{Y}_w\right) = V_{SAM} + V_{NR} . \tag{5.35}$$

There exists virtually no survey such that the condition (5.34) is exactly satisfied. Inevitably, whenever there is nonresponse, some bias is introduced. But calibration on strong auxiliary information may go a long way toward eliminating the conditional nonresponse bias $B_{NR|s}$. When this is the case, the inferences (e.g. confidence intervals and so on) made by acting as if (5.34) is true will still be approximately valid. It can been worked under the the assumption that (5.34) holds approximately, so that $V_{pq}\left(\hat{Y}_w\right) \approx V_{pq}^{0}\left(\hat{Y}_w\right) = V_{SAM} + V_{NR}$.

It is diffucult to obtain expressions for variances of point estimators $\hat{Y}_{ws}$ and $\hat{Y}_{wU}$ because of their complexity. So there exists a question that how variance esimates should be constructed. The aim is to obtain variance estimators for $\hat{Y}_{ws}$ and $\hat{Y}_{wU}$ that will work reasonably well and moreover, they sould be simple to calculate.

Finally, Särndal and Lundström(1999) imposed the condition that the variance estimators must be model free that they must be constructed without recourse to a nonresponse mechanism model. A strength of their approach is precisely the fact that point estimators are derived without appealing to such models.

Särndal and Lundström (1999) relied on analogy with the estimator (5.7) for two-phase sampling. An appropriate variance estimator for (5.7) is given by formula (9.7.22) in Särndal, Swensson and Wretman (1991).

To estimate the variance of $\hat{Y}_w$, Särndal and Lundström (1999) proposed to use formula (9.7.22) of Särndal, Swensson and Wretman (1991) as follows:

- ✓ replace $\pi_{k|s_a}$ by θ_k and then replace θ_k by $\hat{\theta}_k = 1/\nu_{sk}$, ν_{sk} is given by (5.26).
- ✓ it is assumed elements to respond independeently, so that , $\Pr(k \,\&\, s \in r|s) = \theta_{kl} = \theta_k \theta_l$ for

all $k \neq l$.

At the end of this, the variance estimator of $\hat{Y}_w$ for info-U is

$$\hat{V}(\hat{Y}_{wU}) = \hat{V}_{SAM} + \hat{V}_{NR} . \tag{5.36}$$

Explicit form of this varaince estimator is

$$\begin{aligned}\hat{V}(\hat{Y}_{wU}) = &\sum_r \sum_r (d_k d_l - d_{kl})(g_k \nu_{sk} e_k)(g_l \nu_{sl} e_l) - \sum_r d_k (d_k - 1)\nu_{sk}(\nu_{sk} - 1)(g_k e_k)^2 \\ &+ \sum_r d_k^2 \nu_{sk}(\nu_{sk} - 1) e_k^{\,2}\end{aligned} \tag{5.37}$$

where ν_{sk} is given by (5.30),

$$e_k = y_k - x_k' \hat{B}_\nu \tag{5.38}$$

$$\hat{B}_\nu = \left(\sum_r d_k \nu_{sk} q_k x_k x_k' \right)^{-1} \sum_r d_k \nu_{sk} q_k x_k y_k \tag{5.39}$$

and g_k is given by (3.16).

A variance estimator for $\hat{Y}_w$ for info-S is

$$\hat{V}(\hat{Y}_{ws}) = \hat{V}_{SAM} + \hat{V}_{NR} \tag{5.40}$$

Explicit form of this variance estimator is

$$\hat{V}\left(\hat{Y}_{ws}\right)=\sum_{r}\sum_{r}\left(d_k d_l-d_{kl}\right)\left(\nu_{sk}y_k\right)\left(\nu_{sl}y_l\right)-\sum_{r}d_k\left(d_k-1\right)\nu_{sk}\left(\nu_{sk}-1\right)\left(y_k\right)^2 \\ +\sum_{r}d_k^2\nu_{sk}\left(\nu_{sk}-1\right)e_k^{\,2} \qquad (5.41)$$

g_k and e_k is given, respectively by (3.12) and (5.34).

Also, the variance estimator can be written for domain estimators $\hat{Y}_{dw}$ given by (5.31) and $\hat{Y}_{dws}$ given by (5.32). An appropriate variance estiamtor for $\hat{Y}_{dw}$ follows easily by replacing y_k by y_{dk} throughout the calculations defined by (5.36) to (5.39). That is, in (5.37) it must be replaced e_k by

$$e_{dk}=y_{dk}-x_k'\hat{B}_{d\nu} \qquad (5.42)$$

where

$$\hat{B}_{d\nu}=\left(\sum_{r}d_k\nu_{sk}q_kx_kx_k'\right)^{-1}\sum_{r}d_k\nu_{sk}q_kx_ky_{dk} \qquad (5.43)$$

For large response sets, the nonresponse bias of $\hat{Y}_w$ is

$$B_{pq}\left(\hat{Y}_w\right)\approx-\sum_{U}\left(1-\theta_k\right)E_k^{\theta} \qquad (5.44)$$

where $E_k^{\theta}=y_k-x_k'\,B_U^{\theta}$ and $B_U^{\theta}=\left(\sum\theta_kq_kx_kx_k'\right)^{-1}\sum_{U}q_kx_ky_k$

5.2.2.8 Examples of calibration estimators

Many methodologists are accustomed to specific formulas corresponding to particular methods for nonresponse reweighting. Therefore, the objective is to show that the calibration approach reproduces formulas that many readers are familiar with. Thus, the explicit form of (5.25) and (5.29) are derived by Särndal and Lundström (1999) for some simple specifications of the auxiliary information and showed that commonly used estimators are obtained. They started with the simplest forms of $\boldsymbol{x}_k$, then gradually increased the auxiliary information content and thereby also the complexity of the formulas.

<u>*The Simplest Auxiliary vector:*</u> The simplest formulation of the auxiliary vector is $\boldsymbol{x}_k=1$ for all k. This vector recognises no differences among elements. Specifying also $q_k=1$ for all k. Then (5.26) gives the weight for all k,

$$v_k = \frac{n}{m} \tag{5.45}$$

and the calibration estimator (5.25) becomes

$$\hat{Y}_w = \frac{N}{m}\sum_r y_k = \hat{Y}_{EXP} \tag{5.46}$$

As seen from that, when auxiliary vector is $\boldsymbol{x}_k=1$, the calibration estimator becomes the traditional expansion estimator. For this estimator bias expression for large response sets, Särndal and Lundström(1999) found that

$$B_{pq} = \left(\hat{Y}_{EXP}\right) \approx N\left(\frac{\sum_U \theta_k y_k}{\sum_U \theta_k} - \bar{Y}\right) \tag{5.47}$$

where $\bar{Y} = \frac{\sum_U y_k}{N}$. And the variance estimator

$$\hat{V}\left(\hat{Y}_{EXP}\right) = N^2\left(1-\frac{m}{N}\right)\frac{S_r^2}{m}. \tag{5.48}$$

<u>One-way classification</u>*:* In this case the target population is divided into non-overlapping and exhaustive groups, U_p , *p = 1,…,P* , based on a specified classification criterion, for example age by sex groups. The auxiliary vector form is $\boldsymbol{x_k} = \left(\gamma_{1k},...,\gamma_{pk},...,\gamma_{Pk}\right)'$ where

$$\gamma_{pk} = \begin{cases} 1 & if \quad k \in U_p \\ 0 & otherwise \end{cases} \tag{5.49}$$

The component of the key vector totals are denoted as follows:

$$\sum_U \boldsymbol{x_k} = \left(N_1,...,N_p,...,N_P\right)'$$

and

$$\sum_s \boldsymbol{x_k} = \left(n_1,...,n_p,...,n_P\right)'$$

and

$$\sum_r x_k = \left(m_1,...,m_p,...,m_P\right)'.$$

When $q_k = 1$ for all *k*, the weight is obtained from (5.22)

$$v_k = \frac{N_p n}{N m_p} \tag{5.50}$$

for $k \in r_p$, and the calibration estimator (5.25) becomes

$$\hat{Y}_w = \sum_{p=1}^{P} N_p \bar{y}_{r_p} = \hat{Y}_{PST} \tag{5.51}$$

where $\bar{y}_{r_p} = \frac{1}{m_p}\sum_{r_p} y_k$ and m_p is the number of respondents in group p. This estimator commonly called the poststratified estimator and denotes as $\hat{Y}_{PST}$.

When knowledge of the auxiliary vactor $\boldsymbol{x_k} = (\gamma_{1k},...,\gamma_{pk},...,\gamma_{Pk})'$ is limited to the elements of the sample s, the calibtarion estimator (5.29) becomes

$$\hat{Y}_{ws} = \sum_{p=1}^{P} \hat{N}_p \bar{y}_{r_p} = \hat{Y}_{WCL} \tag{5.52}$$

with $\hat{N}_p = \frac{N}{n} n_p$ and n_p is the number of sampled elements in group p. This estimator known as the weighting class estimator dented as $\hat{Y}_{WCE}$ .Under the this specifications bias expression is as follows:

$$B_{pq}(\hat{Y}_{WCL}) = B_{pq}(\hat{Y}_{PST}) \approx \sum_{p,=1}^{P} N_p \left(\frac{\sum_{U_p} \theta_k y_k}{\sum_{U_p} \theta_k} - \bar{Y}_p \right) \tag{5.53}$$

where $\bar{Y}_p = \frac{\sum_{U_p} y_k}{N_p}$. An the variance estimator is ,

$$\hat{V}(\hat{Y}_{PST}) = \sum_{p=1}^{P} N_p{}^2 \left(1 - \frac{m_p}{N_p}\right) \frac{S_r^2}{m_p} . \tag{5.54}$$

<u>*A single quantitative variable*</u> *:* It is assumed that a quantitative auxiliary variable $\boldsymbol{x_k}$ is available. For example, the number of employees of enterprise k in a business

survey, k = 1,... N. It is assumed that its population total, $\sum_{u} x_k$, known. If this is the only auxiliary variable, the auxiliary vector is uni-dimensional, $\boldsymbol{x_k} = x_k$. When q_k is specified as $q_k = x_k^{-1}$, then the calibration estimator (5.9) is

$$\hat{Y}_w = \left(\sum_{U} x_k\right)\frac{\bar{y}_r}{\bar{x}_r} = \hat{Y}_{RA} \tag{5.55}$$

where $\bar{y}_r = \frac{1}{m}\sum_{r} y_k$ and $\bar{x}_r = \frac{1}{m}\sum_{r} x_k$. This estimator has the well known form of a ratio estimator and denoted as $\hat{Y}_{RA}$.

With the same information it can be alternatively formulated the auxiliary vector as $\boldsymbol{x_k} = (1, x_k)'$. When q_k =1 for all k, the calibration estimator (5.29) becomes,

$$\hat{Y}_w = N\left\{\bar{y}_r + \left(\bar{X} - \bar{x}_r\right)\hat{B}\right\} = \hat{Y}_{REG} \tag{5.56}$$

where $\bar{X} = \frac{1}{N}\sum_{U} x_k$ and $\hat{B} = \left[\sum_{r} y_k x_k - \frac{1}{m}\sum_{r} y_k\right] / \left[\sum_{r} x_k^2 - \frac{1}{m}\left(\sum_{r} x_k\right)^2\right]$. This estimator has the well known form of a regression estimator and denoted as $\hat{Y}_{REG}$.

One –way classification and a quantitative variable: The auxiliary information concerns a P-valued categorical variable and a quantitative variable, x, that may be an indicator of the size of an element. It is assumed that every sampled elemet k is placed into the appropriate group, that its value x_k is known. And fore each group the size N_p is known and the x total, $\sum_{U_p} y_k$. There are more than one way to use this information. The auxiliary vector is defined as

$$\boldsymbol{x_k} = \left(\gamma_{1k} x_k, ..., \gamma_{pk} x_k, ..., \gamma_{Pk} x_k\right)' \tag{5.57}$$

where γ_{pk} is defined by (5.49). It leads to a well known estimator, because if $q_k = x_k^{-1}$ is let (5.29) becomes

$$\hat{Y}_w = \sum_{p=1}^{P}\left(\sum_{U_p} x_k\right)\frac{\overline{y}_{r_p}}{\overline{x}_{r_p}} = \hat{Y}_{SEPRA} \tag{5.58}$$

Where $\overline{y}_{r_p} = \frac{1}{m_p}\sum_{r_p} y_k$ and $\overline{x}_{r_p} = \frac{1}{m_p}\sum_{r_p} x_k$ .Thus, $\hat{Y}_{SEPRA}$ has the form of a separate ratio estimator.

Instead, the auxiliary vector is formulated as

$$\boldsymbol{x_k} = \left(\gamma_{1k},...,\gamma_{pk},...,\gamma_{Pk},\gamma_{1k}x_k,...,\gamma_{pk}x_k,...,\gamma_{Pk}x_k\right)' \tag{5.59}$$

Then if $q_k = 1$ for all k, the estimator (5.29) becomes

$$\hat{Y}_w = \sum_{p=1}^{P} N_p\left\{\overline{y}_{r_p} + \left(\overline{X}_p - \overline{x}_{r_p}\right)\hat{B}_p\right\} = \hat{Y}_{SEPREG} \tag{5.60}$$

with $\overline{X}_p = \frac{1}{N_p}\sum_{U_p} x_k$ and $\hat{B}_p = \frac{Cov_{xyr_p}}{S^2_{xr_p}}$. Covariance and variance term is

$$Cov_{xyr_p} = \frac{1}{m_p - 1}\left[\sum_{r_p} y_k x_k - \frac{1}{m_p}\sum_{r_p} y_k \sum_{r_p} x_k\right]$$

and

$$S^2_{xr_p} = \frac{1}{m_p - 1}\left[\sum_{r_p} x_k^2 - \frac{1}{m_p}\left(\sum_{r_p} x_k\right)^2\right].$$

The estimator (5.60) is another well known form separate regression estimator.

5.2.2.9 Analysis of the nonresponse bias for examples of calibration estimator

The nonresponse bias given by (5.40) will be zero if a certain relation exists between the response probability θ_k and the auxiliary vector $\boldsymbol{x_k}$, namely,

$$\theta_k^{-1} = 1 + q_k \lambda' x_k \qquad \text{for} \quad k \in U \tag{5.61}$$

where λ is a column vector independent of *k*. If (5.61) holds, the right hand side of (5.60) is zero, and the bias $B_{pq}(\hat{Y}_w)$ is thus approximately zero. When the ideal conditions is satisfied , nonresponse bias is approximately zero, $B_{pq}(\hat{Y}_w) \approx 0$, because $E_{\theta k} = 0$ for all k . For many sampling designs, a reduction of the residuals will reduce the variance. Consequently, an auxiliary vector that explains the variation of the study variable is effective in reducing MSE. (Särndal & Lundström, 2001)

Several special cases of the general calibtration estimator was obtained, corresponding to different formulations of the auxiliary vector $\boldsymbol{x_k}$ and the factor q_k given Table 5.1. Särndal and Lundström (1999) revisited these estimators with the purpose of showing how the theoretical result (5.44) can guide the selection of relevant auxiliary information.

Table. 5.1. The specifications of the auxiliary vector $\boldsymbol{x_k}$ and the factor q_k leading to well known estimators (Lundström & Särndal, 2001)

Estimator	Auxiliary vector $\boldsymbol{x_k}$	Factor q_k
EXP	1	1
PST and WCE	$(\gamma_{1k},\ldots,\gamma_{pk},\ldots,\gamma_{Pk})'$	1
RA	$\boldsymbol{x_k}$	x_k^{-1}
REG	$(1,x_k)'$	1
SEPRA	$(\gamma_{1k}x_k,\ldots,\gamma_{pk}x_k,\ldots,\gamma_{Pk}x_k)'$	x_k^{-1}
SEPREG	$(\gamma_{1k},\ldots,\gamma_{pk},\ldots,\gamma_{Pk},\gamma_{1k}x_k,\ldots,\gamma_{pk}x_k,\ldots,\gamma_{Pk}x_k)'$	1

In the section (4.5.1), three principles are explained that the auxiliary vector 's satisfied. For the first principle, each of the six estimators in Table 5.1, the specifications of $\boldsymbol{x_k}$ and q_k into (5.61) and the results are obtained in Table 5.2. It is assumed that the value of θ_k is known for every k and the set of N points $\{(\theta_k^{-1},u_k): k=1,\ldots,N\}$ is examined in orded to see how closely θ_k^{-1} agrees with u_k. And u_k is defined as $u_k = 1+q_k\lambda' x_k$. If the relationship is perfect, so that u_k equals θ_k^{-1} for every k. then the nonresponse bias is totally eliminated. This perfect relationship is stated in the Table 5.2.

Table 5.2. The relationship between θ_k^{-1} and u_k needed to elminate the nonresponse bias for six well-known estimators. a, a_p, b, b_p denote constants. (Lundström & Särndal, 2001)

Estimator	Form of the θ_k^{-1} needed to eliminate bias	Description of the θ_k^{-1} needed to eliminate bias
EXP	$\theta_k^{-1} = a$ for all $k \in U$	constant throughout
PST and WCE	$\theta_k^{-1} = a_p$ for all $k \in U_p$	constant within groups
RA	$\theta_k^{-1} = a$ for all $k \in U$	constant throughout
REG	$\theta_k^{-1} = a + bx_k$	linear in $\boldsymbol{x}_k$
SEPRA	$\theta_k^{-1} = a_p$ for all $k \in U_p$	constant within groups
SEPREG	$\theta_k^{-1} = a_p + b_p x_k$	linear in $\boldsymbol{x}_k$ within groups

The six example of calibration estimators in Table 5.1 and Table 5.2 illustrate the following important principles the more succeed in incorporating important auixuliury information into the auxiliary vector the better are the chances that the nonresponse bias will be reduced to near zero levels.(Lundström & Särndal, 2001)

Table 5.3. The linearship between y_k and the auxiliary vector that eliminates the nonresponse bias for six well-known estimators. α, β, β_p denote constants. (Lundström & Särndal, 2001)

Estimator	Form of the y_k needed to eliminate bias	Description of the y_k needed to eliminate bias
EXP	$y_k = \alpha$ for all $k \in U$	constant throughout
PST and WCE	$y_k = \beta_p$ for all $k \in U_p$	constant within groups
RA	$y_k = \alpha x_k$ for all $k \in U$	linear in $\boldsymbol{x_k}$ through the origin within groups
REG	$y_k = \alpha + \beta x_k$	linear in $\boldsymbol{x_k}$
SEPRA	$y_k = \alpha_p x_k$ for all $k \in U_p$	linear in $\boldsymbol{x_k}$ through the origin within groups
SEPREG	$y_k = \alpha_p + \beta_p x_k$ for all $k \in U_p$	linear in $\boldsymbol{x_k}$ within groups

For the second principle in section (4.5.1) , the auxiliary vector should explain the variation of the most important study variables. If there is perfect relationship , $y_k = \boldsymbol{x}_k' \boldsymbol{\beta}$, holds for all $k \in U$, then $\hat{\boldsymbol{Y}}_w = \boldsymbol{Y}$. All population residuals E_k and the nonresponse bias are then zero. If $y_k = \boldsymbol{x}_k' \boldsymbol{\beta}$ is taken as a starting point for analysis, the six estimator is likely satisfy the pirinciple two.

It is assumed that, it could be examined the N points $\{(y_k, y^0{}_k) : k = 1,...,N\}$, where $y_k = \boldsymbol{x}_k' \boldsymbol{\beta}$. If the relationship is perfect, so that y_k^0 equals y_k for every *k*, then the nonresponse bias is totally eliminated. This perfect relationship is illustruated in Table 5.3.

CHAPTER SIX

APPLICATION

6.1 Introduction

In the previous chapter, the definition of calibration approach to reweighting under nonresponse problem was given. In the presence of available auxiliary information the calibration approach is flexible for the reduction of nonresponse. In this chapter, it is given through the simulation that how alternative specifications of the auxiliary vector $\mathbf{x_k}$ affect the quality of estimators derived by the calibration technique.

At the beginning of the application, a population in the size of N = 1000 was generated by the Minitab macro program. Then, 100 random samples were drawn from this population with sample size of n = 400. Throughout the application section the only one type of parameter estimated is a total for the entire population. The SRS design is used for drawing samples from population. In order to both computation of point estimators and quality measures that relative bias and variance for the population total, generated by different $\mathbf{x_k}$ vector specifications, a macro was written using C++ programming language.

6.2 Application

The population was generated by the Minitab macro program with the size of N = 1000. The study variable y_k is a numerical variable measuring such as expense. The first auxiliary variable, $\mathbf{\Gamma}_{k,}$ is categorical, indicating one out of four possible regions. The second auxiliary variable, denoted x_k, is numerical too and defined such as revenues. In the simulation the square root of x_k was used as second auxiliary variable. Some of variables of this population is given Table 6.1.

Table 6.1 Generated Population

	y_k	x_k	square root of x_k	Γ_k
1	1556	1598	40	4
2	1986	1773	42	2
3	1332	715	27	4
4	1128	1610	40	4
5	1584	441	21	3
6	1555	1566	40	4
7	1572	1295	36	1
8	1035	466	22	4
9	1773	1877	43	3
10	1877	1012	32	3
11	2208	1741	42	3
12	894	1541	39	2
13	1341	1435	38	1
14	1807	1419	38	3
15	1544	1550	39	1
16	1063	1109	33	4
17	1397	1219	35	1
18	1461	1702	41	1
19	1212	1627	40	1
20	1419	1414	38	3
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
981	1333	1519	40	4
982	1862	1982	39	4
983	1306	1446	45	4
984	1413	1356	38	2
985	1672	1394	37	2
986	1516	1365	37	1
987	1251	1398	37	2
988	1759	1583	37	1
989	1526	1673	40	1
990	1283	1699	41	4
991	1373	1641	41	1
992	829	1789	41	4
993	1469	1560	42	4
994	1331	1298	39	3
995	1245	1888	36	2
996	1696	1402	37	2
997	1563	1600	40	3
998	1438	1745	42	4
990	1654	1835	43	4
1000	1282	1858	43	1

It is assumed that 850 of the 1000 data of population was responded. So the population used in the simulation consist of 850 responding elements. The response rate was thus 85 per cent. From the population consisting of the 850 responding elements , repeated simple random samples are drawn. 100 random samples with the size of n = 400 was selected by the following Minitab macro program 'SAMPLING.MTW'.

<u>SAMPLING.MTW</u>

```
SAMPLE 400 C1 CK1
LET K1=K1+1
END
```

The macro was completed when the Minitab macro command was written as follows:

```
MTB > LET K1=2
MTB > EXEC 'C: \MTBWIN\DATA\SAMPLING.MTW' 100
```

Table 6.2 Some key characteristics of the study variable *y*.

Characteristics for the entire population		Group (Regions)			
		1	2	3	4
Total (Y)	1,282,671	353,209	306,795	305,314	317,353
Mean ($\bar{Y}$)	1,509	1,542	1,511	1,511	1,469
Number of elements (N)	850	229	203	202	216

In the chapter five, examples of calibration estimators obtained by the specifications of auxiliary vectors for the population total were given. These point estimators were computed through the C ++ programming language. For response set j *(j=1,..., 100)*, $\hat{Y}_{w_{(j)}}$ is calculated, where $\hat{Y}_{w_{(j)}}$ is the value of the point estimator $\hat{Y}_w$

for response set *j*. Here, $\hat{Y}_w = \hat{Y}_{ws}$ given by (5.29) for Info-S, and $\hat{Y}_w = \hat{Y}_{wU}$ given by (5.25) for Info-U.

Table 6.3 Results of the point estimators with different auxiliary vectors, SRS n = 400

Auxiliary vector	Info Level	Estimators	Results
1	s, U	$\hat{Y}_{EXP}$	See Appendix-2
$(\gamma_{1k},\dots,\gamma_{pk},\dots,\gamma_{Pk})'$	U	$\hat{Y}_{PST}$	See Appendix-3
	s	$\hat{Y}_{WCE}$	See Appendix-4
x_k	U	$\hat{Y}_{RA}$	See Appendix-5
	s		See Appendix-6
$(1, x_k)'$	U	$\hat{Y}_{REG}$	See Appendix-7
	s		See Appendix-8
$(\gamma_{1k}x_k,\dots,\gamma_{pk}x_k,\dots,\gamma_{Pk}x_k)'$	U	$\hat{Y}_{SEPRA}$	See Appendix-9
	s		See Appendix-10
$(\gamma_{1k},\dots,\gamma_{pk},\dots,\gamma_{Pk},\gamma_{1k}x_k,\dots,\gamma_{pk}x_k,\dots,\gamma_{Pk}x_k)'$	U	$\hat{Y}_{SEPREG}$	See Appendix-11
	s		See Appendix-12

And two quality measures were studied for see the affects of auxiliary vectors on estimators. The first is simulation relative bias in per cent,

$$RB_{SIM}\left(\hat{Y}_w\right) = 100\frac{\left\lfloor E_{SIM}\left(\hat{Y}_w\right) - Y\right\rfloor}{Y} \tag{6.1}$$

where $E_{SIM}\left(\hat{Y}_w\right)$ is the simulation expectation of $\hat{Y}_w$ denote as,

$$E_{SIM}\left(\hat{Y}_w\right) = \sum_{j=1}^{100}\frac{\hat{Y}_{w(j)}}{100} \tag{6.2}$$

and the second is simulation variance,

$$V_{SIM}\left(\hat{Y}_w\right) = \sum_{j=1}^{100}\frac{\left[\hat{Y}_{w(j)} - E_{SIM}\left(\hat{Y}_w\right)\right]}{99} \tag{6.3}$$

The simulation results are given in Table 6.4.

Table 6.4 Simulation relative bias and simulation variance for different point estimators, SRS with n = 400

Auxiliary vector	Info Level	Estimators	$RB_{SIM}\left(\hat{Y}_w\right)$	$V_{SIM}\left(\hat{Y}_w\right)x10^6$
1	U	$\hat{Y}_{EXP}$	6.96	9.105
$\left(\gamma_{1k},\ldots,\gamma_{pk},\ldots,\gamma_{Pk}\right)'$	U	$\hat{Y}_{PST}$	0.13	6.346
x_k	U	$\hat{Y}_{RA}$	1.57	1,614
$\left(1, x_k\right)'$	U	$\hat{Y}_{REG}$	0.09	2,492
$\left(\gamma_{1k}x_k,\ldots,\gamma_{pk}x_k,\ldots,\gamma_{Pk}x_k\right)'$	U	$\hat{Y}_{SEPRA}$	0.10	1.372
$\left(\gamma_{1k},\ldots,\gamma_{pk},\ldots,\gamma_{Pk},\gamma_{1k}x_k,\ldots,\gamma_{pk}x_k,\ldots,\gamma_{Pk}x_k\right)'$	U	$\hat{Y}_{SEPREG}$	0.11	1.342

In the simulations, we studied the point estimators generated by different $\mathbf{x}_k$ vector specifications, namely, $\mathbf{x}_k = 1$ for all k, $\mathbf{x}_k = (\gamma_{1k},\dots,\gamma_{pk},\dots,\gamma_{Pk})'$, $\mathbf{x}_k = x_k$, $\mathbf{x}_k = (1, x_k)'$, $\mathbf{x}_k = (\gamma_{1k} x_k,\dots,\gamma_{pk} x_k,\dots,\gamma_{Pk} x_k)'$, $\mathbf{x}_k = (\gamma_{1k},\dots,\gamma_{pk},\dots,\gamma_{Pk},\gamma_{1k} x_k,\dots,\gamma_{pk} x_k,\dots,\gamma_{Pk} x_k)'$ When $\mathbf{x}_k = 1$ there is no auxiliary information. In this reason simulation results for estimator ($\hat{Y}_{EXP}$) has more bias and variance. We compared the other estimators ($\hat{Y}_{PST}, \hat{Y}_{RA}, \hat{Y}_{REG}, \hat{Y}_{SEPRA}, \hat{Y}_{SEPREG}$), which generated by different $\mathbf{x}_k$ vector specifications, with $\mathbf{x}_k = 1$. We would expect the nonresponse bias to diminish with increasing amounts of auxiliary information, and this is confirmed by Table 6.4.

CHAPTER SEVEN
CONCLUSION

The aim of this thesis is to introduce methods for the reduction of bias and errors arising from survey with nonresponse. In the recent years, scientist became increased to concern with the calibration approach to reweigthing method in the presence of nonresponse. Because of this reason, we interested in especially reweighting method with calibration approach. Calibration procedure requires the formulation of a suitable auxiliary vector, through a selection from a possible larger set of auxiliary variables. Different specifications of auxiliary variables leads to different form of calibration estimators.

In application, we studied the point estimators generated by different $\mathbf{x}_k$ vector specifications, namely, $\mathbf{x}_k = 1$ for all k, $\mathbf{x}_k = (\gamma_{1k},...,\gamma_{pk},...,\gamma_{Pk})'$, $\mathbf{x}_k = x_k$, $\mathbf{x}_k = (1, x_k)'$, $\mathbf{x}_k = (\gamma_{1k}x_k,...,\gamma_{pk}x_k,...,\gamma_{Pk}x_k)'$, $\mathbf{x}_k = (\gamma_{1k},...,\gamma_{pk},...,\gamma_{Pk},\gamma_{1k}x_k,...,\gamma_{pk}x_k,...,\gamma_{Pk}x_k)'$ Application shows that the calibration approach is highly flexible in its use of auxiliary information for the reduction of nonresponse bias. When $\mathbf{x}_k = 1$ there is no auxiliary information. In this reason simulation results for estimator ($\hat{Y}_{EXP}$) has more bias and variance. We compared the other estimators ($\hat{Y}_{PST}$, $\hat{Y}_{RA}$, $\hat{Y}_{REG}$, $\hat{Y}_{SEPRA}$, $\hat{Y}_{SEPREG}$), which generated by different $\mathbf{x}_k$ vector specifications, with $\mathbf{x}_k = 1$. We would expect the nonresponse bias to diminish with increasing amounts of auxiliary information, and this is confirmed by Table 6.4.

For instance, as shown in Table 6.4 when auxiliary information are not used its calculated that the relative bias of estimator is [$RB_{SIM}(\hat{Y}_{EXP}) = 6.96$ %] and variance is [$V_{SIM}(\hat{Y}_{EXP}) = 9{,}105 \times 10^6$]. When auxiliary vector , $\mathbf{x}_k = (\gamma_{1k},...,\gamma_{pk},...,\gamma_{Pk})'$, is used in the estimation stage then the relative bias is obtained [$RB_{SIM}(\hat{Y}_{PST})$ =0.13%] and variance is [$V_{SIM}(\hat{Y}_{PST})$ =$6,346x10^6$]. This denotes that using the auxiliary information decreases the nonresponse bias and variance.

Calibration approach may play considerable role in reduction of nonresponse bias and it is better suited for a routine treatment of nonresponse in organization.

REFERENCES

Alavi, A., & Beamont, F.J. (2003) Evaluation and adjustment for nonresponse in the Canadian Labour Force Survey. *Statistics Canada International Symposium Series-Proceedings.*

Bethlehem, J.G. (1988) Reduction of nonresponse bias through regression estimation. Journal of Official Statistics,4 (3) 251-260.

Deville, J.C., & Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association* 87, 376-382.

Ekholm, A., & Laaksonen, S. (1991). Weighting via response modeling in the Finnish Household Budget Survey. *Journal of Official Statistics*,11, 181-204.

Hansen, M.H., Hurwitz, W.N. (1946). The problem of nonresponse in sample surveys. Journal of American Statistical Association, 41, 517-529.

Kish, L. (1969) *Survey Sampling*. New York, Wiley.

Kalton, G. and Cervantes, I.F. (2003). Weighting methods. Journal of Official Statistics, 19 (2), 81-97.

Lundström, S., and Särndal, C. E. (1999). Calibration as a standard method for treatment of nonresponse. Journal of Official Statistics, 15, 305-327.

Lundström, S. (1997). Calibration as a standard method for treatment of nonresponse. Doctoral dissertation, Stockholm University.

Lundström, S., and Särndal, C. E. (2001). Estimation in the presence of nonresponse and frame imperfections. Printed in Sweden, SCB-tyrck,Örebro.

Lohr, S. L. (1999). *Sampling: Design and Analysis*, Duxbury Press

Särndal, C. E., Estevao, V. M. (2000). A functional form to calibration. Journal of Official Statisitics, 16 (4), 379-399.

Särndal, C. E.,Swensson, B.(1987) A general view of estimation for two-phases of selection with applications to two-phase sampling and nonresponse. *International Statistical Review*., 55 (3), 279-294.

Särndal, C. E.,Swensson, B., Wretman, J. (1991). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Son, C. K., Jung, Y. M. (2004). The calibration of Horvitz-Thompson variance estimator under the unit nonresponse. *Proceedings of the Spring Conference*, Korean Statistical Society.

Teksoy, N. (1991). *Cevaplamama hatalarının ve ikamelerin tahminler üzerindeki etkileri*. DİE Uzmanlık Tezi.

Théberge, A. (1999). Extensions of calibration estimators in survey sampling. *Journal of the American Statistical Association*, 94 (446) 635-644.

Wu, C.(1999). The effective Use of Complete Auxiliary Information from survey data. Doctoral dissertation, Simon Fraser University.

Wu, C. and Sitter, R.R. (2001). A model-calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association*, 96, 453.

Wu, C. and Luan, Y. (2003). Optimal calibration estimators under the two-phase sampling. Journal of Official Statistics,19 (2),119-131.

Wu, C. (2003). *Optimal calibration estimators in survey sampling*. *Biometrika* , 90 (4), 937-951.

Appendix -1 C++ Program for calculation of estimators, biases and variances for 100 samples.

```
#include<stdio.h>
#include<stdlib.h>
#include<math.h>
main(){

float yk[1000],xk[1000],blg[1000];
float y_val[100][400],x_val[100][400],blg_val[100][400];
float var_yk, var_xk,bolge;
float cov1[100],cov2[100],cov3[100],cov4[100];
float var1[100],var2[100],var3[100],var4[100];
float yrp1_ort[100],yrp2_ort[100],yrp3_ort[100],yrp4_ort[100];
float xrp1_ort[100],xrp2_ort[100],xrp3_ort[100],xrp4_ort[100];
float RBsim_Y[11],Esim_Y[11], VarSim_Y[11];
float y_pst[100];
float Yexp[100];
float Y_wce[100];
float Y_rau[100],xr_ort[100],yr_ort[100],Y_ras[100];
float Y_regu[100],Beta_sapka[100],Y_regs[100];
float yk_carp_xk_top[100], xk_kare_top[100];
float top_x[100],top_y[100];
float x_blgorn_top[100][4];
float Y_seprau[100], Y_sepras[100];
float Y_sepregu[100], Y_sepregs[100];
float Beta_sapka1[100],Beta_sapka2[100],Beta_sapka3[100],Beta_sapka4[100];
int m[100][4],M[100];
int n[100][4];
int Np_sapka[100][4];
int toplam_y,toplam_x;
int i=0;
FILE *cfPtr,*cfPtr1,*cfPtr2;

/*******************************************************************************************/
/************************READING DATA FROM POPULATION FILE*******************/
/******************************************************************************************/

if((cfPtr=fopen("b1000_kitle. dat","r"))==NULL)
        printf("Dosyada veri yok\n");
else        {
  fscanf(cfPtr,"%f%f%f",&var_yk,&var_xk,&bolge);
  while(var_yk!=float(-1))     {
        //printf("%.2f %.2f %.2f\n",var_yk,var_xk,bolge);
    yk[i]=var_yk;   xk[i]=var_xk;   blg[i]=bolge;
        fscanf(cfPtr,"%f%f%f",&var_yk,&var_xk,&bolge);
        i=i+1;

  }
  fclose(cfPtr);
```

Appendix -1 Continued

```
   //printf("\nyk[0]=%f\n blg[999]=%f,\nblg[998]=%f",yk[0],blg[999],blg[998]);
}

/*********************************************************************************************/
/************************* READING DATA FROM SAMPLE FILE ***********************/
/*********************************************************************************************/

if((cfPtr1=fopen("ilk_50.dat","r"))==NULL)
        printf("Dosyada veri yok\n");
else        {
        i=0;

fscanf(cfPtr1,"%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f
%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f
%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f
%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f
%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f",
   &blg_val[0][i],&y_val[0][i],&x_val[0][i],&blg_val[1][i],&y_val[1][i],&x_val[1][i],
   &blg_val[2][i],&y_val[2][i],&x_val[2][i],&blg_val[3][i],&y_val[3][i],&x_val[3][i],
   &blg_val[4][i],&y_val[4][i],&x_val[4][i],&blg_val[5][i],&y_val[5][i],&x_val[5][i],
   &blg_val[6][i],&y_val[6][i],&x_val[6][i],&blg_val[7][i],&y_val[7][i],&x_val[7][i],
   &blg_val[8][i],&y_val[8][i],&x_val[8][i],&blg_val[9][i],&y_val[9][i],&x_val[9][i],
   &blg_val[10][i],&y_val[10][i],&x_val[10][i],&blg_val[11][i],&y_val[11][i],&x_val[11][i],
   &blg_val[12][i],&y_val[12][i],&x_val[12][i],&blg_val[13][i],&y_val[13][i],&x_val[13][i],
   &blg_val[14][i],&y_val[14][i],&x_val[14][i],&blg_val[15][i],&y_val[15][i],&x_val[15][i],
   &blg_val[16][i],&y_val[16][i],&x_val[16][i],&blg_val[17][i],&y_val[17][i],&x_val[17][i],
   &blg_val[18][i],&y_val[18][i],&x_val[18][i],&blg_val[19][i],&y_val[19][i],&x_val[19][i],
   &blg_val[20][i],&y_val[20][i],&x_val[20][i],&blg_val[21][i],&y_val[21][i],&x_val[21][i],
   &blg_val[22][i],&y_val[22][i],&x_val[22][i],&blg_val[23][i],&y_val[23][i],&x_val[23][i],
   &blg_val[24][i],&y_val[24][i],&x_val[24][i],&blg_val[25][i],&y_val[25][i],&x_val[25][i],
   &blg_val[26][i],&y_val[26][i],&x_val[26][i],&blg_val[27][i],&y_val[27][i],&x_val[27][i],
   &blg_val[28][i],&y_val[28][i],&x_val[28][i],&blg_val[29][i],&y_val[29][i],&x_val[29][i],
   &blg_val[30][i],&y_val[30][i],&x_val[30][i],&blg_val[31][i],&y_val[31][i],&x_val[31][i],
   &blg_val[32][i],&y_val[32][i],&x_val[32][i],&blg_val[33][i],&y_val[33][i],&x_val[33][i],
   &blg_val[34][i],&y_val[34][i],&x_val[34][i],&blg_val[35][i],&y_val[35][i],&x_val[35][i],
   &blg_val[36][i],&y_val[36][i],&x_val[36][i],&blg_val[37][i],&y_val[37][i],&x_val[37][i],
   &blg_val[38][i],&y_val[38][i],&x_val[38][i],&blg_val[39][i],&y_val[39][i],&x_val[39][i],
   &blg_val[40][i],&y_val[40][i],&x_val[40][i],&blg_val[41][i],&y_val[41][i],&x_val[41][i],
   &blg_val[42][i],&y_val[42][i],&x_val[42][i],&blg_val[43][i],&y_val[43][i],&x_val[43][i],
   &blg_val[44][i],&y_val[44][i],&x_val[44][i],&blg_val[45][i],&y_val[45][i],&x_val[45][i],
   &blg_val[46][i],&y_val[46][i],&x_val[46][i],&blg_val[47][i],&y_val[47][i],&x_val[47][i],
   &blg_val[48][i],&y_val[48][i],&x_val[48][i],&blg_val[49][i],&y_val[49][i],&x_val[49][i]);
   while(i!=399)
     {
     i=i+1;

fscanf(cfPtr1,"%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f
%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f
%f
```

Appendix -1 Continued

```
%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f
%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f
%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f",
  &blg_val[0][i],&y_val[0][i],&x_val[0][i],&blg_val[1][i],&y_val[1][i],&x_val[1][i],
  &blg_val[2][i],&y_val[2][i],&x_val[2][i],&blg_val[3][i],&y_val[3][i],&x_val[3][i],
  &blg_val[4][i],&y_val[4][i],&x_val[4][i],&blg_val[5][i],&y_val[5][i],&x_val[5][i],
  &blg_val[6][i],&y_val[6][i],&x_val[6][i],&blg_val[7][i],&y_val[7][i],&x_val[7][i],
  &blg_val[8][i],&y_val[8][i],&x_val[8][i],&blg_val[9][i],&y_val[9][i],&x_val[9][i],
  &blg_val[10][i],&y_val[10][i],&x_val[10][i],&blg_val[11][i],&y_val[11][i],&x_val[11][i],
  &blg_val[12][i],&y_val[12][i],&x_val[12][i],&blg_val[13][i],&y_val[13][i],&x_val[13][i],
  &blg_val[14][i],&y_val[14][i],&x_val[14][i],&blg_val[15][i],&y_val[15][i],&x_val[15][i],
  &blg_val[16][i],&y_val[16][i],&x_val[16][i],&blg_val[17][i],&y_val[17][i],&x_val[17][i],
  &blg_val[18][i],&y_val[18][i],&x_val[18][i],&blg_val[19][i],&y_val[19][i],&x_val[19][i],
  &blg_val[20][i],&y_val[20][i],&x_val[20][i],&blg_val[21][i],&y_val[21][i],&x_val[21][i],
  &blg_val[22][i],&y_val[22][i],&x_val[22][i],&blg_val[23][i],&y_val[23][i],&x_val[23][i],
  &blg_val[24][i],&y_val[24][i],&x_val[24][i],&blg_val[25][i],&y_val[25][i],&x_val[25][i],
  &blg_val[26][i],&y_val[26][i],&x_val[26][i],&blg_val[27][i],&y_val[27][i],&x_val[27][i],
  &blg_val[28][i],&y_val[28][i],&x_val[28][i],&blg_val[29][i],&y_val[29][i],&x_val[29][i],
  &blg_val[30][i],&y_val[30][i],&x_val[30][i],&blg_val[31][i],&y_val[31][i],&x_val[31][i],
  &blg_val[32][i],&y_val[32][i],&x_val[32][i],&blg_val[33][i],&y_val[33][i],&x_val[33][i],
  &blg_val[34][i],&y_val[34][i],&x_val[34][i],&blg_val[35][i],&y_val[35][i],&x_val[35][i],
  &blg_val[36][i],&y_val[36][i],&x_val[36][i],&blg_val[37][i],&y_val[37][i],&x_val[37][i],
  &blg_val[38][i],&y_val[38][i],&x_val[38][i],&blg_val[39][i],&y_val[39][i],&x_val[39][i],
  &blg_val[40][i],&y_val[40][i],&x_val[40][i],&blg_val[41][i],&y_val[41][i],&x_val[41][i],
  &blg_val[42][i],&y_val[42][i],&x_val[42][i],&blg_val[43][i],&y_val[43][i],&x_val[43][i],
  &blg_val[44][i],&y_val[44][i],&x_val[44][i],&blg_val[45][i],&y_val[45][i],&x_val[45][i],
  &blg_val[46][i],&y_val[46][i],&x_val[46][i],&blg_val[47][i],&y_val[47][i],&x_val[47][i],
  &blg_val[48][i],&y_val[48][i],&x_val[48][i],&blg_val[49][i],&y_val[49][i],&x_val[49][i]);
  }
  fclose(cfPtr1);
  //printf("\ny[0][0]=%f \n y[0][399]=%f",y_val[0][0],y_val[49][399]);
}

if((cfPtr2=fopen("ikinci_50.dat","r"))==NULL)
        printf("Dosyada veri yok\n");
else        {
        i=0;

fscanf(cfPtr1,"%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f
%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f
%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f
%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f
%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f",
  &blg_val[50][i],&y_val[50][i],&x_val[50][i],&blg_val[51][i],&y_val[51][i],&x_val[51][i],
  &blg_val[52][i],&y_val[52][i],&x_val[52][i],&blg_val[53][i],&y_val[53][i],&x_val[53][i],
  &blg_val[54][i],&y_val[54][i],&x_val[54][i],&blg_val[55][i],&y_val[55][i],&x_val[55][i],
  &blg_val[56][i],&y_val[56][i],&x_val[56][i],&blg_val[57][i],&y_val[57][i],&x_val[57][i],
  &blg_val[58][i],&y_val[58][i],&x_val[58][i],&blg_val[59][i],&y_val[59][i],&x_val[59][i],
```

Appendix -1 Continued

```
  &blg_val[60][i],&y_val[60][i],&x_val[60][i],&blg_val[61][i],&y_val[61][i],&x_val[61][i],
  &blg_val[62][i],&y_val[62][i],&x_val[62][i],&blg_val[63][i],&y_val[63][i],&x_val[63][i],
  &blg_val[64][i],&y_val[64][i],&x_val[64][i],&blg_val[65][i],&y_val[65][i],&x_val[65][i],
  &blg_val[66][i],&y_val[66][i],&x_val[66][i],&blg_val[67][i],&y_val[67][i],&x_val[67][i],
  &blg_val[68][i],&y_val[68][i],&x_val[68][i],&blg_val[69][i],&y_val[69][i],&x_val[69][i],
  &blg_val[70][i],&y_val[70][i],&x_val[70][i],&blg_val[71][i],&y_val[71][i],&x_val[71][i],
  &blg_val[72][i],&y_val[72][i],&x_val[72][i],&blg_val[73][i],&y_val[73][i],&x_val[73][i],
  &blg_val[74][i],&y_val[74][i],&x_val[74][i],&blg_val[75][i],&y_val[75][i],&x_val[75][i],
  &blg_val[76][i],&y_val[76][i],&x_val[76][i],&blg_val[77][i],&y_val[77][i],&x_val[77][i],
  &blg_val[78][i],&y_val[78][i],&x_val[78][i],&blg_val[79][i],&y_val[79][i],&x_val[79][i],
  &blg_val[80][i],&y_val[80][i],&x_val[80][i],&blg_val[81][i],&y_val[81][i],&x_val[81][i],
  &blg_val[82][i],&y_val[82][i],&x_val[82][i],&blg_val[83][i],&y_val[83][i],&x_val[83][i],
  &blg_val[84][i],&y_val[84][i],&x_val[84][i],&blg_val[85][i],&y_val[85][i],&x_val[85][i],
  &blg_val[86][i],&y_val[86][i],&x_val[86][i],&blg_val[87][i],&y_val[87][i],&x_val[87][i],
  &blg_val[88][i],&y_val[88][i],&x_val[88][i],&blg_val[89][i],&y_val[89][i],&x_val[89][i],
  &blg_val[90][i],&y_val[90][i],&x_val[90][i],&blg_val[91][i],&y_val[91][i],&x_val[91][i],
  &blg_val[92][i],&y_val[92][i],&x_val[92][i],&blg_val[93][i],&y_val[93][i],&x_val[93][i],
  &blg_val[94][i],&y_val[94][i],&x_val[94][i],&blg_val[95][i],&y_val[95][i],&x_val[95][i],
  &blg_val[96][i],&y_val[96][i],&x_val[96][i],&blg_val[97][i],&y_val[97][i],&x_val[97][i],
  &blg_val[98][i],&y_val[98][i],&x_val[98][i],&blg_val[99][i],&y_val[99][i],&x_val[99][i]);
  while(i!=399)
    {
    i=i+1;

fscanf(cfPtr1,"%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f
%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f
%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f
%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f
%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f%f",
  &blg_val[50][i],&y_val[50][i],&x_val[50][i],&blg_val[51][i],&y_val[51][i],&x_val[51][i],
  &blg_val[52][i],&y_val[52][i],&x_val[52][i],&blg_val[53][i],&y_val[53][i],&x_val[53][i],
  &blg_val[54][i],&y_val[54][i],&x_val[54][i],&blg_val[55][i],&y_val[55][i],&x_val[55][i],
  &blg_val[56][i],&y_val[56][i],&x_val[56][i],&blg_val[57][i],&y_val[57][i],&x_val[57][i],
  &blg_val[58][i],&y_val[58][i],&x_val[58][i],&blg_val[59][i],&y_val[59][i],&x_val[59][i],
  &blg_val[60][i],&y_val[60][i],&x_val[60][i],&blg_val[61][i],&y_val[61][i],&x_val[61][i],
  &blg_val[62][i],&y_val[62][i],&x_val[62][i],&blg_val[63][i],&y_val[63][i],&x_val[63][i],
  &blg_val[64][i],&y_val[64][i],&x_val[64][i],&blg_val[65][i],&y_val[65][i],&x_val[65][i],
  &blg_val[66][i],&y_val[66][i],&x_val[66][i],&blg_val[67][i],&y_val[67][i],&x_val[67][i],
  &blg_val[68][i],&y_val[68][i],&x_val[68][i],&blg_val[69][i],&y_val[69][i],&x_val[69][i],
  &blg_val[70][i],&y_val[70][i],&x_val[70][i],&blg_val[71][i],&y_val[71][i],&x_val[71][i],
  &blg_val[72][i],&y_val[72][i],&x_val[72][i],&blg_val[73][i],&y_val[73][i],&x_val[73][i],
  &blg_val[74][i],&y_val[74][i],&x_val[74][i],&blg_val[75][i],&y_val[75][i],&x_val[75][i],
  &blg_val[76][i],&y_val[76][i],&x_val[76][i],&blg_val[77][i],&y_val[77][i],&x_val[77][i],
  &blg_val[78][i],&y_val[78][i],&x_val[78][i],&blg_val[79][i],&y_val[79][i],&x_val[79][i],
  &blg_val[80][i],&y_val[80][i],&x_val[80][i],&blg_val[81][i],&y_val[81][i],&x_val[81][i],
  &blg_val[82][i],&y_val[82][i],&x_val[82][i],&blg_val[83][i],&y_val[83][i],&x_val[83][i],
  &blg_val[84][i],&y_val[84][i],&x_val[84][i],&blg_val[85][i],&y_val[85][i],&x_val[85][i],
  &blg_val[86][i],&y_val[86][i],&x_val[86][i],&blg_val[87][i],&y_val[87][i],&x_val[87][i],
```

Appendix -1 Continued

```
  &blg_val[88][i],&y_val[88][i],&x_val[88][i],&blg_val[89][i],&y_val[89][i],&x_val[89][i],
  &blg_val[90][i],&y_val[90][i],&x_val[90][i],&blg_val[91][i],&y_val[91][i],&x_val[91][i],
  &blg_val[92][i],&y_val[92][i],&x_val[92][i],&blg_val[93][i],&y_val[93][i],&x_val[93][i],
  &blg_val[94][i],&y_val[94][i],&x_val[94][i],&blg_val[95][i],&y_val[95][i],&x_val[95][i],
  &blg_val[96][i],&y_val[96][i],&x_val[96][i],&blg_val[97][i],&y_val[97][i],&x_val[97][i],
  &blg_val[98][i],&y_val[98][i],&x_val[98][i],&blg_val[99][i],&y_val[99][i],&x_val[99][i]);
  }
  fclose(cfPtr2);
  //printf("\ny[99][399]=%f",y_val[99][399]);
  }

/*******************************************************************************************/
/************COMPUTATION RESPONDING SIZE FOR 100 SAMPLES*************/
/*******************************************************************************************/

for(int i=0;i<100;i++){
    m[i][0]=0;     m[i][1]=0;     m[i][2]=0;     m[i][3]=0;
    n[i][0]=0; n[i][1]=0; n[i][2]=0; n[i][3]=0;

x_blgorn_top[i][0]=0;x_blgorn_top[i][1]=0;x_blgorn_top[i][2]=0;x_blgorn_top[i][3]=0;
        for(int j=0;j<400;j++) {
        if(blg_val[i][j]==1){
        x_blgorn_top[i][0]=x_blgorn_top[i][0]+x_val[i][j];
      n[i][0]=n[i][0]+1;
        if(y_val[i][j]!=0){
        m[i][0]=m[i][0]+1;
      }
    }
    if(blg_val[i][j]==2){
        x_blgorn_top[i][1]=x_blgorn_top[i][1]+x_val[i][j];
        n[i][1]=n[i][1]+1;
        if(y_val[i][j]!=0){
        m[i][1]=m[i][1]+1;
      }
    }
    if(blg_val[i][j]==3){
        x_blgorn_top[i][2]=x_blgorn_top[i][2]+x_val[i][j];
        n[i][2]=n[i][2]+1;
        if(y_val[i][j]!=0){
        m[i][2]=m[i][2]+1;
      }
        }
    if(blg_val[i][j]==4){
        x_blgorn_top[i][3]=x_blgorn_top[i][3]+x_val[i][j];
        n[i][3]=n[i][3]+1;
        if(y_val[i][j]!=0){
        m[i][3]=m[i][3]+1;
      }
```

Appendix -1 Continued

```
      }
   }
   M[i]=m[i][0]+m[i][1]+m[i][2]+m[i][3];
   //printf("\nx_blgorn_top[%d][0]=%f",i,x_blgorn_top[i][0]);
   //printf("\nM[%d]=%d",i,M[i]);
  /* printf("\nm[%d][1]=%d",i,m[i][0]);
  printf("\nm[%d][2]=%d",i,m[i][1]);
   printf("\nm[%d][3]=%d",i,m[i][2]);
   printf("\nm[%d][4]=%d",i,m[i][3]); */
/*   printf("\nx_blgorn_top[%d][0]=%f",i,x_blgorn_top[i][0]);
   printf("\nx_blgorn_top[%d][1]=%f",i,x_blgorn_top[i][1]);
   printf("\nx_blgorn_top[%d][2]=%f",i,x_blgorn_top[i][2]);
   printf("\nx_blgorn_top[%d][3]=%f",i,x_blgorn_top[i][3]);*/
 }
 /*for(int t=0;t<25;t++){
 printf("\nx_blgorn_top[%d][0]=%f",t,x_blgorn_top[t][0]);
   printf("\nx_blgorn_top[%d][1]=%f",t,x_blgorn_top[t][1]);
   printf("\nx_blgorn_top[%d][2]=%f",t,x_blgorn_top[t][2]);
   printf("\nx_blgorn_top[%d][3]=%f",t,x_blgorn_top[t][3]);
   }
 /*  printf("\nm[%d][1]=%d",t,m[t][0]);
   printf("\nm[%d][2]=%d",t,m[t][1]);
   printf("\nm[%d][3]=%d",t,m[t][2]);
   printf("\nm[%d][4]=%d",t,m[t][3]);
   }
   /*for(int t=94;t<100;t++){
   printf("\nn[%d][1]=%d",t,n[t][0]);
   printf("\nn[%d][2]=%d",t,n[t][1]);
   printf("\nn[%d][3]=%d",t,n[t][2]);
   printf("\nn[%d][4]=%d",t,n[t][3]);    */
   //}
//printf("m[50][1]=%d\tm[50][2]=%d\tm[99][3]=%d\tm[99][4]=%d",m[50][0],m[50][1],m[
99][2],m[99][3]);

/*******************************************************************************************/
 /*******COMPUTATION OF N̂p VALUES FOR FOUR GROUP (REGION) **********/
/*******************************************************************************************/

//printf("\nn[1][1]=%d\nn[1][2]=%d\nn[1][3]=%d\nn[1][4]=%d",n[0][0],n[0][1],n[0][2],n[0
][3]);
 for(int k=0;k<100;k++){
          Np_sapka[k][0]=int(2.125*n[k][0]);
   Np_sapka[k][1]=int(2.125*n[k][1]);
   Np_sapka[k][2]=int(2.125*n[k][2]);
   Np_sapka[k][3]=int(2.125*n[k][3]);

 }
 /*for(int k=0;k<40;k++)
```

Appendix -1 Continued

```
          printf("\nNp_sapka[%d][1]=%d\nNp_sapka[%d][2]=%d\nNp_sapka[%d][3]=%
d\nNp_sapka[%d][4]=%d\n",
  k,Np_sapka[k][0],k,Np_sapka[k][1],k,Np_sapka[k][2],k,Np_sapka[k][3]);
```

/***/
/**********COMPUTATION OF POINT ESTIMATORS $\hat{Y}_{EXP}$, $\hat{Y}_{RA_U}$ AND $\hat{Y}_{RA_S}$ *********/
/***/

```
for(int i=0;i<100;i++){
          toplam_y=0; toplam_x=0;
          for(int j=0;j<400;j++){
          toplam_y=toplam_y+y_val[i][j];
     toplam_x=toplam_x+x_val[i][j];
  }
  //printf("\ntoplam_y_%d=%d",i,toplam_y);
  //printf("\ntoplam_x_%d=%d",i,toplam_y);
  //printf("\nM_Toplam[%d]=%d",i,M_Toplam[i]);
  Yexp[i]=float(850/float(M_Toplam[i])*toplam_y);
  yr_ort[i]=toplam_y/M_Toplam[i];
  xr_ort[i]=toplam_x/M_Toplam[i];
  Y_rau[i]=32776*yr_ort[i]/xr_ort[i];
  Y_ras[i]=toplam_x*yr_ort[i]/xr_ort[i];
  //printf("\nYexp[%d]=%f",i,Yexp[i]);
  //printf("\ntoplam_x=%d",toplam_x);
  //printf("\nY_ras[%d]=%f",i,Y_ras[i]);
  //printf("\nyr_ort[%d]=%f",i,yr_ort[i]);
          //printf("\nxr_ort[%d]=%f",i,xr_ort[i]);
  //printf("\nY_rau[%d]=%f",i,Y_rau[i]);
}
```

/***/
/*************COMPUTATION OF POINT ESTIMATORS $\hat{Y}_{REG_U}$ AND $\hat{Y}_{REG_S}$ ***********/
/***/

```
for(int i=0;i<100;i++){
          top_x[i]=0; top_y[i]=0;
          for(int j=0;j<400;j++){
          top_x[i]=top_x[i]+x_val[i][j];
     top_y[i]=top_y[i]+y_val[i][j];
  }
  //printf("\ntop_x[%d]=%f",i,top_x[i]);
  //printf("\ntop_y[%d]=%f",i,top_y[i]);

}
//printf("\ntop_x[0]=%f",top_x[0]);
```

Appendix -1 Continued

```
//printf("\ntop_y[0]=%f",top_y[0]);
for(int t=0;t<100;t++){
          yk_carp_xk_top[t]=0;xk_kare_top[t]=0;
          for(int z=0;z<400;z++){
      yk_carp_xk_top[t]=yk_carp_xk_top[t]+y_val[t][z]*x_val[t][z];
      xk_kare_top[t]=xk_kare_top[t]+x_val[t][z]*x_val[t][z];
  }
  //printf("\nyk_carp_xk_top[%d]=%f",t,yk_carp_xk_top[t]);
  //printf("\nxk_kare_top[%d]=%f",t,xk_kare_top[t]);
}
//printf("\nyk_carp_xk_top[0]=%f",yk_carp_xk_top[0]);
//printf("\nxk_kare_top[0]=%f",xk_kare_top[0]);
for(int i=0;i<100;i++ ){
          Beta_sapka[i]=(yk_carp_xk_top[i]-top_x[i]*top_y[i]/M[i])/(xk_kare_top[i]-
top_x[i]*top_x[i]/M[i]);
          //printf("\nBeta_sapka[%d]=%f",i,Beta_sapka[i]);
 }
 //printf("\nBeta_sapka[0]=%f",Beta_sapka[0]);
 for(int j=0;j<100;j++){
          Y_regu[j]=850*(yr_ort[j]+(39-xr_ort[j])*Beta_sapka[j]);
  //printf("\nYregu[%d]=%f",j,Y_regu[j]);
 }
 for(int i=0;i<100;i++){
          Y_regs[i]=400*(yr_ort[i]+(top_x[i]/M [i]-xr_ort[i])*Beta_sapka[i]);
  //printf("\nornek_ort=%f",top_x[i]/M [i]) ;
  //printf("\nYregs[%d]=%f",i,Y_regs[i]);
 }
```

/***/
/**********COMPUTATION OF POINT ESTIMATORS $\hat{Y}_{SEPRA_U}$ AND $\hat{Y}_{SEPRA_S}$ ***********/
/***/

```
for(int i=0;i<100;i++){
          yrp1_ort[i]=0;yrp2_ort[i]=0;yrp3_ort[i]=0;yrp4_ort[i]=0;
  xrp1_ort[i]=0;xrp2_ort[i]=0;xrp3_ort[i]=0;xrp4_ort[i]=0;
          for(int j=0;j<400;j++){
          if(blg_val[i][j]==1){
        yrp1_ort[i]=yrp1_ort[i]+y_val[i][j];
        xrp1_ort[i]=xrp1_ort[i]+x_val[i][j];
     }
     if(blg_val[i][j]==2){
        yrp2_ort[i]=yrp2_ort[i]+y_val[i][j];
        xrp2_ort[i]=xrp2_ort[i]+x_val[i][j];
     }
     if(blg_val[i][j]==3){
        yrp3_ort[i]=yrp3_ort[i]+y_val[i][j];
```

Appendix -1 Continued

```
        xrp3_ort[i]=xrp3_ort[i]+x_val[i][j];
      }
      if(blg_val[i][j]==4){
        yrp4_ort[i]=yrp4_ort[i]+y_val[i][j];
        xrp4_ort[i]=xrp4_ort[i]+x_val[i][j];
      }
   }

          yrp1_ort[i]=yrp1_ort[i]/float(m[i][0]);  xrp1_ort[i]=xrp1_ort[i]/float(m[i][0]);
   yrp2_ort[i]=yrp2_ort[i]/float(m[i][1]);  xrp2_ort[i]=xrp2_ort[i]/float(m[i][0]);
   yrp3_ort[i]=yrp3_ort[i]/float(m[i][2]);  xrp3_ort[i]=xrp3_ort[i]/float(m[i][0]);
   yrp4_ort[i]=yrp4_ort[i]/float(m[i][3]);  xrp4_ort[i]=xrp4_ort[i]/float(m[i][0]);

/* printf("\nyrp1_ort[%d]=%f",t,yrp1_ort[t]);
printf("\nyrp2_ort[%d]=%f",t,yrp2_ort[t]);
printf("\nyrp3_ort[%d]=%f",t,yrp3_ort[t]);
printf("\nyrp4_ort[%d]=%f",t,yrp4_ort[t]);     */

/*printf("\nxrp1_ort[%d]=%f",i,xrp1_ort[i]);
printf("\nxrp2_ort[%d]=%f",i,xrp2_ort[i]);
printf("\nxrp3_ort[%d]=%f",i,xrp3_ort[i]);
printf("\nxrp4_ort[%d]=%f",i,xrp4_ort[i]);
}  */
}
/*for(int t=0;t<25;t++){
 printf("\nxrp1_ort[%d]=%f",t,xrp1_ort[t]);
printf("\nxrp2_ort[%d]=%f",t,xrp2_ort[t]);
printf("\nxrp3_ort[%d]=%f",t,xrp3_ort[t]);
printf("\nxrp4_ort[%d]=%f",t,xrp4_ort[t]);
} */

 //printf("\nm[0][1]%d=",m[0][0]);
 //printf("\nyrp1_ort[0]=%f",yrp1_ort[0]);
 for(int i=0;i<100;i++){

          Y_seprau[i]=8848*yrp1_ort[i]/xrp1_ort[i]+7840*yrp2_ort[i]/xrp2_ort[i]+7841*yr
p3_ort[i]/xrp3_ort[i]+8269*yrp4_ort[i]/xrp4_ort[i];
   //printf("\nY_seprau[%d]=%f",i,Y_seprau[i]);
 }
 for(int j=0;j<100;j++){

          Y_sepras[j]=850/400*(x_blgorn_top[j][0]*yrp1_ort[j]/xrp1_ort[j]+x_blgorn_top[j
][1]*yrp2_ort[j]/xrp2_ort[j]+x_blgorn_top[j][2]*yrp3_ort[j]/xrp3_ort[j]+x_blgorn_top[j][3]
*yrp4_ort[j]/xrp4_ort[j]);
  //printf("\nY_sepras[%d]=%f",j,Y_sepras[j]);
 }
```

Appendix -1 Continued

/***/

/*****************COMPUTATION OF POINT ESTIMATORS $\hat{Y}_{PST}$ *********************/

/***/

```
for(int i=0;i<100;i++){
        y_pst[i]=229*yrp1_ort[i]+203*yrp2_ort[i]+202*yrp3_ort[i]+216*yrp4_ort[i];
  //printf("\ny_pst[%d]=%f",i,y_pst[i]);
}
/*printf("\nyrp1_ort[%d]=%f",0,yrp1_ort[0]);
printf("\nyrp2_ort[%d]=%f",0,yrp2_ort[0]);
printf("\nyrp3_ort[%d]=%f",0,yrp3_ort[0]);
printf("\nyrp4_ort[%d]=%f",0,yrp4_ort[0]);    */
```

/***/

/*****************COMPUTATION OF POINT ESTIMATORS $\hat{Y}_{WCE}$ *********************/

/***/

```
for(int i=0;i<100;i++){
        Y_wce[i]=float(Np_sapka[i][0])*yrp1_ort[i]+float(Np_sapka[i][1])*yrp2_ort[i]+fl
oat(Np_sapka[i][2])*yrp3_ort[i]+float(Np_sapka[i][3])*yrp4_ort[i];
  //printf("\nY_wce[%d]=%f",i,Y_wce[i]);
}
/*printf("\nNp_sapka[0][0]=%d",Np_sapka[0][0]);
printf("\nNp_sapka[0][1]=%d",Np_sapka[0][1]);
printf("\nNp_sapka[0][2]=%d",Np_sapka[0][2]);
printf("\nNp_sapka[0][3]=%d",Np_sapka[0][3]);        */
```

/***/

/*********COMPUTATION OF POINT ESTIMATORS $\hat{Y}_{SEPRREG_U}$ AND $\hat{Y}_{SEPREG_S}$ ********/

/***/

```
for(int i=0;i<100;i++) {
        cov1[i]=0;cov2[i]=0;cov3[i]=0;cov4[i]=0;
  var1[i]=0; var2[i]=0; var3[i]=0; var4[i]=0;
        for(int j=0;j<400;j++){
        if(blg_val[i][j]==1) {
        if(x_val[i][j]!=0){
                cov1[i]=cov1[i]+(x_val[i][j]-xrp1_ort[i])*(y_val[i][j]-yrp1_ort[i]);
      var1[i]=var1[i]+ (x_val[i][j]-xrp1_ort[i])*(x_val[i][j]-xrp1_ort[i]);
      }
    }
    if(blg_val[i][j]==2)
        if(x_val[i][j]!=0){
                cov2[i]=cov2[i]+(x_val[i][j]-xrp2_ort[i])*(y_val[i][j]-yrp2_ort[i]);
```

Appendix -1 Continued

```
        var2[i]= var2[i]+(x_val[i][j]-xrp2_ort[i])*(x_val[i][j]-xrp2_ort[i]);
        }
      if(blg_val[i][j]==3)
          if(x_val[i][j]!=0){
                  cov3[i]=cov3[i]+(x_val[i][j]-xrp3_ort[i])*(y_val[i][j]-yrp3_ort[i]);
        var3[i]=var3[i]+ (x_val[i][j]-xrp3_ort[i])*(x_val[i][j]-xrp3_ort[i]);
        }
      if(blg_val[i][j]==4)
          if(x_val[i][j]!=0){
                  cov4[i]=cov4[i]+(x_val[i][j]-xrp4_ort[i])*(y_val[i][j]-yrp4_ort[i]);
        var4[i]=var4[i]+(x_val[i][j]-xrp4_ort[i])*(x_val[i][j]-xrp4_ort[i]);
        }
   }
   cov1[i]=cov1[i]/(m[i][0]-1); var1[i]=var1[i]/(m[i][0]-1);
   cov2[i]=cov2[i]/(m[i][1]-1); var2[i]=var2[i]/(m[i][1]-1);
   cov3[i]=cov3[i]/(m[i][2]-1); var3[i]=var3[i]/(m[i][2]-1);
   cov4[i]=cov4[i]/(m[i][3]-1); var4[i]=var4[i]/(m[i][3]-1);
   /*printf("\ncov4[%d]=%f",i,cov4[i]);
   printf("\nvar4[%d]=%f",i,var4[i]);*/
}
 for(int n=0;n<100;n++){
          Beta_sapka1[n]=cov1[n]/var1[n];
   Beta_sapka2[n]=cov2[n]/var2[n];
   Beta_sapka3[n]=cov3[n]/var3[n];
   Beta_sapka4[n]=cov4[n]/var4[n];
   //printf("\nBeta_sapka4[%d]=%f",n,Beta_sapka4[n]);
 }

 for(int i=0;i<100;i++){
          Y_sepregu[i]=229*(yrp1_ort[i]+(39-
xrp1_ort[i])*Beta_sapka1[i])+203*(yrp2_ort[i]+(39-
xrp2_ort[i])*Beta_sapka2[i])+202*(yrp3_ort[i]+(39-
xrp3_ort[i])*Beta_sapka3[i])+216*(yrp4_ort[i]+(38-xrp4_ort[i])*Beta_sapka4[i]);
          //printf("\nY_sepregu[%d]=%f",i,Y_sepregu[i]);
 }
 for(int i=0;i<100;i++){
          Y_sepregs[i]=Np_sapka[i][0]*(yrp1_ort[i]+(x_blgorn_top[i][0]/n[i][0]-
xrp1_ort[i])*Beta_sapka1[i])+Np_sapka[i][1]*(yrp2_ort[i]+(x_blgorn_top[i][1]/n[i][1]-
xrp2_ort[i])*Beta_sapka2[i])+ Np_sapka[i][2]*(yrp3_ort[i]+(x_blgorn_top[i][2]/n[i][2]-
xrp3_ort[i])*Beta_sapka3[i])+ Np_sapka[i][3]*(yrp4_ort[i]+(x_blgorn_top[i][3]/n[i][3]-
xrp4_ort[i])*Beta_sapka4[i]);
    //printf("\nY_sepregs[%d]=%f",i,Y_sepregs[i]);
 }
```

Appendix -1 Continued

```
/*********************************************************************************************/
/******COMPUTATION OF SIMULATION RELATIVE BIAS AND SIMULATION******/
              VARIANCE FOR DIFFERENT POINT ESTIMATORS
/*********************************************************************************************/

Esim_Y[0]=0;
for(int i=0;i<100;i++)
        Esim_Y[0]=Esim_Y[0]+Yexp[i];
Esim_Y[0]=Esim_Y[0]/100;
printf("\nYexp_Esim=%f",Esim_Y[0]);
Esim_Y[1]=0;
for(int i=0;i<100;i++)
        Esim_Y[1]=Esim_Y[1]+y_pst[i];
Esim_Y[1]=Esim_Y[1]/100;
printf("\nYpst_Esim=%f",Esim_Y[1]);
Esim_Y[2]=0;
for(int i=0;i<100;i++)
        Esim_Y[2]=Esim_Y[2]+Y_wce[i];
Esim_Y[2]=Esim_Y[2]/100;
printf("\nYwce_Esim=%f",Esim_Y[2]);
Esim_Y[3]=0;
for(int i=0;i<100;i++)
        Esim_Y[3]=Esim_Y[3]+Y_rau[i];
Esim_Y[3]=Esim_Y[3]/100;
printf("\nYrau_Esim=%f",Esim_Y[3]);
Esim_Y[4]=0;
for(int i=0;i<100;i++)
        Esim_Y[4]=Esim_Y[4]+Y_ras[i];
Esim_Y[4]=Esim_Y[4]/100;
printf("\nYras_Esim=%f",Esim_Y[4]);

Esim_Y[5]=0;
for(int i=0;i<100;i++)
        Esim_Y[5]=Esim_Y[5]+Y_regu[i];
Esim_Y[5]=Esim_Y[5]/100;
printf("\nYregu_Esim=%f",Esim_Y[5]);
Esim_Y[6]=0;
for(int i=0;i<100;i++)
        Esim_Y[6]=Esim_Y[6]+Y_regs[i];
Esim_Y[6]=Esim_Y[6]/100;
printf("\nYregs_Esim=%f",Esim_Y[6]);
Esim_Y[7]=0;
for(int i=0;i<100;i++)
        Esim_Y[7]=Esim_Y[7]+Y_seprau[i];
Esim_Y[7]=Esim_Y[7]/100;
printf("\nYseprau_Esim=%f",Esim_Y[7]);
Esim_Y[8]=0;
for(int i=0;i<100;i++)
```

Appendix -1 Continued

```
        Esim_Y[8]=Esim_Y[8]+Y_sepras[i];
Esim_Y[8]=Esim_Y[8]/100;
printf("\nYsepras_Esim=%f",Esim_Y[8]);
Esim_Y[9]=0;
for(int i=0;i<100;i++)
        Esim_Y[9]=Esim_Y[9]+Y_sepregu[i];
Esim_Y[9]=Esim_Y[9]/100;
printf("\nYsepregu_Esim=%f",Esim_Y[9]);
Esim_Y[10]=0;
for(int i=0;i<100;i++)
        Esim_Y[10]=Esim_Y[10]+Y_sepregs[i];
Esim_Y[10]=Esim_Y[10]/100;
printf("\nYsepregs_Esim=%f",Esim_Y[10]);


for(int i=0;i<11;i++){
        RBsim_Y[i]=0;
        RBsim_Y[i]=100*(Esim_Y[i]-1282671)/1282671;
}
printf("\nYexp_RBsim=%f",RBsim_Y[0]);
printf("\nYpst_RBsim=%f",RBsim_Y[1]);
printf("\nYwce_RBsim=%f",RBsim_Y[2]);
printf("\nYrau_RBsim=%f",RBsim_Y[3]);
printf("\nYras_RBsim=%f",RBsim_Y[4]);
printf("\nYregu_RBsim=%f",RBsim_Y[5]);
printf("\nYregs_RBsim=%f",RBsim_Y[6]);
printf("\nYseprau_RBsim=%f",RBsim_Y[7]);
printf("\nYsepras_RBsim=%f",RBsim_Y[8]);
printf("\nYsepregu_RBsim=%f",RBsim_Y[9]);
printf("\nYsepregs_RBsim=%f",RBsim_Y[10]);

VarSim_Y[0]=0;
for(int i=0;i<100;i++)
        VarSim_Y[0]=(VarSim_Y[0]+(Yexp[i]-Esim_Y[0])*(Yexp[i]-Esim_Y[0]))/99;
printf("\nVarexp_sim=%f",VarSim_Y[0]);

VarSim_Y[1]=0;
for(int i=0;i<100;i++)
        VarSim_Y[1]=(VarSim_Y[1]+(y_pst[i]-Esim_Y[1])*(y_pst[i]-Esim_Y[1]))/99;
printf("\nVarpst_sim=%f",VarSim_Y[1]);

VarSim_Y[2]=0;
for(int i=0;i<100;i++)
        VarSim_Y[2]=(VarSim_Y[2]+(Y_wce[i]-Esim_Y[2])*(Y_wce[i]-Esim_Y[2]))/99;
printf("\nVarwce_sim=%f",VarSim_Y[2]);
```

Appendix -1 Continued

```
VarSim_Y[3]=0;
for(int i=0;i<100;i++)
        VarSim_Y[3]=(VarSim_Y[3]+(Y_rau[i]-Esim_Y[3])*(Y_rau[i]-Esim_Y[3]))/99;
printf("\nVarrau_sim=%f",VarSim_Y[3]);

VarSim_Y[4]=0;
for(int i=0;i<100;i++)
        VarSim_Y[4]=(VarSim_Y[4]+(Y_ras[i]-Esim_Y[4])*(Y_ras[i]-Esim_Y[4]))/99;
printf("\nVarras_sim=%f",VarSim_Y[4]);

VarSim_Y[5]=0;
for(int i=0;i<100;i++)
        VarSim_Y[5]=(VarSim_Y[5]+(Y_regu[i]-Esim_Y[5])*(Y_regu[i]-
Esim_Y[5]))/99;
printf("\nVarregu_sim=%f",VarSim_Y[5]);

VarSim_Y[6]=0;
for(int i=0;i<100;i++)
        VarSim_Y[6]=(VarSim_Y[6]+(Y_regs[i]-Esim_Y[6])*(Y_regs[i]-Esim_Y[6]))/99;
printf("\nVarregs_sim=%f",VarSim_Y[6]);

VarSim_Y[7]=0;
for(int i=0;i<100;i++)
VarSim_Y[7]=(VarSim_Y[7]+(Y_seprau[i]-Esim_Y[7])*(Y_seprau[i]-Esim_Y[7]))/99;
printf("\nVarseprau_sim=%f",VarSim_Y[7]);

VarSim_Y[8]=0;
for(int i=0;i<100;i++)
        VarSim_Y[8]=(VarSim_Y[8]+(Y_sepras[i]-Esim_Y[8])*(Y_sepras[i]-
Esim_Y[8]))/99;
printf("\nVarsepras_sim=%f",VarSim_Y[8]);

VarSim_Y[9]=0;
for(int i=0;i<100;i++)
        VarSim_Y[9]=(VarSim_Y[9]+(Y_sepregu[i]-Esim_Y[9])*(Y_sepregu[i]-
Esim_Y[9]))/99;
printf("\nVarsepregu_sim=%f",VarSim_Y[9]);

VarSim_Y[10]=0;
for(int i=0;i<100;i++)
        VarSim_Y[10]=(VarSim_Y[10]+(Y_sepregs[i]-Esim_Y[10])*(Y_sepregs[i]-
Esim_Y[10]))/99;
printf("\nVarsepregs_sim=%f",VarSim_Y[10]);

/*****************************************************************************/
system("PAUSE");
}
```

Appendix-2 Results of Expansion Estimators obtained from simulation with 100 samples for Info-S and Info-U.

Y_EXP[0]=1284327.750000
Y_EXP[1]=1295099.375000
Y_EXP[2]=1310186.500000
Y_EXP[3]=1276198.000000
Y_EXP[4]=1282300.250000
Y_EXP[5]=1275123.625000
Y_EXP[6]=1278435.750000
Y_EXP[7]=1274455.750000
Y_EXP[8]=1298645.500000
Y_EXP[9]=1281395.000000
Y_EXP[10]=1295991.250000
Y_EXP[11]=1274590.000000
Y_EXP[12]=1298392.875000
Y_EXP[13]=1278850.250000
Y_EXP[14]=1281982.625000
Y_EXP[15]=1292483.875000
Y_EXP[16]=1299520.125000
Y_EXP[17]=1302138.125000
Y_EXP[18]=1271565.000000
Y_EXP[19]=1298814.875000
Y_EXP[20]=1251721.250000
Y_EXP[21]=1274220.625000
Y_EXP[22]=1289951.125000
Y_EXP[23]=1293308.250000
Y_EXP[24]=1286392.500000
Y_EXP[25]=1277882.500000
Y_EXP[26]=1278097.000000
Y_EXP[27]=1278341.250000
Y_EXP[28]=1285571.125000
Y_EXP[29]=1287831.250000
Y_EXP[30]=1282052.500000
Y_EXP[31]=1299264.375000
Y_EXP[32]=1265647.500000
Y_EXP[33]=1291732.125000
Y_EXP[34]=1280043.875000
Y_EXP[35]=1287645.375000
Y_EXP[36]=1280206.875000
Y_EXP[37]=1293311.375000
Y_EXP[38]=1292399.875000
Y_EXP[39]=1285392.500000
Y_EXP[40]=1281607.125000
Y_EXP[41]=1280187.750000
Y_EXP[42]=1285355.875000
Y_EXP[43]=1273180.375000
Y_EXP[44]=1290006.625000
Y_EXP[45]=1283962.750000
Y_EXP[46]=1273564.875000
Y_EXP[47]=1276912.500000
Y_EXP[48]=1270963.750000
Y_EXP[49]=1292081.625000
Y_EXP[50]=1287721.875000
Y_EXP[51]=1279823.500000
Y_EXP[52]=1289914.375000
Y_EXP[53]=1286253.625000
Y_EXP[54]=1316353.000000
Y_EXP[55]=1277000.500000
Y_EXP[56]=1283210.750000
Y_EXP[57]=1286783.250000
Y_EXP[58]=1270323.750000
Y_EXP[59]=1282437.500000
Y_EXP[60]=1290884.000000
Y_EXP[61]=1274307.625000
Y_EXP[62]=1287436.875000
Y_EXP[63]=1279089.875000
Y_EXP[64]=1281373.750000
Y_EXP[65]=1286635.125000
Y_EXP[66]=1317073.750000
Y_EXP[67]=1286383.000000
Y_EXP[68]=1263000.125000
Y_EXP[69]=1291838.750000
Y_EXP[70]=1271780.125000
Y_EXP[71]=1296461.250000
Y_EXP[72]=1282376.875000
Y_EXP[73]=1275238.500000
Y_EXP[74]=1260076.875000
Y_EXP[75]=1293765.750000
Y_EXP[76]=1281838.000000
Y_EXP[77]=1282827.750000
Y_EXP[78]=1286611.625000
Y_EXP[79]=1278157.125000
Y_EXP[80]=1294114.875000
Y_EXP[81]=1291180.625000
Y_EXP[82]=1293595.375000
Y_EXP[83]=1276810.750000
Y_EXP[84]=1271402.250000
Y_EXP[85]=1281163.125000
Y_EXP[86]=1262606.750000
Y_EXP[87]=1290969.125000
Y_EXP[88]=1287819.000000
Y_EXP[89]=1269045.000000
Y_EXP[90]=1287435.625000
Y_EXP[91]=1285265.625000
Y_EXP[92]=1279028.750000
Y_EXP[93]=1276060.625000
Y_EXP[94]=1274124.750000
Y_EXP[95]=1271378.875000
Y_EXP[96]=1290616.875000
Y_EXP[97]=1293044.625000
Y_EXP[98]=1294942.125000
Y_EXP[99]=1280456.750000

Appendix -3 Results of Poststratified Estimators obtained from simulation with 100 samples for Info-U.

Y_PST[0]=1285282.125000
Y_PST[1]=1295380.375000
Y_PST[2]=1308894.125000
Y_PST[3]=1276705.625000
Y_PST[4]=1283470.000000
Y_PST[5]=1275793.750000
Y_PST[6]=1276496.875000
Y_PST[7]=1272002.750000
Y_PST[8]=1300646.500000
Y_PST[9]=1281067.125000
Y_PST[10]=1295356.375000
Y_PST[11]=1272995.375000
Y_PST[12]=1300639.500000
Y_PST[13]=1278824.625000
Y_PST[14]=1281641.125000
Y_PST[15]=1293830.250000
Y_PST[16]=1299637.375000
Y_PST[17]=1297594.250000
Y_PST[18]=1273351.000000
Y_PST[19]=1298835.000000
Y_PST[20]=1252246.750000
Y_PST[21]=1273772.750000
Y_PST[22]=1292573.000000
Y_PST[23]=1294181.625000
Y_PST[24]=1286726.875000
Y_PST[25]=1278985.500000
Y_PST[26]=1278776.500000
Y_PST[27]=1279183.000000
Y_PST[28]=1285616.875000
Y_PST[29]=1288689.750000
Y_PST[30]=1281577.875000
Y_PST[31]=1299666.625000
Y_PST[32]=1264983.875000
Y_PST[33]=1289999.750000
Y_PST[34]=1280260.625000
Y_PST[35]=1289398.750000
Y_PST[36]=1278340.500000
Y_PST[37]=1294861.375000
Y_PST[38]=1292408.250000
Y_PST[39]=1285540.000000
Y_PST[40]=1280964.500000
Y_PST[41]=1278718.250000
Y_PST[42]=1285232.625000
Y_PST[43]=1273363.750000
Y_PST[44]=1290360.000000
Y_PST[45]=1284387.375000
Y_PST[46]=1273011.750000
Y_PST[47]=1276855.375000
Y_PST[48]=1271823.125000
Y_PST[49]=1291200.250000
Y_PST[50]=1286118.375000
Y_PST[51]=1279866.250000
Y_PST[52]=1290183.000000
Y_PST[53]=1285325.375000
Y_PST[54]=1316973.750000
Y_PST[55]=1277266.125000
Y_PST[56]=1284031.625000
Y_PST[57]=1286694.000000
Y_PST[58]=1270533.000000
Y_PST[59]=1283822.125000
Y_PST[60]=1291086.500000
Y_PST[61]=1274162.375000
Y_PST[62]=1286774.750000
Y_PST[63]=1276550.250000
Y_PST[64]=1283448.000000
Y_PST[65]=1286547.000000
Y_PST[66]=1317161.375000
Y_PST[67]=1286731.375000
Y_PST[68]=1263710.250000
Y_PST[69]=1291995.500000
Y_PST[70]=1271461.875000
Y_PST[71]=1296042.875000
Y_PST[72]=1282148.750000
Y_PST[73]=1276268.500000
Y_PST[74]=1260300.875000
Y_PST[75]=1292062.625000
Y_PST[76]=1282199.375000
Y_PST[77]=1284777.750000
Y_PST[78]=1287113.000000
Y_PST[79]=1278749.500000
Y_PST[80]=1293365.750000
Y_PST[81]=1290813.250000
Y_PST[82]=1294025.250000
Y_PST[83]=1276874.375000
Y_PST[84]=1270815.000000
Y_PST[85]=1281845.500000
Y_PST[86]=1262791.125000
Y_PST[87]=1291236.250000
Y_PST[88]=1288219.625000
Y_PST[89]=1267433.750000
Y_PST[90]=1286969.375000
Y_PST[91]=1284443.875000
Y_PST[92]=1280572.000000
Y_PST[93]=1274379.250000
Y_PST[94]=1275169.500000
Y_PST[95]=1274606.250000
Y_PST[96]=1290238.750000
Y_PST[97]=1291841.625000
Y_PST[98]=1292623.000000
Y_PST[99]=1280469.750000

Appendix - 4 Results of Weighting Class Estimators obtained from simulation with 100 samples for Info-S.

Y_WCE[0]=1280919.875000
Y_WCE[1]=1292065.875000
Y_WCE[2]=1308944.500000
Y_WCE[3]=1273277.500000
Y_WCE[4]=1280425.875000
Y_WCE[5]=1272173.375000
Y_WCE[6]=1275219.125000
Y_WCE[7]=1270866.875000
Y_WCE[8]=1297613.625000
Y_WCE[9]=1279148.125000
Y_WCE[10]=1292878.000000
Y_WCE[11]=1272338.000000
Y_WCE[12]=1295206.000000
Y_WCE[13]=1275743.375000
Y_WCE[14]=1280422.875000
Y_WCE[15]=1290024.375000
Y_WCE[16]=1298353.125000
Y_WCE[17]=1298005.500000
Y_WCE[18]=1269429.500000
Y_WCE[19]=1297215.375000
Y_WCE[20]=1250342.500000
Y_WCE[21]=1270946.000000
Y_WCE[22]=1286875.750000
Y_WCE[23]=1290534.375000
Y_WCE[24]=1285257.000000
Y_WCE[25]=1275674.250000
Y_WCE[26]=1275019.125000
Y_WCE[27]=1276299.625000
Y_WCE[28]=1284183.375000
Y_WCE[29]=1286249.750000
Y_WCE[30]=1278942.875000
Y_WCE[31]=1297537.875000
Y_WCE[32]=1262675.875000
Y_WCE[33]=1290075.750000
Y_WCE[34]=1276930.875000
Y_WCE[35]=1285429.000000
Y_WCE[36]=1278852.000000
Y_WCE[37]=1290432.125000
Y_WCE[38]=1291121.875000
Y_WCE[39]=1282492.625000
Y_WCE[40]=1278340.000000
Y_WCE[41]=1275363.875000
Y_WCE[42]=1283332.625000
Y_WCE[43]=1269778.750000
Y_WCE[44]=1286850.125000
Y_WCE[45]=1280674.500000
Y_WCE[46]=1271771.125000
Y_WCE[47]=1275656.125000
Y_WCE[48]=1270069.625000
Y_WCE[49]=1289141.750000
Y_WCE[50]=1286036.000000
Y_WCE[51]=1278193.375000
Y_WCE[52]=1286408.000000
Y_WCE[53]=1283451.125000
Y_WCE[54]=1313807.000000
Y_WCE[55]=1274347.125000
Y_WCE[56]=1280537.625000
Y_WCE[57]=1284076.875000
Y_WCE[58]=1269467.625000
Y_WCE[59]=1280087.125000
Y_WCE[60]=1287743.125000
Y_WCE[61]=1272637.000000
Y_WCE[62]=1284091.875000
Y_WCE[63]=1274470.375000
Y_WCE[64]=1280763.250000
Y_WCE[65]=1284825.000000
Y_WCE[66]=1313801.625000
Y_WCE[67]=1284593.000000
Y_WCE[68]=1259877.875000
Y_WCE[69]=1290479.375000
Y_WCE[70]=1270166.125000
Y_WCE[71]=1294759.250000
Y_WCE[72]=1281017.250000
Y_WCE[73]=1274852.500000
Y_WCE[74]=1258184.000000
Y_WCE[75]=1290329.750000
Y_WCE[76]=1280415.125000
Y_WCE[77]=1281845.000000
Y_WCE[78]=1285340.250000
Y_WCE[79]=1276365.125000
Y_WCE[80]=1291541.125000
Y_WCE[81]=1287517.875000
Y_WCE[82]=1291963.375000
Y_WCE[83]=1273811.000000
Y_WCE[84]=1268508.375000
Y_WCE[85]=1279205.000000
Y_WCE[86]=1259797.125000
Y_WCE[87]=1289414.875000
Y_WCE[88]=1284627.625000
Y_WCE[89]=1265406.250000
Y_WCE[90]=1285493.125000
Y_WCE[91]=1282817.000000
Y_WCE[92]=1276194.500000
Y_WCE[93]=1270986.125000
Y_WCE[94]=1272588.750000
Y_WCE[95]=1269135.875000
Y_WCE[96]=1288954.125000
Y_WCE[97]=1291162.875000
Y_WCE[98]=1292810.250000
Y_WCE[99]=1276166.250000

Appendix -5 Results of Ratio Estimators obtained from simulation with 100 samples for Info-S.

Y_RA[0]=467027.093750
Y_RA[1]=529122.250000
Y_RA[2]=529901.250000
Y_RA[3]=521044.500000
Y_RA[4]=510934.218750
Y_RA[5]=506013.156250
Y_RA[6]=510726.750000
Y_RA[7]=520310.781250
Y_RA[8]=531114.687500
Y_RA[9]=522334.125000
Y_RA[10]=518360.531250
Y_RA[11]=515498.218750
Y_RA[12]=509817.093750
Y_RA[13]=513813.906250
Y_RA[14]=516370.937500
Y_RA[15]=520920.000000
Y_RA[16]=535242.312500
Y_RA[17]=514778.593750
Y_RA[18]=518017.500000
Y_RA[19]=535403.187500
Y_RA[20]=494553.250000
Y_RA[21]=512855.250000
Y_RA[22]=534582.812500
Y_RA[23]=493964.750000
Y_RA[24]=509960.625000
Y_RA[25]=522411.156250
Y_RA[26]=521541.000000
Y_RA[27]=512839.406250
Y_RA[28]=519133.250000
Y_RA[29]=528894.500000
Y_RA[30]=517323.375000
Y_RA[31]=516102.093750
Y_RA[32]=505371.781250
Y_RA[33]=506666.437500
Y_RA[34]=510274.218750
Y_RA[35]=512967.093750
Y_RA[36]=524880.625000
Y_RA[37]=522823.750000
Y_RA[38]=520280.000000
Y_RA[39]=522594.937500
Y_RA[40]=511824.781250
Y_RA[41]=516161.687500
Y_RA[42]=516268.406250
Y_RA[43]=507798.156250
Y_RA[44]=520969.750000
Y_RA[45]=516380.250000
Y_RA[46]=514208.218750
Y_RA[47]=521510.218750
Y_RA[48]=518450.250000
Y_RA[49]=515200.000000
Y_RA[50]=509142.250000
Y_RA[51]=509917.750000
Y_RA[52]=514342.843750
Y_RA[53]=512469.031250
Y_RA[54]=517154.218750
Y_RA[55]=504830.093750
Y_RA[56]=520009.343750
Y_RA[57]=509841.187500
Y_RA[58]=495890.062500
Y_RA[59]=507203.906250
Y_RA[60]=503057.218750
Y_RA[61]=519206.250000
Y_RA[62]=527111.062500
Y_RA[63]=497309.468750
Y_RA[64]=517456.218750
Y_RA[65]=515813.562500
Y_RA[66]=523847.343750
Y_RA[67]=515574.656250
Y_RA[68]=501539.218750
Y_RA[69]=537686.000000
Y_RA[70]=503640.218750
Y_RA[71]=508266.437500
Y_RA[72]=507045.156250
Y_RA[73]=508539.468750
Y_RA[74]=508794.000000
Y_RA[75]=519843.093750
Y_RA[76]=517720.218750
Y_RA[77]=507182.843750
Y_RA[78]=520511.812500
Y_RA[79]=525456.687500
Y_RA[80]=515877.906250
Y_RA[81]=511942.968750
Y_RA[82]=516299.437500
Y_RA[83]=504237.218750
Y_RA[84]=503460.906250
Y_RA[85]=523563.531250
Y_RA[86]=509784.875000
Y_RA[87]=531859.250000
Y_RA[88]=530808.187500
Y_RA[89]=517331.375000
Y_RA[90]=524948.812500
Y_RA[91]=515273.687500
Y_RA[92]=512705.687500
Y_RA[93]=509747.500000
Y_RA[94]=512079.468750
Y_RA[95]=484852.093750
Y_RA[96]=519076.093750
Y_RA[97]=535311.937500
Y_RA[98]=519944.187500
Y_RA[99]=514100.843750

Appendix - 6 Results of Ratio Estimators obtained from simulation with 100 samples for Info-U.

Y_RA[0]=1302414.750000
Y_RA[1]=1313627.625000
Y_RA[2]=1329153.000000
Y_RA[3]=1294652.000000
Y_RA[4]=1300689.625000
Y_RA[5]=1293789.500000
Y_RA[6]=1297239.625000
Y_RA[7]=1292927.000000
Y_RA[8]=1317077.625000
Y_RA[9]=1299827.125000
Y_RA[10]=1314490.125000
Y_RA[11]=1292927.000000
Y_RA[12]=1317077.625000
Y_RA[13]=1297239.625000
Y_RA[14]=1300689.625000
Y_RA[15]=1311040.000000
Y_RA[16]=1317940.250000
Y_RA[17]=1320527.750000
Y_RA[18]=1289476.875000
Y_RA[19]=1317940.250000
Y_RA[20]=1269638.750000
Y_RA[21]=1292927.000000
Y_RA[22]=1308452.375000
Y_RA[23]=1311902.500000
Y_RA[24]=1305002.375000
Y_RA[25]=1296377.000000
Y_RA[26]=1296377.000000
Y_RA[27]=1296377.000000
Y_RA[28]=1304139.750000
Y_RA[29]=1306727.375000
Y_RA[30]=1300689.625000
Y_RA[31]=1317940.250000
Y_RA[32]=1283439.125000
Y_RA[33]=1310177.500000
Y_RA[34]=1298102.125000
Y_RA[35]=1305864.875000
Y_RA[36]=1298964.625000
Y_RA[37]=1311902.500000
Y_RA[38]=1311040.000000
Y_RA[39]=1304139.750000
Y_RA[40]=1299827.125000
Y_RA[41]=1298964.625000
Y_RA[42]=1304139.750000
Y_RA[43]=1291201.875000
Y_RA[44]=1308452.375000
Y_RA[45]=1302414.750000
Y_RA[46]=1292064.375000
Y_RA[47]=1295514.500000
Y_RA[48]=1289476.875000
Y_RA[49]=1311040.000000
Y_RA[50]=1305864.875000
Y_RA[51]=1298102.125000
Y_RA[52]=1308452.375000
Y_RA[53]=1305002.375000
Y_RA[54]=1335190.750000
Y_RA[55]=1295514.500000
Y_RA[56]=1301552.250000
Y_RA[57]=1305002.375000
Y_RA[58]=1288614.375000
Y_RA[59]=1300689.625000
Y_RA[60]=1309315.000000
Y_RA[61]=1292927.000000
Y_RA[62]=1305864.875000
Y_RA[63]=1297239.625000
Y_RA[64]=1299827.125000
Y_RA[65]=1305002.375000
Y_RA[66]=1336053.250000
Y_RA[67]=1305002.375000
Y_RA[68]=1280851.625000
Y_RA[69]=1310177.500000
Y_RA[70]=1290339.375000
Y_RA[71]=1315352.625000
Y_RA[72]=1300689.625000
Y_RA[73]=1293789.500000
Y_RA[74]=1278264.000000
Y_RA[75]=1312765.000000
Y_RA[76]=1300689.625000
Y_RA[77]=1301552.250000
Y_RA[78]=1305002.375000
Y_RA[79]=1296377.000000
Y_RA[80]=1312765.000000
Y_RA[81]=1310177.500000
Y_RA[82]=1311902.500000
Y_RA[83]=1295514.500000
Y_RA[84]=1289476.875000
Y_RA[85]=1299827.125000
Y_RA[86]=1280851.625000
Y_RA[87]=1309315.000000
Y_RA[88]=1306727.375000
Y_RA[89]=1286889.250000
Y_RA[90]=1341158.500000
Y_RA[91]=1304139.750000
Y_RA[92]=1297239.625000
Y_RA[93]=1294652.000000
Y_RA[94]=1292064.375000
Y_RA[95]=1289476.875000
Y_RA[96]=1309315.000000
Y_RA[97]=1311902.500000
Y_RA[98]=1313627.625000
Y_RA[99]=1298964.625000

Appendix - 7 Results of Regression Estimators obtained from simulation with 100 samples for Info-S.

Y_REG[0]=604915.625000
Y_REG[1]=608742.937500
Y_REG[2]=615965.812500
Y_REG[3]=601717.875000
Y_REG[4]=601225.875000
Y_REG[5]=600035.375000
Y_REG[6]=599378.812500
Y_REG[7]=598938.437500
Y_REG[8]=609028.562500
Y_REG[9]=602655.312500
Y_REG[10]=610272.062500
Y_REG[11]=599286.875000
Y_REG[12]=611088.875000
Y_REG[13]=603835.125000
Y_REG[14]=603510.750000
Y_REG[15]=607266.625000
Y_REG[16]=611085.625000
Y_REG[17]=612464.437500
Y_REG[18]=597578.000000
Y_REG[19]=606147.687500
Y_REG[20]=587586.125000
Y_REG[21]=600674.125000
Y_REG[22]=604615.750000
Y_REG[23]=609644.250000
Y_REG[24]=605341.312500
Y_REG[25]=600904.125000
Y_REG[26]=601819.312500
Y_REG[27]=599561.000000
Y_REG[28]=603444.125000
Y_REG[29]=606873.625000
Y_REG[30]=603788.437500
Y_REG[31]=610252.625000
Y_REG[32]=595559.062500
Y_REG[33]=607746.812500
Y_REG[34]=602255.375000
Y_REG[35]=604680.875000
Y_REG[36]=600209.812500
Y_REG[37]=607527.375000
Y_REG[38]=607944.687500
Y_REG[39]=603448.375000
Y_REG[40]=603380.687500
Y_REG[41]=602064.187500
Y_REG[42]=603953.562500
Y_REG[43]=598390.937500
Y_REG[44]=605903.687500
Y_REG[45]=604125.875000
Y_REG[46]=598289.750000
Y_REG[47]=598549.375000
Y_REG[48]=598490.375000
Y_REG[49]=608219.500000
Y_REG[50]=606395.875000
Y_REG[51]=603176.312500
Y_REG[52]=606126.625000
Y_REG[53]=606217.312500
Y_REG[54]=619692.125000
Y_REG[55]=600690.875000
Y_REG[56]=605977.875000
Y_REG[57]=605184.937500
Y_REG[58]=597341.375000
Y_REG[59]=600821.812500
Y_REG[60]=607528.375000
Y_REG[61]=599984.187500
Y_REG[62]=605210.375000
Y_REG[63]=601756.687500
Y_REG[64]=603806.125000
Y_REG[65]=605213.625000
Y_REG[66]=620825.812500
Y_REG[67]=606013.500000
Y_REG[68]=595283.250000
Y_REG[69]=605817.812500
Y_REG[70]=598782.000000
Y_REG[71]=608899.000000
Y_REG[72]=602742.250000
Y_REG[73]=599514.937500
Y_REG[74]=594732.125000
Y_REG[75]=609195.562500
Y_REG[76]=603143.812500
Y_REG[77]=603098.187500
Y_REG[78]=605067.562500
Y_REG[79]=600547.687500
Y_REG[80]=608344.125000
Y_REG[81]=608598.062500
Y_REG[82]=609726.250000
Y_REG[83]=602660.125000
Y_REG[84]=597728.625000
Y_REG[85]=603003.375000
Y_REG[86]=595255.687500
Y_REG[87]=609272.625000
Y_REG[88]=605380.437500
Y_REG[89]=598538.062500
Y_REG[90]=608121.375000
Y_REG[91]=605227.062500
Y_REG[92]=601901.312500
Y_REG[93]=601655.875000
Y_REG[94]=602637.437500
Y_REG[95]=598466.687500
Y_REG[96]=607295.500000
Y_REG[97]=608570.500000
Y_REG[98]=609846.250000
Y_REG[99]=604842.875000

Appendix -8 Results of Regression Estimators obtained from simulation with 100 samples for Info-U.

Y_REG[0]=1287140.750000
Y_REG[1]=1290908.000000
Y_REG[2]=1306963.375000
Y_REG[3]=1280761.500000
Y_REG[4]=1275479.000000
Y_REG[5]=1276948.875000
Y_REG[6]=1272113.000000
Y_REG[7]=1271671.625000
Y_REG[8]=1292993.875000
Y_REG[9]=1280533.500000
Y_REG[10]=1297853.500000
Y_REG[11]=1272621.375000
Y_REG[12]=1300820.250000
Y_REG[13]=1285857.500000
Y_REG[14]=1282584.500000
Y_REG[15]=1288253.000000
Y_REG[16]=1298125.500000
Y_REG[17]=1301540.625000
Y_REG[18]=1269515.625000
Y_REG[19]=1285695.000000
Y_REG[20]=1246682.500000
Y_REG[21]=1278094.625000
Y_REG[22]=1282841.125000
Y_REG[23]=1296701.375000
Y_REG[24]=1287339.875000
Y_REG[25]=1276807.750000
Y_REG[26]=1283524.000000
Y_REG[27]=1273832.625000
Y_REG[28]=1279280.375000
Y_REG[29]=1291855.500000
Y_REG[30]=1285465.000000
Y_REG[31]=1292377.250000
Y_REG[32]=1265990.875000
Y_REG[33]=1291912.750000
Y_REG[34]=1280035.750000
Y_REG[35]=1283957.000000
Y_REG[36]=1273681.875000
Y_REG[37]=1289357.625000
Y_REG[38]=1291756.250000
Y_REG[39]=1280636.625000
Y_REG[40]=1283298.750000
Y_REG[41]=1278396.500000
Y_REG[42]=1280559.250000
Y_REG[43]=1270983.625000
Y_REG[44]=1285606.625000
Y_REG[45]=1284098.750000
Y_REG[46]=1270561.125000
Y_REG[47]=1270765.500000
Y_REG[48]=1272708.000000
Y_REG[49]=1292687.250000
Y_REG[50]=1290344.750000
Y_REG[51]=1282454.250000
Y_REG[52]=1286337.375000
Y_REG[53]=1290083.750000
Y_REG[54]=1320194.750000
Y_REG[55]=1276304.375000
Y_REG[56]=1289454.500000
Y_REG[57]=1285906.875000
Y_REG[58]=1264702.250000
Y_REG[59]=1271631.500000
Y_REG[60]=1291409.625000
Y_REG[61]=1279566.750000
Y_REG[62]=1285689.875000
Y_REG[63]=1280138.625000
Y_REG[64]=1285316.000000
Y_REG[65]=1286115.500000
Y_REG[66]=1323861.875000
Y_REG[67]=1290123.875000
Y_REG[68]=1266403.000000
Y_REG[69]=1285344.125000
Y_REG[70]=1275916.500000
Y_REG[71]=1290073.500000
Y_REG[72]=1279166.500000
Y_REG[73]=1272743.000000
Y_REG[74]=1264662.250000
Y_REG[75]=1295038.500000
Y_REG[76]=1281655.625000
Y_REG[77]=1281133.125000
Y_REG[78]=1285550.625000
Y_REG[79]=1275655.750000
Y_REG[80]=1291979.125000
Y_REG[81]=1295766.125000
Y_REG[82]=1296680.625000
Y_REG[83]=1282711.125000
Y_REG[84]=1269878.500000
Y_REG[85]=1281832.375000
Y_REG[86]=1265486.750000
Y_REG[87]=1295695.250000
Y_REG[88]=1285523.375000
Y_REG[89]=1273977.125000
Y_REG[90]=1298113.500000
Y_REG[91]=1287323.625000
Y_REG[92]=1280367.375000
Y_REG[93]=1280958.875000
Y_REG[94]=1284355.500000
Y_REG[95]=1272316.125000
Y_REG[96]=1290757.375000
Y_REG[97]=1293323.500000
Y_REG[98]=1296800.750000
Y_REG[99]=1288650.000000

Appendix - 9 Results of Separate Ratio Estimators obtained from simulation with 100 samples for Info-S.

Y_SEPRA[0]=919132.062500
Y_SERA[1]=1145824.625000
Y_SEPRA[2]=1145872.625000
Y_SEPRA[3]=996888.500000
Y_SEPRA[4]=953533.937500
Y_SEPRA[5]=1044577.750000
Y_SEPRA[6]=1166152.750000
Y_SEPRA[7]=970231.625000
Y_SEPRA[8]=1063561.875000
Y_SEPRA[9]=1144441.375000
Y_SEPRA[10]=1218851.125000
Y_SEPRA[11]=1077537.375000
Y_SEPRA[12]=941753.375000
Y_SEPRA[13]=1081902.250000
Y_SEPRA[14]=1145788.375000
Y_SEPRA[15]=961803.625000
Y_SEPRA[16]=990727.187500
Y_SEPRA[17]=1135999.625000
Y_SEPRA[18]=1041721.500000
Y_SEPRA[19]=1308258.750000
Y_SEPRA[20]=1013878.375000
Y_SEPRA[21]=1055638.500000
Y_SEPRA[22]=1154879.250000
Y_SEPRA[23]=974639.312500
Y_SEPRA[24]=1054279.500000
Y_SEPRA[25]=1060668.000000
Y_SEPRA[26]=1130435.625000
Y_SEPRA[27]=1181429.250000
Y_SEPRA[28]=1089103.875000
Y_SEPRA[29]=1080980.625000
Y_SEPRA[30]=1097615.250000
Y_SEPRA[31]=1137526.125000
Y_SEPRA[32]=1142105.625000
Y_SEPRA[33]=995865.000000
Y_SEPRA[34]=1096791.750000
Y_SEPRA[35]=994219.312500
Y_SEPRA[36]=987482.000000
Y_SEPRA[37]=1194929.250000
Y_SEPRA[38]=1107384.125000
Y_SEPRA[39]=1113074.875000
Y_SEPRA[40]=1146085.250000
Y_SEPRA[41]=939289.875000
Y_SEPRA[42]=1077373.125000
Y_SEPRA[43]=862808.937500
Y_SEPRA[44]=1153920.000000
Y_SEPRA[45]=1123376.625000
Y_SEPRA[46]=1066140.125000
Y_SEPRA[47]=1153695.375000
Y_SEPRA[48]=1005824.437500
Y_SEPRA[49]=1094939.000000
Y_SEPRA[50]=859902.687500
Y_SEPRA[51]=1059886.625000
Y_SEPRA[52]=1129426.250000
Y_SEPRA[53]=980046.625000
Y_SEPRA[54]=1028151.562500
Y_SEPRA[55]=1058291.375000
Y_SEPRA[56]=1124413.750000
Y_SEPRA[57]=1029781.812500
Y_SEPRA[58]=969211.500000
Y_SEPRA[59]=1002502.687500
Y_SEPRA[60]=1142661.625000
Y_SEPRA[61]=1116331.375000
Y_SEPRA[62]=1150607.500000
Y_SEPRA[63]=1045792.875000
Y_SEPRA[64]=905149.875000
Y_SEPRA[65]=1162456.500000
Y_SEPRA[66]=1054394.500000
Y_SEPRA[67]=1053945.500000
Y_SEPRA[68]=1021565.125000
Y_SEPRA[69]=1094684.875000
Y_SEPRA[70]=1113118.875000
Y_SEPRA[71]=1060901.125000
Y_SEPRA[72]=1121834.750000
Y_SEPRA[73]=1057242.750000
Y_SEPRA[74]=1079877.500000
Y_SEPRA[75]=1155529.375000
Y_SEPRA[76]=1085787.250000
Y_SEPRA[77]=1039942.562500
Y_SEPRA[78]=1090312.125000
Y_SEPRA[79]=926522.437500
Y_SEPRA[80]=1083481.000000
Y_SEPRA[81]=1105992.500000
Y_SEPRA[82]=1022969.875000
Y_SEPRA[83]=1045913.187500
Y_SEPRA[84]=1087951.125000
Y_SEPRA[85]=1111041.750000
Y_SEPRA[86]=1105451.125000
Y_SEPRA[87]=1141787.875000
Y_SEPRA[88]=1175201.250000
Y_SEPRA[89]=1122505.875000
Y_SEPRA[90]=1114776.375000
Y_SEPRA[91]=1003138.750000
Y_SEPRA[92]=1109681.000000
Y_SEPRA[93]=1222443.500000
Y_SEPRA[94]=1237152.500000
Y_SEPRA[95]=1163511.250000
Y_SEPRA[96]=1153306.250000
Y_SEPRA[97]=973217.312500
Y_SEPRA[98]=1166485.625000
Y_SEPRA[99]=1156250.875000

Appendix -10 Results of Separate Ratio Estimators obtained from simulation with 100 samples for Info-U.

Y_SEPRA[0]=1291402.250000
Y_SEPRA[1]=1431480.500000
Y_SEPRA[2]=1450631.250000
Y_SEPRA[3]=1244453.125000
Y_SEPRA[4]=1221974.125000
Y_SEPRA[5]=1348050.000000
Y_SEPRA[6]=1491954.875000
Y_SEPRA[7]=1218010.875000
Y_SEPRA[8]=1329875.000000
Y_SEPRA[9]=1428379.125000
Y_SEPRA[10]=1560745.625000
Y_SEPRA[11]=1355271.875000
Y_SEPRA[12]=1224069.125000
Y_SEPRA[13]=1369785.750000
Y_SEPRA[14]=1451465.125000
Y_SEPRA[15]=1218064.250000
Y_SEPRA[16]=1224471.000000
Y_SEPRA[17]=1479914.625000
Y_SEPRA[18]=1316507.250000
Y_SEPRA[19]=1641930.125000
Y_SEPRA[20]=1305016.375000
Y_SEPRA[21]=1332992.750000
Y_SEPRA[22]=1424380.125000
Y_SEPRA[23]=1301852.875000
Y_SEPRA[24]=1367281.625000
Y_SEPRA[25]=1326909.625000
Y_SEPRA[26]=1414269.500000
Y_SEPRA[27]=1512895.625000
Y_SEPRA[28]=1368901.000000
Y_SEPRA[29]=1353763.250000
Y_SEPRA[30]=1395902.125000
Y_SEPRA[31]=1459441.250000
Y_SEPRA[32]=1453847.875000
Y_SEPRA[33]=1290113.625000
Y_SEPRA[34]=1399497.375000
Y_SEPRA[35]=1271777.500000
Y_SEPRA[36]=1223277.125000
Y_SEPRA[37]=1509981.625000
Y_SEPRA[38]=1397161.500000
Y_SEPRA[39]=1389904.500000
Y_SEPRA[40]=1461701.875000
Y_SEPRA[41]=1187886.750000
Y_SEPRA[42]=1362339.125000
Y_SEPRA[43]=1111506.875000
Y_SEPRA[44]=1452315.250000
Y_SEPRA[45]=1436661.875000
Y_SEPRA[46]=1346948.250000
Y_SEPRA[47]=1439256.250000
Y_SEPRA[48]=1255100.375000
Y_SEPRA[49]=1399505.375000
Y_SEPRA[50]=1114371.250000
Y_SEPRA[51]=1353352.125000
Y_SEPRA[52]=1438997.750000
Y_SEPRA[53]=1250265.625000
Y_SEPRA[54]=1360038.750000
Y_SEPRA[55]=1359042.750000
Y_SEPRA[56]=1410793.000000
Y_SEPRA[57]=1318180.125000
Y_SEPRA[58]=1272174.625000
Y_SEPRA[59]=1291842.375000
Y_SEPRA[60]=1501500.000000
Y_SEPRA[61]=1396735.750000
Y_SEPRA[62]=1442108.000000
Y_SEPRA[63]=1373191.125000
Y_SEPRA[64]=1149966.875000
Y_SEPRA[65]=1484028.875000
Y_SEPRA[66]=1353351.125000
Y_SEPRA[67]=1337278.375000
Y_SEPRA[68]=1306645.000000
Y_SEPRA[69]=1336384.125000
Y_SEPRA[70]=1448260.125000
Y_SEPRA[71]=1386455.375000
Y_SEPRA[72]=1444225.875000
Y_SEPRA[73]=1348186.875000
Y_SEPRA[74]=1360850.000000
Y_SEPRA[75]=1479798.500000
Y_SEPRA[76]=1365814.750000
Y_SEPRA[77]=1347277.250000
Y_SEPRA[78]=1368519.250000
Y_SEPRA[79]=1153192.250000
Y_SEPRA[80]=1386790.750000
Y_SEPRA[81]=1417140.000000
Y_SEPRA[82]=1304173.375000
Y_SEPRA[83]=1345360.375000
Y_SEPRA[84]=1416332.500000
Y_SEPRA[85]=1389686.875000
Y_SEPRA[86]=1399191.375000
Y_SEPRA[87]=1409582.000000
Y_SEPRA[88]=1452041.375000
Y_SEPRA[89]=1409501.625000
Y_SEPRA[90]=1425643.000000
Y_SEPRA[91]=1274009.500000
Y_SEPRA[92]=1431877.250000
Y_SEPRA[93]=1570395.375000
Y_SEPRA[94]=1578299.500000
Y_SEPRA[95]=1582006.750000
Y_SEPRA[96]=1459656.125000
Y_SEPRA[97]=1198489.000000
Y_SEPRA[98]=1480787.000000
Y_SEPRA[99]=1466307.250000

Appendix -11 Results of Separate Regression Estimators obtained from simulation with 100 samples for Info-S.

Y_SEPREG[0]=1235690.000000
Y_SEPREG[1]=1303312.125000
Y_SEPREG[2]=1314335.500000
Y_SEPREG[3]=1247055.875000
Y_SEPREG[4]=1313392.750000
Y_SEPREG[5]=1236245.625000
Y_SEPREG[6]=1308602.125000
Y_SEPREG[7]=1281673.750000
Y_SEPREG[8]=1295049.250000
Y_SEPREG[9]=1269113.500000
Y_SEPREG[10]=1288677.750000
Y_SEPREG[11]=1268032.125000
Y_SEPREG[12]=1289066.625000
Y_SEPREG[13]=1258002.125000
Y_SEPREG[14]=1296432.500000
Y_SEPREG[15]=1307618.125000
Y_SEPREG[16]=1301829.250000
Y_SEPREG[17]=1296793.500000
Y_SEPREG[18]=1267847.500000
Y_SEPREG[19]=1285146.875000
Y_SEPREG[20]=1292321.250000
Y_SEPREG[21]=1259808.250000
Y_SEPREG[22]=1296853.250000
Y_SEPREG[23]=1282830.500000
Y_SEPREG[24]=1274724.125000
Y_SEPREG[25]=1272473.875000
Y_SEPREG[26]=1259651.250000
Y_SEPREG[27]=1294463.625000
Y_SEPREG[28]=1303880.250000
Y_SEPREG[29]=1268915.375000
Y_SEPREG[30]=1268206.875000
Y_SEPREG[31]=1308697.375000
Y_SEPREG[32]=1248191.250000
Y_SEPREG[33]=1276570.125000
Y_SEPREG[34]=1275039.500000
Y_SEPREG[35]=1332085.250000
Y_SEPREG[36]=1317025.375000
Y_SEPREG[37]=1289021.000000
Y_SEPREG[38]=1293345.250000
Y_SEPREG[39]=1286184.000000
Y_SEPREG[40]=1280788.250000
Y_SEPREG[41]=1274345.375000
Y_SEPREG[42]=1308062.000000
Y_SEPREG[43]=1269789.000000
Y_SEPREG[44]=1286842.375000
Y_SEPREG[45]=1278294.000000
Y_SEPREG[46]=1268805.500000
Y_SEPREG[47]=1278315.500000
Y_SEPREG[48]=1264402.125000
Y_SEPREG[49]=1291236.375000
Y_SEPREG[50]=1267340.500000
Y_SEPREG[51]=1266215.250000
Y_SEPREG[52]=1316758.000000
Y_SEPREG[53]=1264796.500000
Y_SEPREG[54]=1292180.625000
Y_SEPREG[55]=1278086.000000
Y_SEPREG[56]=1273513.750000
Y_SEPREG[57]=1286910.750000
Y_SEPREG[58]=1294385.375000
Y_SEPREG[59]=1337283.625000
Y_SEPREG[60]=1313582.375000
Y_SEPREG[61]=1254554.875000
Y_SEPREG[62]=1275445.125000
Y_SEPREG[63]=1262179.625000
Y_SEPREG[64]=1258793.500000
Y_SEPREG[65]=1286349.375000
Y_SEPREG[66]=1288612.500000
Y_SEPREG[67]=1262284.500000
Y_SEPREG[68]=1244802.625000
Y_SEPREG[69]=1306006.375000
Y_SEPREG[70]=1262316.375000
Y_SEPREG[71]=1322208.625000
Y_SEPREG[72]=1280708.500000
Y_SEPREG[73]=1284186.875000
Y_SEPREG[74]=1255200.750000
Y_SEPREG[75]=1272657.250000
Y_SEPREG[76]=1285434.875000
Y_SEPREG[77]=1287111.875000
Y_SEPREG[78]=1280081.250000
Y_SEPREG[79]=1282801.500000
Y_SEPREG[80]=1302772.500000
Y_SEPREG[81]=1272063.375000
Y_SEPREG[82]=1264942.000000
Y_SEPREG[83]=1251992.875000
Y_SEPREG[84]=1244176.125000
Y_SEPREG[85]=1280653.375000
Y_SEPREG[86]=1253156.375000
Y_SEPREG[87]=1290925.500000
Y_SEPREG[88]=1313680.375000
Y_SEPREG[89]=1236510.500000
Y_SEPREG[90]=1270459.750000
Y_SEPREG[91]=1268002.000000
Y_SEPREG[92]=1272565.000000
Y_SEPREG[93]=1256452.000000
Y_SEPREG[94]=1246662.875000
Y_SEPREG[95]=1264149.625000
Y_SEPREG[96]=1284374.500000
Y_SEPREG[97]=1277157.875000
Y_SEPREG[98]=1275761.250000
Y_SEPREG[99]=1265816.750000

Appendix - 12 Results of Separate Regression Estimators obtained from simulation with 100 samples for Info-U.

Y_SEPREG[0]=1274673.500000
Y_SEPREG[1]=1289335.875000
Y_SEPREG[2]=1305351.000000
Y_SEPREG[3]=1276190.250000
Y_SEPREG[4]=1286484.500000
Y_SEPREG[5]=1269585.250000
Y_SEPREG[6]=1271853.375000
Y_SEPREG[7]=1269989.750000
Y_SEPREG[8]=1296678.125000
Y_SEPREG[9]=1278449.250000
Y_SEPREG[10]=1297747.125000
Y_SEPREG[11]=1270148.250000
Y_SEPREG[12]=1291605.375000
Y_SEPREG[13]=1289537.500000
Y_SEPREG[14]=1289134.125000
Y_SEPREG[15]=1300001.875000
Y_SEPREG[16]=1299859.875000
Y_SEPREG[17]=1295882.000000
Y_SEPREG[18]=1272502.000000
Y_SEPREG[19]=1293795.000000
Y_SEPREG[20]=1258480.750000
Y_SEPREG[21]=1278293.875000
Y_SEPREG[22]=1286990.500000
Y_SEPREG[23]=1298361.875000
Y_SEPREG[24]=1287499.750000
Y_SEPREG[25]=1277741.375000
Y_SEPREG[26]=1288129.000000
Y_SEPREG[27]=1281379.625000
Y_SEPREG[28]=1276901.375000
Y_SEPREG[29]=1285983.750000
Y_SEPREG[30]=1290811.750000
Y_SEPREG[31]=1294980.000000
Y_SEPREG[32]=1266003.500000
Y_SEPREG[33]=1293341.250000
Y_SEPREG[34]=1281716.500000
Y_SEPREG[35]=1296605.375000
Y_SEPREG[36]=1286708.000000
Y_SEPREG[37]=1286724.500000
Y_SEPREG[38]=1293243.500000
Y_SEPREG[39]=1276796.250000
Y_SEPREG[40]=1285399.625000
Y_SEPREG[41]=1283856.375000
Y_SEPREG[42]=1283253.875000
Y_SEPREG[43]=1280589.000000
Y_SEPREG[44]=1282505.375000
Y_SEPREG[45]=1286013.750000
Y_SEPREG[46]=1265075.500000
Y_SEPREG[47]=1263668.125000
Y_SEPREG[48]=1273088.375000
Y_SEPREG[49]=1292210.875000
Y_SEPREG[50]=1283720.875000
Y_SEPREG[51]=1283118.875000
Y_SEPREG[52]=1288199.750000
Y_SEPREG[53]=1280552.875000
Y_SEPREG[54]=1315632.625000
Y_SEPREG[55]=1275255.625000
Y_SEPREG[56]=1298603.375000
Y_SEPREG[57]=1284934.250000
Y_SEPREG[58]=1271084.750000
Y_SEPREG[59]=1281321.000000
Y_SEPREG[60]=1296672.250000
Y_SEPREG[61]=1279927.125000
Y_SEPREG[62]=1281569.875000
Y_SEPREG[63]=1274452.750000
Y_SEPREG[64]=1278867.875000
Y_SEPREG[65]=1285376.000000
Y_SEPREG[66]=1320853.125000
Y_SEPREG[67]=1290925.000000
Y_SEPREG[68]=1274669.500000
Y_SEPREG[69]=1283176.500000
Y_SEPREG[70]=1269636.500000
Y_SEPREG[71]=1288213.625000
Y_SEPREG[72]=1275851.375000
Y_SEPREG[73]=1269952.125000
Y_SEPREG[74]=1267190.375000
Y_SEPREG[75]=1290852.375000
Y_SEPREG[76]=1283985.000000
Y_SEPREG[77]=1282517.000000
Y_SEPREG[78]=1285245.375000
Y_SEPREG[79]=1279894.125000
Y_SEPREG[80]=1292141.375000
Y_SEPREG[81]=1295685.500000
Y_SEPREG[82]=1285116.000000
Y_SEPREG[83]=1284968.875000
Y_SEPREG[84]=1260604.000000
Y_SEPREG[85]=1287823.250000
Y_SEPREG[86]=1267576.125000
Y_SEPREG[87]=1305098.125000
Y_SEPREG[88]=1293131.000000
Y_SEPREG[89]=1272392.500000
Y_SEPREG[90]=1297979.375000
Y_SEPREG[91]=1280844.250000
Y_SEPREG[92]=1287412.625000
Y_SEPREG[93]=1278142.500000
Y_SEPREG[94]=1282494.625000
Y_SEPREG[95]=1276088.625000
Y_SEPREG[96]=1290801.500000
Y_SEPREG[97]=1294795.375000
Y_SEPREG[98]=1296287.625000
Y_SEPREG[99]=1292287.250000

Appendix-13 Sampling sizes for four groups obtained for 100 samples.

n[0][1]=97
n[0][2]=102
n[0][3]=93
n[0][4]=108
n[1][1]=107
n[1][2]=93
n[1][3]=105
n[1][4]=95
n[2][1]=105
n[2][2]=99
n[2][3]=113
n[2][4]=83
n[3][1]=96
n[3][2]=109
n[3][3]=102
n[3][4]=93
n[4][1]=104
n[4][2]=102
n[4][3]=91
n[4][4]=103
n[5][1]=99
n[5][2]=110
n[5][3]=104
n[5][4]=87
n[6][1]=119
n[6][2]=90
n[6][3]=104
n[6][4]=87
n[7][1]=92
n[7][2]=107
n[7][3]=106
n[7][4]=95
n[8][1]=110
n[8][2]=95
n[8][3]=104
n[8][4]=91
n[9][1]=116
n[9][2]=93
n[9][3]=89
n[9][4]=102
n[10][1]=112
n[10][2]=95
n[10][3]=90
n[10][4]=103
n[11][1]=106
n[11][2]=97
n[11][3]=90
n[11][4]=107
n[12][1]=93
n[12][2]=109
n[12][3]=97
n[12][4]=101
n[13][1]=107
n[13][2]=101
n[13][3]=91
n[13][4]=101
n[14][1]=106
n[14][2]=101
n[14][3]=89
n[14][4]=104
n[15][1]=99
n[15][2]=96
n[15][3]=103
n[15][4]=102
n[16][1]=97
n[16][2]=96
n[16][3]=103
n[16][4]=104
n[17][1]=107
n[17][2]=109
n[17][3]=99
n[17][4]=85
n[18][1]=106
n[18][2]=110
n[18][3]=83
n[18][4]=101
n[19][1]=121
n[19][2]=88
n[19][3]=104
n[19][4]=87
n[20][1]=104
n[20][2]=98
n[20][3]=91
n[20][4]=107
n[21][1]=107
n[21][2]=102
n[21][3]=92
n[21][4]=99
n[22][1]=107
n[22][2]=102
n[22][3]=101
n[22][4]=90
n[23][1]=101
n[23][2]=111
n[23][3]=91
n[23][4]=97
n[24][1]=107
n[24][2]=107
n[24][3]=97
n[24][4]=89
n[25][1]=98
n[25][2]=110
n[25][3]=87
n[25][4]=105
n[26][1]=106
n[26][2]=103
n[26][3]=91
n[26][4]=100
n[27][1]=114
n[27][2]=102
n[27][3]=95
n[27][4]=89
n[28][1]=107
n[28][2]=96
n[28][3]=99
n[28][4]=98
n[29][1]=107
n[29][2]=104
n[29][3]=85
n[29][4]=104
n[30][1]=108
n[30][2]=107
n[30][3]=95
n[30][4]=90
n[31][1]=113
n[31][2]=89
n[31][3]=88
n[31][4]=110
n[32][1]=112
n[32][2]=95
n[32][3]=90
n[32][4]=103
n[33][1]=98
n[33][2]=106
n[33][3]=90
n[33][4]=106
n[34][1]=110
n[34][2]=84
n[34][3]=94
n[34][4]=112
n[35][1]=104
n[35][2]=100
n[35][3]=86
n[35][4]=110
n[36][1]=96
n[36][2]=105
n[36][3]=99
n[36][4]=100
n[37][1]=113
n[37][2]=95
n[37][3]=85
n[37][4]=107
n[38][1]=104
n[38][2]=104
n[38][3]=95
n[38][4]=97
n[39][1]=110
n[39][2]=94
n[39][3]=98
n[39][4]=98
n[40][1]=114
n[40][2]=98
n[40][3]=95
n[40][4]=93
n[41][1]=95
n[41][2]=101
n[41][3]=101
n[41][4]=103

Appendix-13 Continued

n[42][1]=105
n[42][2]=93
n[42][3]=96
n[42][4]=106
n[43][1]=85
n[43][2]=115
n[43][3]=101
n[43][4]=99
n[44][1]=108
n[44][2]=92
n[44][3]=100
n[44][4]=100
n[45][1]=106
n[45][2]=103
n[45][3]=104
n[45][4]=87
n[46][1]=105
n[46][2]=89
n[46][3]=109
n[46][4]=97
n[47][1]=106
n[47][2]=105
n[47][3]=90
n[47][4]=99
n[48][1]=102
n[48][2]=96
n[48][3]=97
n[48][4]=105
n[49][1]=106
n[49][2]=107
n[49][3]=92
n[49][4]=95
n[50][1]=90
n[50][2]=99
n[50][3]=104
n[50][4]=107
n[51][1]=102
n[51][2]=98
n[51][3]=96
n[51][4]=104
n[52][1]=117
n[52][2]=89
n[52][3]=95
n[52][4]=99
n[53][1]=93
n[53][2]=94
n[53][3]=109
n[53][4]=104
n[54][1]=103
n[54][2]=106
n[54][3]=112
n[54][4]=79
n[55][1]=107
n[55][2]=101
n[55][3]=95
n[55][4]=97
n[56][1]=110
n[56][2]=96
n[56][3]=94
n[56][4]=100
n[57][1]=98
n[57][2]=99
n[57][3]=102
n[57][4]=101
n[58][1]=96
n[58][2]=97
n[58][3]=96
n[58][4]=111
n[59][1]=101
n[59][2]=101
n[59][3]=100
n[59][4]=98
n[60][1]=112
n[60][2]=103
n[60][3]=101
n[60][4]=84
n[61][1]=109
n[61][2]=106
n[61][3]=88
n[61][4]=97
n[62][1]=114
n[62][2]=103
n[62][3]=86
n[62][4]=97
n[63][1]=111
n[63][2]=98
n[63][3]=86
n[63][4]=105
n[64][1]=97
n[64][2]=98
n[64][3]=96
n[64][4]=109
n[65][1]=112
n[65][2]=98
n[65][3]=106
n[65][4]=84
n[66][1]=106
n[66][2]=106
n[66][3]=85
n[66][4]=103
n[67][1]=108
n[67][2]=97
n[67][3]=89
n[67][4]=106
n[68][1]=101
n[68][2]=93
n[68][3]=102
n[68][4]=104
n[69][1]=106
n[69][2]=96
n[69][3]=101
n[69][4]=97
n[70][1]=106
n[70][2]=112
n[70][3]=90
n[70][4]=92
n[71][1]=106
n[71][2]=98
n[71][3]=105
n[71][4]=91
n[72][1]=113
n[72][2]=98
n[72][3]=91
n[72][4]=98
n[73][1]=105
n[73][2]=96
n[73][3]=101
n[73][4]=98
n[74][1]=114
n[74][2]=97
n[74][3]=93
n[74][4]=96
n[75][1]=107
n[75][2]=109
n[75][3]=91
n[75][4]=93
n[76][1]=105
n[76][2]=104
n[76][3]=93
n[76][4]=98
n[77][1]=104
n[77][2]=105
n[77][3]=102
n[77][4]=89
n[78][1]=112
n[78][2]=96
n[78][3]=95
n[78][4]=97
n[79][1]=89
n[79][2]=107
n[79][3]=100
n[79][4]=104
n[80][1]=113
n[80][2]=100
n[80][3]=88
n[80][4]=99
n[81][1]=111
n[81][2]=96
n[81][3]=92
n[81][4]=101
n[82][1]=106
n[82][2]=91
n[82][3]=98
n[82][4]=105
n[83][1]=105
n[83][2]=95
n[83][3]=103
n[83][4]=97

Appendix-13 Continued

n[84][1]=111
n[84][2]=98
n[84][3]=106
n[84][4]=85
n[85][1]=106
n[85][2]=105
n[85][3]=84
n[85][4]=105
n[86][1]=109
n[86][2]=84
n[86][3]=100
n[86][4]=107
n[87][1]=107
n[87][2]=97
n[87][3]=96
n[87][4]=100
n[88][1]=112
n[88][2]=103
n[88][3]=87
n[88][4]=98
n[89][1]=106

n[89][2]=106
n[89][3]=95
n[89][4]=93
n[90][1]=106
n[90][2]=96
n[90][3]=96
n[90][4]=102
n[91][1]=96
n[91][2]=102
n[91][3]=98
n[91][4]=104
n[92][1]=102
n[92][2]=110
n[92][3]=84
n[92][4]=104
n[93][1]=117
n[93][2]=94
n[93][3]=86
n[93][4]=103
n[94][1]=122
n[94][2]=90

n[94][3]=88
n[94][4]=100
n[95][1]=118
n[95][2]=106
n[95][3]=93
n[95][4]=83
n[96][1]=113
n[96][2]=96
n[96][3]=94
n[96][4]=97
n[97][1]=99
n[97][2]=104
n[97][3]=97
[97][4]=100
n[98][1]=112
n[98][2]=90
n[98][3]=91
n[98][4]=107
n[99][1]=112
n[99][2]=95
n[99][3]=99
n[99][4]=94

Appendix-14 Sampling sizes for four groups obtained for 100 samples.

m[0][1]=76
m[0][2]=83
m[0][3]=68
m[0][4]=78
m[1][1]=94
m[1][2]=78
m[1][3]=93
m[1][4]=80
m[2][1]=93
m[2][2]=83
m[2][3]=93
m[2][4]=72
m[3][1]=83
m[3][2]=91
m[3][3]=87
m[3][4]=81
m[4][1]=79
m[4][2]=90
m[4][3]=80
m[4][4]=84
m[5][1]=87
m[5][2]=90
m[5][3]=88
m[5][4]=72
m[6][1]=97
m[6][2]=79
m[6][3]=85
m[6][4]=72
m[7][1]=81
m[7][2]=91
m[7][3]=94
m[7][4]=76
m[8][1]=87
m[8][2]=78
m[8][3]=95
m[8][4]=81
m[9][1]=95
m[9][2]=84
m[9][3]=78
m[9][4]=83
m[10][1]=100
m[10][2]=77
m[10][3]=73
m[10][4]=85
m[11][1]=90
m[11][2]=86
m[11][3]=74
m[11][4]=90
m[12][1]=77
m[12][2]=88
m[12][3]=84
m[12][4]=83
m[13][1]=90
m[13][2]=86
m[13][3]=75
m[13][4]=85
m[14][1]=95
m[14][2]=82
m[14][3]=71
m[14][4]=87
m[15][1]=79
m[15][2]=84
m[15][3]=84
m[15][4]=92
m[16][1]=81
m[16][2]=90
m[16][3]=87
m[16][4]=89
m[17][1]=93
m[17][2]=91
m[17][3]=80
m[17][4]=66
m[18][1]=87
m[18][2]=100
m[18][3]=70
m[18][4]=83
m[19][1]=107
m[19][2]=71
m[19][3]=91
m[19][4]=74
m[20][1]=86
m[20][2]=82
m[20][3]=75
m[20][4]=88
m[21][1]=88
m[21][2]=85
m[21][3]=77
m[21][4]=87
m[22][1]=95
m[22][2]=87
m[22][3]=87
m[22][4]=77
m[23][1]=80
m[23][2]=89
m[23][3]=71
m[23][4]=79
m[24][1]=87
m[24][2]=97
m[24][3]=79
m[24][4]=72
m[25][1]=88
m[25][2]=93
m[25][3]=71
m[25][4]=88
m[26][1]=94
m[26][2]=91
m[26][3]=75
m[26][4]=85
m[27][1]=98
m[27][2]=88
m[27][3]=73
m[27][4]=74
m[28][1]=90
m[28][2]=81
m[28][3]=82
m[28][4]=86
m[29][1]=89
m[29][2]=96
m[29][3]=69
m[29][4]=91
m[30][1]=91
m[30][2]=95
m[30][3]=80
m[30][4]=74
m[31][1]=93
m[31][2]=78
m[31][3]=75
m[31][4]=89
m[32][1]=96
m[32][2]=75
m[32][3]=78
m[32][4]=85
m[33][1]=82
m[33][2]=82
m[33][3]=75
m[33][4]=91
m[34][1]=91
m[34][2]=75
m[34][3]=79
m[34][4]=88
m[35][1]=82
m[35][2]=86
m[35][3]=75
m[35][4]=90
m[36][1]=82
m[36][2]=88
m[36][3]=85
m[36][4]=87
m[37][1]=98
m[37][2]=82

Appendix-14 Continued

m[37][3]=73
m[37][4]=86
m[38][1]=91
m[38][2]=84
m[38][3]=79
m[38][4]=84
m[39][1]=92
m[39][2]=83
m[39][3]=81
m[39][4]=84
m[40][1]=95
m[40][2]=86
m[40][3]=76
m[40][4]=78
m[41][1]=78
m[41][2]=89
m[41][3]=87
m[41][4]=85
m[42][1]=89
m[42][2]=81
m[42][3]=78
m[42][4]=90
m[43][1]=72
m[43][2]=89
m[43][3]=89
m[43][4]=84
m[44][1]=95
m[44][2]=79
m[44][3]=84
m[44][4]=81
m[45][1]=93
m[45][2]=88
m[45][3]=88
m[45][4]=69
m[46][1]=89
m[46][2]=77
m[46][3]=93
m[46][4]=78
m[47][1]=96
m[47][2]=85
m[47][3]=77
m[47][4]=82
m[48][1]=84
m[48][2]=84
m[48][3]=81
m[48][4]=93
m[49][1]=90
m[49][2]=89
m[49][3]=73
m[49][4]=81
m[50][1]=71
m[50][2]=83
m[50][3]=91
m[50][4]=87
m[51][1]=88
m[51][2]=82
m[51][3]=75
m[51][4]=87
m[52][1]=93
m[52][2]=79
m[52][3]=79
m[52][4]=84
m[53][1]=81
m[53][2]=77
m[53][3]=87
m[53][4]=89
m[54][1]=83
m[54][2]=93
m[54][3]=91
m[54][4]=65
m[55][1]=88
m[55][2]=82
m[55][3]=78
m[55][4]=83
m[56][1]=93
m[56][2]=83
m[56][3]=79
m[56][4]=83
m[57][1]=85
m[57][2]=81
m[57][3]=83
m[57][4]=86
m[58][1]=81
m[58][2]=73
m[58][3]=81
m[58][4]=96
m[59][1]=83
m[59][2]=86
m[59][3]=87
m[59][4]=76
m[60][1]=94
m[60][2]=84
m[60][3]=77
m[60][4]=71
m[61][1]=93
m[61][2]=92
m[61][3]=75
m[61][4]=85
m[62][1]=95
m[62][2]=95
m[62][3]=73
m[62][4]=79
m[63][1]=87
m[63][2]=86
m[63][3]=67
m[63][4]=89
m[64][1]=75
m[64][2]=89
m[64][3]=81
m[64][4]=94
m[65][1]=96
m[65][2]=79
m[65][3]=88
m[65][4]=74
m[66][1]=85
m[66][2]=91
m[66][3]=72
m[66][4]=87
m[67][1]=87
m[67][2]=83
m[67][3]=76
m[67][4]=91
m[68][1]=86
m[68][2]=81
m[68][3]=78
m[68][4]=87
m[69][1]=90
m[69][2]=82
m[69][3]=90
m[69][4]=86
m[70][1]=93
m[70][2]=96
m[70][3]=74
m[70][4]=72
m[71][1]=87
m[71][2]=80
m[71][3]=91
m[71][4]=72
m[72][1]=93
m[72][2]=84
m[72][3]=75
m[72][4]=81
m[73][1]=88
m[73][2]=76
m[73][3]=85
m[73][4]=86
m[74][1]=91
m[74][2]=88
m[74][3]=75
m[74][4]=82
m[75][1]=95
m[75][2]=95
m[75][3]=70

Appendix-14 Continued

m[75][4]=76
m[76][1]=90
m[76][2]=82
m[76][3]=84
m[76][4]=80
m[77][1]=86
m[77][2]=89
m[77][3]=84
m[77][4]=71
m[78][1]=90
m[78][2]=86
m[78][3]=80
m[78][4]=83
m[79][1]=77
m[79][2]=90
m[79][3]=86
m[79][4]=90
m[80][1]=89
m[80][2]=88
m[80][3]=71
m[80][4]=86
m[81][1]=91
m[81][2]=76
m[81][3]=82
m[81][4]=84
m[82][1]=84
m[82][2]=78
m[82][3]=79
m[82][4]=92
m[83][1]=87
m[83][2]=77
m[83][3]=86
m[83][4]=80
m[84][1]=91
m[84][2]=81
m[84][3]=92
m[84][4]=67
m[85][1]=92
m[85][2]=84
m[85][3]=75
m[85][4]=92
m[86][1]=93
m[86][2]=71
m[86][3]=81
m[86][4]=91
m[87][1]=94
m[87][2]=84
m[87][3]=78
m[87][4]=87
m[88][1]=97
m[88][2]=87
m[88][3]=77
m[88][4]=84
m[89][1]=94
m[89][2]=92
m[89][3]=83
m[89][4]=72
m[90][1]=92
m[90][2]=83
m[90][3]=80
m[90][4]=83
m[91][1]=83
m[91][2]=89
m[91][3]=82
m[91][4]=83
m[92][1]=92
m[92][2]=94
m[92][3]=67
m[92][4]=85
m[93][1]=102
m[93][2]=75
m[93][3]=73
m[93][4]=85
m[94][1]=103
m[94][2]=75
m[94][3]=75
m[94][4]=83
m[95][1]=97
m[95][2]=83
m[95][3]=76
m[95][4]=63
m[96][1]=95
m[96][2]=83
m[96][3]=79
m[96][4]=81
m[97][1]=80
m[97][2]=93
m[97][3]=84
m[97][4]=88
m[98][1]=96
m[98][2]=71
m[98][3]=81
m[98][4]=88
m[99][1]=96
m[99][2]=81
m[99][3]=77
m[99][4]=82

Appendix – 15 Size of Responding Elements for each of 100 samples.

M[0]=305
M[1]=345
M[2]=341
M[3]=342
M[4]=333
M[5]=337
M[6]=333
M[7]=342
M[8]=341
M[9]=340
M[10]=335
M[11]=340
M[12]=332
M[13]=336
M[14]=335
M[15]=339
M[16]=347
M[17]=330
M[18]=340
M[19]=343
M[20]=331
M[21]=337
M[22]=346
M[23]=319
M[24]=335
M[25]=340
M[26]=345
M[27]=333
M[28]=339
M[29]=345
M[30]=340
M[31]=335
M[32]=334
M[33]=330
M[34]=333
M[35]=333
M[36]=342
M[37]=339
M[38]=338
M[39]=340
M[40]=335
M[41]=339
M[42]=338
M[43]=334
M[44]=339
M[45]=338
M[46]=337
M[47]=340
M[48]=342
M[49]=333
M[50]=332
M[51]=332
M[52]=335
M[53]=334
M[54]=332
M[55]=331
M[56]=338
M[57]=335
M[58]=331
M[59]=332
M[60]=326
M[61]=345
M[62]=342
M[63]=329
M[64]=339
M[65]=337
M[66]=335
M[67]=337
M[68]=332
M[69]=348
M[70]=335
M[71]=330
M[72]=333
M[73]=335
M[74]=336
M[75]=336
M[76]=336
M[77]=330
M[78]=339
M[79]=343
M[80]=334
M[81]=333
M[82]=333
M[83]=330
M[84]=331
M[85]=343
M[86]=336
M[87]=343
M[88]=345
M[89]=341
M[90]=338
M[91]=337
M[92]=338
M[93]=335
M[94]=336
M[95]=319
M[96]=338
M[97]=345
M[98]=336
M[99]=336