UNIVERSITY OF SOUTHAMPTON

FACULTY OF MATHEMATICAL

STUDIES

MATHEMATICS

ROBUST ESTIMATION OF THE REGRESSION COEFFICIENTS IN COMPLEX SURVEYS

BY

EDWARD GACHANGI NJENGA

Thesis submitted for the degree of Doctor of Philosophy
November, 1990.

ACKNOWLEDGEMENTS

I would like to extend my sincere gratitude to Professor T.M.F Smith for supervising this piece of research and to Dr Chris Skinner,Dr D.J.Holmes and Mr Ray Brown for their invaluable help throughout my research period.

I cannot forget my dear wife Nyandia and my two sons Njenga and Rukwaro for their patience,tolerance and encouragement throughout my stay here in Southampton.

Lastly I thank my employer Kenyatta university,Nairobi ,Kenya for enabling me to undertake the project and the British Council for funding it.

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MATHEMATICAL STUDIES

MATHEMATICS

DOCTOR OF PHILOSOPHY

ROBUST ESTIMATION OF THE REGRESSION COEFFICIENTS IN COMPLEX SURVEYS

BY EDWARD GACHANGI NJENGA

This thesis is concerned with the search for robust efficient procedures for estimating the regression coefficient in the marginal distribution of the survey variables for data collected from complex surveys.

Parametric model based procedures which take into account the structure of the population by assuming a parametric model are found to be very sensitive to the violations of these underlying parametric model assumptions.However these procedures perform very well when the parametric model assumptions hold. Randomization based estimators ,which takes into account the population structure through the selection probabilities,are found to be robust unconditionally but their conditi o nal and efficiency properties may be poor in some circumstances.

We propose nonparametric model based procedures which do not make any explicit assumptions about the distribution describing the population structure.One nonparametric procedure, namely the Nadaraya-Watson kernel estimator of the regression coefficient is the most efficient and robust when the structure of the population is unknown.However the estimator is biased when the population is linear and homoscedastic.

The validity of the theoretical results was assessed in a series of simulation studies based on a variety of stratified sampling schemes.

CONTENTS

Page number

CHAPTER 1

INTRODUCTION

1.1 PROBLEM OF ANALYSING DATA FROM COMPLEX SURVEYS

Most data collected from the field are multivariate in nature and usually involve complex designs including stratification and multistage sampling.Once the data has been collected they can be analysed either descriptively,using simple statistics like the means,totals etc.or causally using more complex statistics.The traditional approach to inference from survey data has concentrated on the descriptive analysis of the data using these simple statistics to estimate parameters in the finite population.This type of inference is called inference to *descriptive* targets or descriptive inference.

However one may be interested in establishing relationships between sets of variables rather than just simple description.Thus the interest is to explain the data rather than to describe it.This type of inference is called inference to *analytic* targets or analytic inference.Regression analysis is one multivariate method for studying such relationships. McKennell [1969],for example,describes a survey of residents around Heathrow airport which measured the annoyance due to noise caused by aircraft.The population was stratified on the basis of number and loudness of the aircraft.The first objective was to describe the data using a descriptive statistic,the mean noise annoyance. The second objective was to use the survey results to predict the noise annoyance around other possible sites for constructing a third London airport.The first objective is achieved using the well known randomization methods.These methods involves the selection of a sample using a known sampling scheme like simple random sampling.Since the only source of randomization is the sampling scheme,the randomization distribution is determined by the sampling scheme over all the possible samples which can be selected from the finite population.Randomization methods are based on this distribution averaged over all the possible samples and the randomization based estimators are chosen using properties such as p-unbiasedness,which will be defined later in the chapter.The second objective is to explain the relationship between the units in the survey data and those in the population around a proposed third airport.We see that in this case the parameter of interest is not in the survey population but in another population most likely different from the survey population,and so is an example of analytic inference.

If the parameters in the survey population are related to the parameters in the other population through a superpopulation model , then these parameters are called superpopulation parameters.Since data collected in surveys involves the use of complex designs related to the finite population,how does one adjust for the effects of the design when estimating superpopulation parameters?Most analysts of survey data avoid this problem by assuming that the data are generated as independent observations from a common distribution, such as the multivariate normal distribution.However since complex designs use stratification and clustering this assumption may not be true as these data may not be independent.see for example Skinner, Holt and Smith [1989]. Most standard statistical packages like BMDP,SPSS and GLIM which are widely used for the analysis of data from complex surveys do not take into account the structure in the population, especially if there are clusters in the population. Stratification usually can be built into these packages.For clustered populations the units within each cluster are not independent hence the use of these standard statistical packages, which ignore the presence of clusters in the population will lead to inflation of the true variance relative to the ordinary least squares variance see Kish and Frankel [1974]. The dilemma is that classical methods of multivariate analysis assume identically and independently distributed (i.i.d) samples from the superpopulation models and classical methods of survey analysis employ probability weights, do not assume that the observations are i.i.d. and provide only descriptive inferences.

This thesis will address the following questions:

(i) Can randomization methods be used to estimate superpopulation parameters?

(ii) How can model based procedures be adapted to take into account the information in the design e.g stratification and size measures?

(iii) How robust are the model based methods to the violations of the model assumptions?

(iv) Is it possible to find more robust estimators than the already existing ones?

Kish and Frankel [1974],Binder [1983] suggest solutions to (i).Holt et.al [1980] suggested solutions to (ii) and Nathan and Holt[1980] suggest a compromise solution. Royall [1976b],Skinner [1981] suggested predictors of the finite population mean and variance which takes into account the presence of clusters in the population.Pferfferman and Holmes [1985],Holmes [1987] have investigated the robustness of model based procedures and found that they are not robust to the violations of the model assumptions.

A massive literature exists examining various approaches to statistical inference for analysing sample survey data.Example of some of these approaches are Bayesian,Model based,randomisation etc.In our literature review we will develop a global model which embraces all these approaches to statistical inference.We will show that under some specific assumptions we can get each particular aproach as a special case of the global model.

1.2 GENERAL FRAMEWORK

Through out this work we will denote matrices by capital letters, vectors by lower case letters and parameters by greek letters. Let U denote a set of N identifiable units of a finite population labelled 1...N. Associated with the i^{th} unit in the finite population is a p×1 vector of unknown values to be measured in the survey, called survey values, denoted by $\underset{\sim}{y}_i=[y_{1i}..y_{pi}]^T$ and a qx1 vector of values assumed known for each member of the finite population, called design values, and denoted by $\underset{\sim}{z}_i=[z_{1i}..z_{qi}]^T$. We will denote the Nxp and Nxq dimensional matrices of the survey and design values by $\underset{\sim}{Y}_U=[\underset{\sim}{y}_1..\underset{\sim}{y}_N]^T$ and $\underset{\sim}{Z}_U=[\underset{\sim}{z}_1..\underset{\sim}{z}_N]^T$ respectively.

For example let i denote a factory producing a certain item, and let $\underset{\sim}{y}_i=(y_{1i},y_{2i})^T$ denotes a 2x1 vector representing the output of an item in the i^{th} factory and $\underset{\sim}{z}_i=(z_{1i},z_{2i})^T$ denotes a 2x1 vector of design values, where;

y_{1i}=value of the output of the item in the i^{th} factory.

y_{2i}=volume of the output of the item in the i^{th} factory.

z_{1i}=size of the i^{th} factory.

z_{2i}=last year's turnover of the item in the i^{th} factory.

To distinguish between a random variable and its realized value we will denote the random variable by the superscript ~. Let $\underset{\sim}{\tilde{Y}}_U=[\underset{\sim}{\tilde{y}}_1..\underset{\sim}{\tilde{y}}_N]^T$ denote a N×p dimensional matrix of survey variables and $\underset{\sim}{\tilde{Z}}_U=[\underset{\sim}{\tilde{z}}_1..\underset{\sim}{\tilde{z}}_N]^T$ denote a N×q dimensional matrix of design variables. In a model based approach the finite population matrix of values $\underset{\sim}{W}=(\underset{\sim}{Y}_U,\underset{\sim}{Z}_U)$ are treated as the realized values of the random matrix $\underset{\sim}{\tilde{W}}=(\underset{\sim}{\tilde{Y}}_U,\underset{\sim}{\tilde{Z}}_U)$. The degenerate case when $\underset{\sim}{\tilde{W}}=\underset{\sim}{W}$ with probability one represents the situation when the finite population values are fixed. The randomization approach to inference is based on this degenerate case of the random model. We assume that the vectors of variables $\tilde{y}_i$ and $\tilde{z}_i$ are distributed as random vectors $\tilde{y}=[\tilde{y}_1..\tilde{y}_p]^T$ and $\tilde{z}=[\tilde{z}_1..\tilde{z}_q]^T$,with a common probability distribution $f(y;\theta)$, where θ is an unknown parameter indexing the marginal distribution of $\tilde{y}$ and $f(z;\phi)$,where ϕ is an unknown parameter indexing the marginal distribution of $\tilde{z}$.The objective is to construct a model for $\tilde{Y}_U$ so that it accomodates all the available background information about the population structure e.g stratification,clustering etc.Inference is then based on the probability distribution specified by the model.

1.2.1 SAMPLE SELECTION

From the finite population U a sample of size n is selected using a known selection scheme,we denote the sample labels by s and the nonsample labels by $\bar{s}$.Let Y_s and Z_s denote the matrices of survey values and design values of dimension nxp and nxq included in the sample respectively.Similarly, the matrices $Y_{\bar{s}}$ and $Z_{\bar{s}}$ of dimension (N-n)xp and (N-n)xq denote the survey and design values not included in the sample respectively, then $Y_U=(Y_s^T,Y_{\bar{s}}^T)^T$ and $Z_U=(Z_s^T,Z_{\bar{s}}^T)^T$.

We summarize the notation given in this section in the following table:

TABLE 1.1 Summary of the notations to be used in the thesis.

RANDOM VARIABLE	REALIZED VALUE
$\tilde{\underset{\sim}{y}}=[\tilde{y}_1.....\tilde{y}_p]^T_{px1}$ vector	$\underset{\sim}{y}=[y_1.....y_p]^T_{px1}$ vector
$\tilde{\underset{\sim}{z}}=[\tilde{z}_1.....\tilde{z}_q]^T_{qx1}$ vector	$\underset{\sim}{z}=[z_1.....z_q]^T_{qx1}$ vector
$\tilde{\underset{\sim}{y}}_i=[\tilde{y}_{1i}....\tilde{y}_{pi}]^T_{px1}$ vector	$\underset{\sim}{y}_i=[y_{1i}....y_{pi}]^T_{px1}$ vector
$\tilde{\underset{\sim}{z}}_i=[\tilde{z}_{1i}....\tilde{z}_{qi}]^T_{qx1}$ vector	$\underset{\sim}{z}_i=[z_{1i}....z_{qi}]^T_{qx1}$ vector
$\tilde{\underset{\sim}{Y}}_U=[\tilde{\underset{\sim}{y}}_1.....\tilde{\underset{\sim}{y}}_N]^T_{Nxp}$ matrix	$\underset{\sim}{Y}_U=[\underset{\sim}{y}_1.....\underset{\sim}{y}_N]^T_{Nxp}$ matrix
$\tilde{\underset{\sim}{Z}}_U=[\tilde{\underset{\sim}{z}}_1.....\tilde{\underset{\sim}{z}}_N]^T_{Nxq}$ matrix	$\underset{\sim}{Z}_U=[\underset{\sim}{z}_1.....\underset{\sim}{z}_N]^T_{Nxq}$ matrix
$\tilde{\underset{\sim}{Y}}_S=[\tilde{\underset{\sim}{y}}_1.....\tilde{\underset{\sim}{y}}_n]^T_{nxp}$ matrix	$\underset{\sim}{Y}_S=[\underset{\sim}{y}_1.....\underset{\sim}{y}_n]^T_{nxp}$ matrix
$\tilde{\underset{\sim}{Z}}_S=[\tilde{\underset{\sim}{z}}_1.....\tilde{\underset{\sim}{z}}_n]^T_{nxq}$ matrix	$\underset{\sim}{Z}_S=[\underset{\sim}{z}_1.....\underset{\sim}{z}_n]^T_{nxq}$ matrix
$\tilde{\underset{\sim}{Y}}_{\bar{S}}=[\tilde{\underset{\sim}{y}}_1...\tilde{\underset{\sim}{y}}_{N-n}]^T_{N-nxp}$ matrix	$\underset{\sim}{Y}_{\bar{S}}=[\underset{\sim}{y}_1...\underset{\sim}{y}_{N-n}]^T_{N-nxp}$ matrix
$\tilde{\underset{\sim}{Z}}_{\bar{S}}=[\tilde{\underset{\sim}{z}}_1...\tilde{\underset{\sim}{z}}_{N-n}]^T_{N-nxp}$ matrix $\tilde{s}$ (sample)	$\underset{\sim}{Z}_{\bar{S}}=[\underset{\sim}{z}_1...\underset{\sim}{z}_{N-n}]^T_{N-nxp}$ matrix s

In the next section we will construct a global model for $\tilde{\underset{\sim}{Y}}_U$ by modelling all the possible mechanisms which generate the population values and the sample values.

1.3 GLOBAL MODEL

In general the joint distribution of the survey variable $\tilde{\underset{\sim}{Y}}_U$ and the design variable $\tilde{\underset{\sim}{Z}}_U$ before selection can be written as

$$f(\underset{\sim}{Y}_U,\underset{\sim}{Z}_U,;\underset{\sim}{\lambda},\underset{\sim}{\phi})= f(\underset{\sim}{Y}_U|\ \underset{\sim}{Z}_U;\underset{\sim}{\lambda})\ f(\underset{\sim}{Z}_U;\underset{\sim}{\phi}), \qquad [1.1]$$

where $\underset{\sim}{\phi}$ and $\underset{\sim}{\lambda}$ are unknown but distinct parameters indexing this joint distribution.

Let $I(s)=[I_1(s)......I_N(s)]$ denote the indicator function which identifies those units of U which are included in the sample s

$$I_i(s)= \begin{cases} 1 & \text{if } i\varepsilon s, \\ 0 & \text{otherwise.} \end{cases}$$

In general the rule of evaluating I(s),i.e selecting the sample s, could depend on the values of the design variable and the values of the survey variable .For example in Quota sampling interviewers have the freedom to choose the individual whom to interview and may base their decision on the values $\underset{\sim}{Y}_U$; see Smith [1983]. Also in non response,the respondent chooses whether to reply or not,this again could be related to the values of the survey variable.For example there is usually a higher nonreponse to questions relating to income,level of taxation etc,in the high income group than in the low income group;see Little [1982].We represent this general mechanism of selecting the sample as;

$$f(s|\underset{\sim}{Z}_U,\underset{\sim}{Y}_U;\underset{\sim}{\delta}), \qquad [1.2]$$

where $\underset{\sim}{\delta}$ are unknown parameters indexing this conditional distribution,assumed distinct from the parameters $\underset{\sim}{\phi}$ and $\underset{\sim}{\lambda}$.

After selection the marginal distribution of $\underset{\sim}{\tilde{Z}}_U$ is modified into another distribution which we denote by

$$f^*(\underset{\sim}{Z}_U;\underset{\sim}{\phi};\underset{\sim}{\delta})=f(s|\underset{\sim}{Z}_U,\underset{\sim}{Y};\underset{\sim}{\delta})f(\underset{\sim}{Z}_U;\underset{\sim}{\phi}). \qquad [1.3]$$

The joint distribution of $\underset{\sim}{\tilde{Y}}_U,\underset{\sim}{\tilde{Z}}_U$ and $\tilde{s}$ is given by

$$f(\underset{\sim}{Y}_U,\underset{\sim}{Z}_U,s;\underset{\sim}{\lambda},\underset{\sim}{\delta},\underset{\sim}{\phi})= f(\underset{\sim}{Y}_U|\underset{\sim}{Z}_U;\underset{\sim}{\lambda})\, f^*(\underset{\sim}{Z}_U;\underset{\sim}{\phi};\underset{\sim}{\delta})$$

$$= f(\underset{\sim}{Y}_U|\underset{\sim}{Z}_U;\underset{\sim}{\lambda})f(s|\underset{\sim}{Z}_U,\underset{\sim}{Y}_U;\underset{\sim}{\delta})\, f(\underset{\sim}{Z}_U;\underset{\sim}{\phi}). \qquad [1.4]$$

[1.4] represents the global model which embraces all possible mechanisms of generating samples and population values. Inference can now be based on the probability distributions generated by this model.

A special case of [1.2] is when the selection mechanism depends only on the design variable,i.e

$$f(s|\underset{\sim}{Z}_U,\underset{\sim}{Y}_U;\underset{\sim}{\delta})=f(s|\underset{\sim}{Z}_U).$$

This condition ensures that the sample selection mechanism depends only on the design variables and excludes sampling schemes such as quota sampling.

If this condition is satisfied then using [1.4] the unknown parameters $\underset{\sim}{\lambda}$ can be estimated from the conditional distribution of $\underset{\sim}{\tilde{Y}}_U$ given $\underset{\sim}{\tilde{Z}}_U=\underset{\sim}{Z}_U$ and $\underset{\sim}{\phi}$ from the marginal distribution of $\underset{\sim}{\tilde{Z}}_U$.The parameter of interest is in the marginal distribution of $\underset{\sim}{\tilde{y}}$ which is a function of both $\underset{\sim}{\lambda}$ and $\underset{\sim}{\phi}$,and we denote this parameter by $\underset{\sim}{\theta}$,i.e $\underset{\sim}{\theta}=g(\underset{\sim}{\lambda},\underset{\sim}{\phi})$,where g denotes a continous function.

EXAMPLE

Suppose we have three variables y_1,y_2 and z where y_1 is a dependent variable , y_2 is an independent variable and z is the design variable.In considering regression analysis we have two cases ;

(i) In the first case the design variable z is considered as another independent variable so that the regression equation is;

$$y_1=\alpha_{1.2z}+\beta_{12.z}y_2+\beta_{1z.2}z\ +\varepsilon_{1.2z}$$

The parameters of interest are in the conditional distribution of y_1 given z and y_2 can be estimated using least squares method of estimation.

(ii) In the second case the design variable is used in the design and is not considered as an independent variable in the regression.For example in the study of the relationship between income (y_1) and educational attainment(y_2),the tax paid by the individuals (z) might be used as a stratifying variable in the design.Thus the design variable does not appear explicitly in the regression relationship. The regression equation is given by;

$$y_1=\alpha_{12}+\beta_{12}y_2+\varepsilon_{12}.$$

We see that in this case the parameters of interest are α_{12} and β_{12} which are in the marginal distribution of y.

If we consider the case where the variables $\tilde{y}_1, \tilde{y}_2$ and $\tilde{z}$ are jointly distributed as multivariate normal variables then the parameter of interest is $\underset{\sim}{\theta}=(\mu_y,\Sigma_{yy})$, indexing the marginal distribution of $\tilde{y}$. The parameter $\underset{\sim}{\phi}=(\mu_z,\Sigma_{zz})$ indexes the marginal distribution of $\tilde{z}$. The parameter indexing the conditional distribution of $\tilde{y}_1$ given $\underset{\sim}{z}$ is $\underset{\sim}{\lambda}=(\mu_{y.z},\Sigma_{yy.z},B_{yz})$ where,

$$\mu_{y.z}=\mu_y-B_{yz}\mu_z \ ,$$

$$\Sigma_{yy.z}=\Sigma_{yy}-B_{yz}\Sigma_{zz}B^T_{yz}$$

$$B_{yz} \ =\Sigma_{yz}\Sigma^{-1}_{zz}.$$

From these expressions we see that $\underset{\sim}{\theta}=g(\underset{\sim}{\lambda},\underset{\sim}{\phi})$, i.e $\underset{\sim}{\theta}$ is a function of $\underset{\sim}{\lambda}$ and $\underset{\sim}{\phi}$.

By assuming multivariate normality, Smith[1981] derived the maximum likelihood estimators of the parameters $\underset{\sim}{\lambda}$, denoted by $\hat{\underset{\sim}{\lambda}}$ as, $\hat{\underset{\sim}{\lambda}}=(\hat{\mu}_{y.z},\hat{\Sigma}_{yy.z},\hat{B}_{yz})$,

where

$$\hat{\mu}_{y.z}=\hat{\mu}_y-\hat{B}_{yz}\hat{\mu}_z \ ,$$

$$\hat{\Sigma}_{yy.z}=\hat{\Sigma}_{yy}-\hat{B}_{yz}\hat{\Sigma}_{zz}B^T_{yz}$$

$$\hat{B}_{yz} \ =\hat{\Sigma}_{yz}\hat{\Sigma}^{-1}_{zz}.$$

If $\hat{\underset{\sim}{\theta}}$ denotes the maximum likelihood estimate of $\underset{\sim}{\theta}$, then from these expressions we see that $\hat{\underset{\sim}{\theta}}=g(\hat{\underset{\sim}{\lambda}},\hat{\underset{\sim}{\phi}})$.

In complex surveys the design variable may denote strata or cluster membership used solely at the design stage to improve the efficiency of the estimators and so is not included explicitly in the regression relationship. So our parameters of interest are $\underset{\sim}{\theta}$ and not $\underset{\sim}{\lambda}$.

From [1.4] we obtain the joint distribution of the data $d_s=(\underset{\sim}{Y}_s,\underset{\sim}{Z}_U,s)$ by integratingoutthe nonsample values from this joint probability distributionof the dataand the nonsample units $Y_{\bar{s}}$

$$f(\underset{\sim}{d_s};\underset{\sim}{\lambda},\underset{\sim}{\phi},\underset{\sim}{\delta}) \propto \int f(\underset{\sim}{Y_U}|\underset{\sim}{Z_U};\underset{\sim}{\lambda}) f(s|\ \underset{\sim}{Z_U},\underset{\sim}{Y_U};\underset{\sim}{\delta}) f(\underset{\sim}{Z_U};\underset{\sim}{\phi}) d\underset{\sim}{Y_{\bar{s}}}. \qquad [1.5]$$

We now consider how we can make inference about analytic and descriptive targets.

(i) ANALYTICAL TARGETS

For analytical targets in inference the parameters of interest are the superpopulation parameters such as $\underset{\sim}{\theta}$ which cannot be expressed as a function of the values of survey variables $\underset{\sim}{\tilde{Y}_U}$ attached to the N units in the finite population.This is because this parameter is in a wider population,the so called superpopulation most likely different from the population being sampled.In order to estimate this parameter we require some assumptions that relate the population being sampled to this parameter of analytic interest.Therefore this inference is based on a model embracing all these assumptions. For example the application of regression analysis in seeking to explain the relationship between the variables is beyond the particular finite population which gave rise to the data as in McKennell's [1969] survey of aircraft noise.

(ii) DESCRIPTIVE TARGETS

For descriptive targets in inference the parameter of interest is a known function of the values of survey variable $\underset{\sim}{\tilde{Y}_U}$ attached to the N units in the finite population. If all the units in the finite population are evaluated then in the absence of measurement errors,there will be no uncertainity in descriptive targets. Means, variances, correlations, differences and regressions within the finite population framework are examples of some descriptive statistics.

There are two competing approaches to inference for each of the above types of targets

(i) Model based approach.

(ii) Randomization approach.

We will review both the approaches to inference for each of the above types of targets and try to interconnect them through the global model.

1.3.1 MODEL BASED APPROACH TO INFERENCE

In the model based approach to inference the finite population is assumed to be generated as a random sample from a wider population the so called superpopulation.A sample is then selected from the finite population using a known sampling scheme.Thus in this approach the finite population values $\underset{\sim}{Y}_U$ and $\underset{\sim}{Z}_U$ are not fixed but are treated as realizations of the random variables $\underset{\sim}{\tilde{Y}}_U$ and $\underset{\sim}{\tilde{Z}}_U$ respectively.Since in model based approach the parameter of interest $\underset{\sim}{\theta}$ is not in the population being sampled,a model embracing all the assumptions relating the parameter $\underset{\sim}{\theta}$ and the population being sampled is assumed.This model provides a relationship between the survey variables $\underset{\sim}{\tilde{Y}}_U$ and its parameter $\underset{\sim}{\theta}$.Thus the observed values $\underset{\sim}{y}_i$,i∊s can provide both estimates of $\underset{\sim}{\theta}$ and the predictions of $\underset{\sim}{y}_i$,i∊$\bar{s}$. Many different models have been proposed.Ericson[1969] and Sugden[1979] proposed models based on exchangeability assumptions,Scott and Smith[1969] made classical normal model assumptions ,Royall[1970,1971] and Royall and Herson [1973a , b] pursue the linear model prediction approach.Smith [1976] and Cassel.et.al[1977] have reviewed these approaches.We now review the model based approach to inference for both descriptive and analytical targets.

1.3.1.1 MODEL BASED APPROACH TO INFERENCE FOR DESCRIPTIVE TARGETS(PREDICTION INFERENCE)

In prediction inference observations on sample values are used to make predictions about non sample values.This inference allows the joint distribution of $\underset{\sim}{\tilde{Y}}_U$ to act as a link between the observed and the unobserved variables. We use the sample values to estimate the parameter $\underset{\sim}{\lambda}$ in the conditional distribution of the sample values given the values of the design variables, which is then used to predict the values not included in the sample. Predictions are made conditional on the sample values actually selected. Therefore if $\underset{\sim}{y}_i$ are the realized values of the survey variables $\underset{\sim}{\tilde{y}}_i$ then prediction is concerned with the use of $\underset{\sim}{y}_i, i\in s$ to predict values of the survey variables $\underset{\sim}{\tilde{y}}_i, i\in\bar{s}$. Then on the basis of these $\underset{\sim}{Y}$ values, some function of the finite population can be estimated, for example the finite population total.

From [1.5] the likelihood function is given by

$$L(\underset{\sim}{\lambda},\underset{\sim}{\phi},\underset{\sim}{\delta}) \propto \int f(\underset{\sim}{Y}_U|\ \underset{\sim}{Z}_U;\underset{\sim}{\lambda})f\ (s|\ \underset{\sim}{Z}_U,\underset{\sim}{Y}_U;\underset{\sim}{\delta})\ f(\underset{\sim}{Z}_U;\underset{\sim}{\phi})d\underset{\sim}{Y}_{\bar{s}}\ .$$

Since the sampling mechanism depends on the values of the survey variables as well as the design variables it is difficult to evaluate this likelihood function therefore to simplify it we make the following assumptions;

ASSUMPTION 1.1

The survey variables $\underset{\sim}{\tilde{Y}}_U$ and the sample variables $\tilde{s}$ are conditionally independent given the design variable $\underset{\sim}{\tilde{Z}}_U=\underset{\sim}{Z}_U$

$$\tilde{s} \perp\!\!\!\perp \underset{\sim}{\tilde{Y}}_U|\underset{\sim}{Z}_U.$$

the symbol $\perp\!\!\!\perp$ denotes independence; see Dawid [1977]

If this condition holds then

$$f(s|\ \underset{\sim}{Z}_U,\underset{\sim}{Y}_U;\underset{\sim}{\delta})=f(s|\ \underset{\sim}{Z}_U).$$

Sample designs of this form are known as *noninformative designs*. Random sampling designs and some purposive sampling designs satisfy this criterion; see Scott[1977]

EXAMPLE

The simplest form of a purposive design is a non randomized design given by

$$f(s|Z_U)=\begin{cases} 1 & \text{if } s=s_o(Z_U), \\ 0 & \text{otherwise.} \end{cases}$$

where $s_o(Z_U)$ is a known function defined in such a way that the sample values are selected to achieve a certain objective.For example Demets and Halperin[1977] in the Framingham heart study considered a population of persons characterised by their initial serum cholestrol level.The objective was to study the effect of dietary cholestrol on serum chorestrol and $s_o(Z_U)$ was chosen in such a way that only persons with high or low initial serum cholestrol levels were included in the sample in order to minimize the variance of the estimated regression coefficient.

ASSUMPTION 1.2 (independence assumption)

The observed variables $\tilde{\underset{\sim}{Y}}_s$ are conditionally independent of the unobserved variables $\tilde{\underset{\sim}{Y}}_{\bar{s}}$ given the design variables $\tilde{\underset{\sim}{Z}}_U=\underset{\sim}{Z}_U$ and $\underset{\sim}{\lambda}$ i.e

$$\tilde{\underset{\sim}{Y}}_s \perp\!\!\!\perp \tilde{\underset{\sim}{Y}}_{\bar{s}} \mid \underset{\sim}{Z}_U;\underset{\sim}{\lambda}.$$

This is a very strong assumption and does not always hold.

EXAMPLE

There are situations when the population units are spread over a large geographical area,and clusters are created by subdividing the area.Suppose unit y_i is selected from cluster c_1 and unit y_j also in cluster c_1 is not selected.Since both units are in the same cluster it might be unreasonable to assume that these two units are uncorrelated .The correlation between the two units may depend on the distance between them.

Thus if $\underset{\sim}{Z}_U$ contains the clustering variable then the sample and nonsample units from the same cluster will be correlated,while if they are from different clusters,then assumption 1.2 may hold.

In such a situation if assumption [1.2] does not hold Scott and Smith [1969], Royall [1976b], Skinner [1981] assumed a superpopulation model which accounts for the groupings in the population and derived predictors based on this model.

ASSUMPTION 1.3

The conditional distribution of the survey variables $\tilde{\underset{\sim}{Y}}_U$ given the design variables $\tilde{\underset{\sim}{Z}}_U=\underset{\sim}{Z}_U$,the conditional distribution of the observed variables $\tilde{\underset{\sim}{Y}}_s$ given $\tilde{\underset{\sim}{Z}}_U=\underset{\sim}{Z}_U$ and the conditional distribution of the unobserved variables $\tilde{\underset{\sim}{Y}}_{\bar{s}}$ given $\tilde{\underset{\sim}{Z}}_U=\underset{\sim}{Z}_U$ are indexed by the same parameter $\underset{\sim}{\lambda}$.In other words the process by which the finite population is partitioned into sample and nonsample units is irrelevant to inference about the parameter $\underset{\sim}{\lambda}$ i.e

$$f(\underset{\sim}{Y}_U \mid \underset{\sim}{Z}_U;\underset{\sim}{\lambda}) = f(\underset{\sim}{Y}_s \mid \underset{\sim}{Z}_U;\underset{\sim}{\lambda}) f(\underset{\sim}{Y}_{\bar{s}} \mid \underset{\sim}{Z}_U;\underset{\sim}{\lambda}),$$

using assumption [1.2].

This assumption does not always hold.

EXAMPLE

Consider a population consisting of men and women.Define a purposive design $f(s|Z_U)$

$$f(s \mid Z_U) = \begin{cases} 1 & \text{if } s=s_o(z), \\ 0 & \text{otherwise,} \end{cases}$$

where

$$s_o(z) = (\text{set of all women}).$$

Let $\underset{\sim}{\lambda}=(\lambda_m,\lambda_w)$ where λ_m corresponds to the parameter for men and λ_w for women.If y denotes a characteristic of interest then after a sample is selected with the known sample design $f(s|Z_U)$ i.e assumption [1.1] holds then assuming that the $\tilde{y}_i$'s are independently distributed then the likelihood function given in [1.5] simplifies to

$$L(\underset{\sim}{\theta}) \propto f(s|Z_U) \prod_{i\in w} f(y_i|Z_U;\lambda_w)\, f(z_i;\phi)$$

$$\propto \begin{cases} \prod_{i\in w} f(y_i|Z_U;\lambda_w) f(z_i;\phi), \\ 0 \qquad \text{otherwise.} \end{cases}$$

We therefore see that λ_m is not in the likelihood hence no inference about this parameter is possible. So in this case the process by which the population was partitioned into men and women is relevant to the inference about $\underset{\sim}{\lambda}$.

Consider the model [1.4], then

$$f(\underset{\sim}{Y}_U,\underset{\sim}{Z}_U,s;\underset{\sim}{\lambda},\underset{\sim}{\phi},\underset{\sim}{\delta})=f(\underset{\sim}{Y}_{\bar{s}},\underset{\sim}{Y}_s,\underset{\sim}{Z}_U,s;\underset{\sim}{\lambda},\underset{\sim}{\phi},\underset{\sim}{\delta})$$

$$=f(\underset{\sim}{Y}_{\bar{s}}|\underset{\sim}{Y}_s,\underset{\sim}{Z}_U,s;\underset{\sim}{\lambda}, \underset{\sim}{\phi},\underset{\sim}{\delta})f(\underset{\sim}{Y}_s|\underset{\sim}{Z}_U,s;\underset{\sim}{\lambda},\underset{\sim}{\phi},\underset{\sim}{\delta})f(\underset{\sim}{Z}_U,s;\underset{\sim}{\lambda},\underset{\sim}{\delta},\underset{\sim}{\phi})f(s|\underset{\sim}{Z}_U;\underset{\sim}{\phi},\underset{\sim}{\delta},\underset{\sim}{\lambda}).$$

If assumptions [1.1],[1.2] and [1.3] holds then simplifying this expression we get

$$f(\underset{\sim}{Y}_{\bar{s}},\underset{\sim}{Y}_s,\underset{\sim}{Z}_U,s;\underset{\sim}{\lambda},\underset{\sim}{\phi})=f(\underset{\sim}{Y}_{\bar{s}}|\underset{\sim}{Z}_U,s;\underset{\sim}{\lambda})f(\underset{\sim}{Y}_s|\underset{\sim}{Z}_U,s;\underset{\sim}{\lambda})f(\underset{\sim}{Z}_U;\underset{\sim}{\phi})f(s|\underset{\sim}{Z}_U), \qquad [1.6]$$

where $f(\underset{\sim}{Y}_{\bar{s}}|\underset{\sim}{Z}_U,s;\underset{\sim}{\lambda})$ is the predictive distribution.

The development of a predictive method of parametric distribution involves the elimination of $\underset{\sim}{\lambda}$ from this distribution in such a way that no available information about $\underset{\sim}{Y}_{\bar{s}}$ is lost.There are various methods which have been advocated in the literature for accomplishing this.In this work we will consider the following two approaches:

(i) Bayesian predictive approach,

(ii) Classical predictive approach.

1.3.1.1.1 BAYESIAN PREDICTIVE APPROACH

In this approach $\underset{\sim}{\lambda}$ is regarded as a realized value of a random variable $\underset{\sim}{\tilde{\lambda}}$ whose distribution reflects subjective beliefs in what the value of $\tilde{\lambda}$ is likely to be. We suppose that before observing the sample values our belief about the value of $\tilde{\lambda}$ could be represented by a prior distribution $f(\lambda)$ and assume that sample design is noninformative i.e assumption 1.1.holds. The posterior distribution of the unobserved variables $\tilde{Y}_{\bar{s}}$ given the design variables $\tilde{Z}_U = Z_U$ and the observed variables $\tilde{Y}_s$ is given by

$$f(Y_{\bar{s}}|Z_U,Y_s)=\int_{\lambda} f(Y_{\bar{s}},\lambda|Z_U,Y_s)d(\lambda)$$

$$=\int_{\lambda} f(Y_{\bar{s}}|Z_U,Y_s,\lambda)f(\lambda|Y_s,Z_U)d(\lambda)$$

$$=\int_{\lambda} f(Y_{\bar{s}}|Z_U,\lambda)f(\lambda|Y_s,Z_U)d(\lambda), \qquad [1.7]$$

using assumption 1.2.

Eq(1.7) is the Bayesian predictive distribution of the unobserved variables $\tilde{Y}_{\bar{s}}$ given the observed variables $\tilde{Y}_s$ and the design variable $\tilde{Z}_U$ implied by the posterior distribution $f(\lambda|Y_s,Z_U)$, see Aitchison and Dunsmore [1975], p19. This posterior distribution $f(\lambda|Y_s,Z_U)$ can be computed.i.e Using Bayes' theorem this posterior distribution is given by

$$f(\lambda|Y_sZ_U)=\frac{f(Y_s|Z_U,\lambda)f(\lambda)}{f(\ Y_s|Z_U)},$$

where $f(Y_s|Z_U)=\int_{\lambda} f(Y_s|Z_U,\lambda)f(\lambda)d(\lambda)$.

We see that all the terms in the distribution $f(\lambda|Y_s,Z_U)$can be calculated on the basis of the sample selected.

The integration of [1.7] is quite straightforward, if the prior distribution and the distribution of the sample data $f(Y_s|Z_U,\lambda)$ are known exactly. If the prior distribution is not known then it is not obvious what to do.

Assuming that the distribution of the sample data is known exactly, Aitchison and Dunsmore [1975] advocated the use of predictive likelihood based on maximum likelihood estimate $\hat{\underset{\sim}{\lambda}}$, to replace the unknown parameter $\underset{\sim}{\lambda}$, Hinckley [1979], Lauritzen [1974] and recently Butler [1986], have advocated the use of predictive likelihood based on the value of a sufficient statistic for the parameter $\underset{\sim}{\lambda}$.

In complex sample surveys the distribution of the sample data is not known exactly, maybe only the first two moments are known. In such a situation neither the Bayesian nor the likelihood predictive approach is appropriate and the classical approach to prediction which does not assume that the exact distribution is known seems more desirable.

In the next subsection we will review the classical approach to prediction.

1.3.1.1.2 CLASSICAL PREDICTIVE APPROACH

Using [1.6] the classical predictive distribution is given by

$$f(\underset{\sim}{Y}_{\bar{s}}|\underset{\sim}{Z}_U,s;\underset{\sim}{\lambda})|_{\underset{\sim}{\lambda}=\hat{\underset{\sim}{\lambda}}} \ . \qquad [1.8]$$

In this work we will relax the exact model assumptions made in the predictive likelihood approach to a model with only linear and homoscedastic assumptions. Then under the linear homoscedastic model, $\hat{\underset{\sim}{\lambda}}$ could be the least squares estimator of $\underset{\sim}{\lambda}$. Since $\underset{\sim}{Z}_U$, s and $\hat{\underset{\sim}{\lambda}}$ are known, the conditional distribution $f(\underset{\sim}{Y}_{\bar{s}}|\underset{\sim}{Z}_U,s;\underset{\sim}{\lambda}|_{\underset{\sim}{\lambda}=\hat{\underset{\sim}{\lambda}}})$ can now be used to predict the unobserved values $\underset{\sim}{y}_i, i\in\bar{s}$. This is the classical approach to prediction which uses the values of the observed variables to infer about the joint distribution between the observed and the unobserved variables. This approach assumes that $\underset{\sim}{\lambda}$ is known to be $\hat{\underset{\sim}{\lambda}}$.

We will review the derivation of the predictors of linear functions of the survey variables in the finite population under the linear Homoscedastic model 1.

LINEAR HOMOSCEDASTIC MODEL 1.

Under this model it is assumed that

$$E_1(\tilde{\underset{\sim}{y}}|\tilde{\underset{\sim}{z}}=\underset{\sim}{z}_i)=\underset{\sim}{m}+\underset{\sim}{H}\underset{\sim}{z}_i,$$

and

$$V_1(\tilde{\underset{\sim}{y}}|\tilde{\underset{\sim}{z}}=\underset{\sim}{z}_i)=\underset{\sim}{K},$$

where $\underset{\sim}{m}$ is a px1 constant vector, $\underset{\sim}{H}$ is a pxq constant matrix and $\underset{\sim}{K}$ is a pxp constant matrix. Thus the conditional expectation of the survey variables $\tilde{\underset{\sim}{y}}$ given the design variables $\tilde{\underset{\sim}{z}}=\underset{\sim}{z}_i$ is a linear function of $\underset{\sim}{z}_i$ and the variance covariance matrix of the survey variables $\tilde{\underset{\sim}{y}}$ given the design variables $\tilde{\underset{\sim}{z}}=\underset{\sim}{z}_i$ is constant and does not depend on $\underset{\sim}{z}_i$.

We also assume that

$$\underset{\sim}{y}_i \perp\!\!\!\perp \underset{\sim}{y}_j|\underset{\sim}{Z}_U \text{ and the parameters for } i\neq j,$$

and $\underset{\sim}{z}_1,..,\underset{\sim}{z}_N$ are identically and independently distributed as $\underset{\sim}{z}$.

We now derive the expressions of $\underset{\sim}{m}$, $\underset{\sim}{K}$ and $\underset{\sim}{H}$ in terms of the parameters in the marginal distribution of $\tilde{\underset{\sim}{y}}$ and of $\tilde{\underset{\sim}{z}}$. Taking expectations and variances of model 1 respectively over the model distribution of $\tilde{\underset{\sim}{z}}$ we get

$$E_z\, E_1(\tilde{\underset{\sim}{y}}|\tilde{\underset{\sim}{z}}=\underset{\sim}{z}_i)= E_z(\underset{\sim}{m})+\underset{\sim}{H}\, E_z(\underset{\sim}{z}_i)$$

$$=\underset{\sim}{m}+\underset{\sim}{H}\, E_z(\underset{\sim}{z}_i). \qquad [1.9]$$

Let $\underset{\sim}{\mu}_y=E_z(E_1(\tilde{\underset{\sim}{y}}|\underset{\sim}{z}_i))$,

$\underset{\sim}{\mu}_z=E_z(\underset{\sim}{z}_i)$.

Then [1.9] becomes

$$\underset{\sim}{\mu}_y = \underset{\sim}{m} + \underset{\sim}{H}\underset{\sim}{\mu}_z . \qquad [1.10]$$

Now

$$\underset{\sim}{\Sigma}_{yy} = V(\underset{\sim}{\tilde{y}})$$

$$= V_z(E_1(\underset{\sim}{\tilde{y}}|\underset{\sim}{z}_i)) + E_z(V_1(\underset{\sim}{\tilde{y}}|\underset{\sim}{z}_i))$$

$$= V_z(\underset{\sim}{m} + \underset{\sim}{H}\ \underset{\sim}{z}_i) + E_z(\underset{\sim}{K}),$$

using [1.7] and [1.8],

$$= \underset{\sim}{H}\ V_z(\underset{\sim}{z}_i)\ \underset{\sim}{H}^T + \underset{\sim}{K} . \qquad [1.11]$$

Let

$V_z(\underset{\sim}{z}) = \underset{\sim}{\Sigma}_{zz}$, then [1.11] becomes

$$\underset{\sim}{\Sigma}_{yy} = \underset{\sim}{H}\ \underset{\sim}{\Sigma}_{zz}\ \underset{\sim}{H}^T + \underset{\sim}{K}. \qquad [1.12]$$

To evaluate $\underset{\sim}{H}$, we consider,

$$\underset{\sim}{\Sigma}_{yz} = Cov(\underset{\sim}{\tilde{y}}, \underset{\sim}{\tilde{z}})$$

$$= E(\underset{\sim}{\tilde{y}}\ \underset{\sim}{\tilde{z}}) - E(\underset{\sim}{\tilde{y}})E(\underset{\sim}{\tilde{z}})$$

$$= E(\underset{\sim}{\tilde{y}}\ \underset{\sim}{\tilde{z}} - E(\underset{\sim}{\tilde{y}})E(\underset{\sim}{\tilde{z}}))$$

$$= E_z E_1((\underset{\sim}{\tilde{y}}\underset{\sim}{\tilde{z}} - E_1(\underset{\sim}{\tilde{y}}|\underset{\sim}{z}_i)E_1(\underset{\sim}{\tilde{z}}|\underset{\sim}{z}_i) + E_1(\underset{\sim}{\tilde{y}}|\underset{\sim}{z}_i)E_1(\underset{\sim}{\tilde{z}}|\underset{\sim}{z}_i) - E(\underset{\sim}{\tilde{y}})E(\underset{\sim}{\tilde{z}})|\underset{\sim}{z}_i)$$

$$= E_z(cov_1(\underset{\sim}{\tilde{y}}, \underset{\sim}{\tilde{z}}|\underset{\sim}{z}_i)) + cov_z(E_1(\underset{\sim}{\tilde{y}}|\underset{\sim}{z}_i), E_1(\underset{\sim}{\tilde{z}}|\underset{\sim}{z}_i)).$$

Now since $E_1(\underset{\sim}{\tilde{z}}|\underset{\sim}{z}_i) = \underset{\sim}{z}_i$ and $cov_1(\underset{\sim}{\tilde{y}}, \underset{\sim}{\tilde{z}}|\underset{\sim}{z}_i) = 0$,

then we get,

$$\underset{\sim}{\Sigma}_{yz} = Cov(E_1(\underset{\sim}{\tilde{y}}|\underset{\sim}{z}_i), \underset{\sim}{z}_i)$$

$$= Cov_z(\underset{\sim}{m} + \underset{\sim}{H}\ \underset{\sim}{z}_i\ , \underset{\sim}{z}_i) \quad \text{Using model 1,}$$

$$= \underset{\sim}{H}\ V_z(\underset{\sim}{z}_i)$$

$$= \underset{\sim}{H} \sum_{zz} .$$

Assuming $\sum_{zz}$ is nonsingular we get

$$\underset{\sim}{H} = \underset{\sim}{\sum}_{yz} \underset{\sim}{\sum}_{zz}^{-1} . \qquad [1.13]$$

$= \underset{\sim}{\beta}_{yz}$ the regression coefficient of $\underset{\sim}{\tilde{y}}$ on $\underset{\sim}{z}$.

As an example of a linear predictor of a function of the survey variables e.g finite population mean under model 1,we state the following theorem proved by Royall and Herson[1973a]

Theorem 1.1(Royall and Herson [1973a])

Under model 1 the minimum variance unbiased predictor of the finite population mean of the survey variables $\underset{\sim}{\tilde{y}}_i$ is given by;

$$\hat{\bar{\underset{\sim}{y}}}_U = \bar{\underset{\sim}{y}}_s + \hat{\underset{\sim}{H}}(\bar{\underset{\sim}{z}}_U - \bar{\underset{\sim}{z}}_s), \qquad [1.14]$$

where

$$\bar{\underset{\sim}{y}}_s = n^{-1}\sum_s \underset{\sim}{y}_i \; , \bar{\underset{\sim}{z}}_s = n^{-1}\sum_s \underset{\sim}{z}_i , \bar{\underset{\sim}{z}}_U = N^{-1}\sum_U \underset{\sim}{z}_i \quad , \qquad [1.15]$$

and

$$\hat{\underset{\sim}{H}} = \underset{\sim}{S}_{yzs} \underset{\sim}{S}_{zzs}^{-1}, \quad \text{where} \quad \underset{\sim}{S}_{yzs} = n^{-1}\sum_s (\underset{\sim}{y}_i - \bar{\underset{\sim}{y}}_s)(\underset{\sim}{z}_i - \bar{\underset{\sim}{z}}_s)^T \text{ and}$$

$$\underset{\sim}{S}_{zzs} = n^{-1}\sum_s (\underset{\sim}{z}_i - \bar{\underset{\sim}{z}}_s)(\underset{\sim}{z}_i - \bar{\underset{\sim}{z}}_s)^T .$$

[1.14] is the regression type estimator of $\bar{\underset{\sim}{y}}_U$ see Cochran [1977].However we note that in Cochran[1977],$\hat{\bar{\underset{\sim}{y}}}_U$ was derived without assuming any model.

When the population is clustered Scott and Smith [1969],Royall [1976b] have derived the estimators of the finite population mean in multistage surveys.

1.3.1.2 MODEL BASED APPROACH TO INFERENCE FOR ANALYTICAL TARGETS

In the previous subsection we have dealt extensively on the model based approach to inference for descriptive targets.Suppose now our parameter of interest is not in the sampled population but in another population most likely different from this sampled population.For example in Britain the A level score is the selection variable for admitting students into the university.Suppose that one student who sat for the A level exam but was not selected challenges this selection criterion and argues that it is not a valid measure of performance in the university.How can the university defend this criterion of selection? Inference to descriptive targets points to descriptive statistics like means,correlation etc based on the selected group of candidates but this will not reflect the performance in the university of those students who were not admitted into the university. Therefore for the university to give a satisfactory defence to this challenge they need to find the relationship of the performance of the selected group with the total number of applicants.We assume that the total number of applicants is a random sample from the population of all the possible applicants.Thus the parameter of interest here is in the population of all the possible applicants.

The question is how can we deduce this relationship when the only data available for the performance is in the selected population? i.e those applicants admitted into the university. If we consider the performance of all the applicants as the survey variable $\underset{\sim}{\tilde{y}}$ and the A level score as the design variable $\underset{\sim}{\tilde{z}}$ then our parameter of interest is in the marginal distribution of $\underset{\sim}{\tilde{y}}$ i.e if θ is the parameter of interest then it indexes $f(\underset{\sim}{y};\theta)$.Let the values of the survey variable of the selected group be denoted by $\underset{\sim}{y}_s$ then the parameter λ

indexes the conditional distribution of $\tilde{y}_s$ given $\underset{\sim}{z}$ i.e $f(\underset{\sim}{y}_s|\underset{\sim}{z};\lambda)$ and ϕ indexes the marginal distribution of $\tilde{\underset{\sim}{z}}$ i.e $f(\underset{\sim}{z};\phi)$. λ is the parameter of interest in predictive type of inference and can be estimated from the sample data without any problems.However our parameter of interest is θ which is not directly observable from the data. The problem is how do we use the conditional distribution $f(\underset{\sim}{y}_s|\underset{\sim}{z};\lambda)$ to make inferences about this parameter θ which is in the marginal distribution of $\tilde{\underset{\sim}{y}}$,that is when θ is a function of both λ and ϕ.This is known as the problem arising due to the effects of selection.

1.3.1.2.1 PEARSON SELECTION EFFECTS

The effects due to selections are encountered in various disciplines like psychological and epidemiological studies, educational testing etc.A solution to the problem arising due to the effect of selection was first given by Pearson[1903] who was studying the effect of natural selection on regression relationships in the case where the population of interest was multivariate normal.Thus he assumed that the joint distribution of the survey variables $\tilde{\underset{\sim}{y}}$ and the design variables $\tilde{\underset{\sim}{z}}$ is multivariate normal. Aitken[1934] reformulated the problem in the language of matrix algebra and Lawley[1943] established an important theorem in which the assumptions of multivariate normality are relaxed. So he extended the Pearson's results to more general distributions with linear regressions. Under the normality assumptions $\underset{\sim}{\lambda}$ and $\underset{\sim}{\phi}$ can be estimated by maximum likelihood estimators $\hat{\underset{\sim}{\lambda}}$ and $\hat{\underset{\sim}{\phi}}$ respectively.Since $\underset{\sim}{\theta} = g(\underset{\sim}{\lambda},\underset{\sim}{\phi})$,then by the invariance property of the maximum likelihood estimators $\hat{\underset{\sim}{\theta}} = g(\hat{\underset{\sim}{\lambda}},\hat{\underset{\sim}{\phi}})$ is a maximum likelihood estimator of θ. To study the effect due to selection we formulate the problem as follows:let $\tilde{\underset{\sim}{z}}=(\tilde{z}_1 \ldots \tilde{z}_q)^T$ be a q dimensional random variable which is used as the basis of selection and $\tilde{y}=(\tilde{y}_1 \ldots \tilde{y}_p)^T$ be a p dimensional random vector of survey variables.We derive the conditional distribution of $\tilde{y}=(\tilde{y}_s,\tilde{y}_{\bar{s}})$ and $\tilde{z}$ after the sample is selected from the finite population U by a known sampling scheme.

Now the joint distribution of $\tilde{y}$ and $\tilde{z}$ is given by

$$f(y,z;\lambda,\phi) = f(y|\ z;\lambda)\ f(z;\phi) \qquad [1.16]$$

and has mean vector μ and variance covariance matrix Σ given by

$$\mu = \begin{bmatrix} \mu_y \\ \mu_z \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yz} \\ \Sigma_{zy} & \Sigma_{zz} \end{bmatrix}$$

If we select a sample s from the finite population U using a known sampling scheme $f(s|z)$,then the joint distribution of $\tilde{y},\tilde{z}$ and s is given by

$$f(y,z,s;\lambda,\phi) = f(s|\ z)\ f(y|\ z;\lambda)\ f(z;\phi). \qquad [1.17]$$

If the inference based on the joint distribution [1.16] is identical to that based on [1.17] then we say that the sampling scheme is *ignorable*.For all noninformative designs the sampling scheme is ignorable in model based inference.However for the informative designs of the form $f(s|\ z,y;\delta)$ the sampling scheme is not ignorable.

After sampling the value of y is y_s we want to find

$$f(\ y_s,z|s;\lambda,\phi) = \frac{f(\ y_s,z\ ,s;\lambda,\phi)}{f(s;\lambda,\ \phi)}. \qquad [1.18]$$

Consider

$$f(\underset{\sim}{s};\underset{\sim}{\lambda},\underset{\sim}{\phi}) = \iint f(\underset{\sim}{s}|\underset{\sim}{z})\, f(\underset{\sim}{y}_s|\underset{\sim}{z};\underset{\sim}{\lambda})\, f(\underset{\sim}{z};\underset{\sim}{\phi})d\underset{\sim}{y}_s d\underset{\sim}{z}$$

$$= \int f(\underset{\sim}{s}|\underset{\sim}{z})\, f(\underset{\sim}{z};\underset{\sim}{\phi})d\underset{\sim}{z} \int f(\underset{\sim}{y}_s|\underset{\sim}{z};\underset{\sim}{\lambda})\, d\underset{\sim}{y}_s$$

$$= \int f(\underset{\sim}{s}|\underset{\sim}{z})\, f(\underset{\sim}{z};\underset{\sim}{\phi})d\underset{\sim}{z}$$

$$= f(s;\underset{\sim}{\phi}), \qquad [1.19]$$

$$\text{since} \int f(\underset{\sim}{y}_s|\underset{\sim}{z};\underset{\sim}{\lambda})\, d\underset{\sim}{y}_s = 1.$$

Substituting (1.19) in (1.18) we get

$$f(\underset{\sim}{y}_s,\underset{\sim}{z}\,|s;\underset{\sim}{\lambda},\underset{\sim}{\phi}) = \frac{f(\underset{\sim}{y}_s,\underset{\sim}{z},s,\underset{\sim}{\lambda},\underset{\sim}{\phi})}{f(s;\phi)} \qquad [1.20]$$

$$= \frac{f(s|\underset{\sim}{z})\, f(\underset{\sim}{y}_s|\underset{\sim}{z};\underset{\sim}{\lambda})\, f(\underset{\sim}{z};\underset{\sim}{\phi})}{f(s;\underset{\sim}{\phi})}$$

$$= f(\underset{\sim}{y}_s|\underset{\sim}{z};\underset{\sim}{\lambda})\, f(\underset{\sim}{z}|s;\underset{\sim}{\phi}). \qquad [1.21]$$

We see from eq[1.16] and [1.21] that after selection the joint distribution of $\underset{\sim}{\tilde{y}}$ and $\underset{\sim}{\tilde{z}}$ given in [1.16] is modified into another joint distribution based on the sample given in [1.21].Thus under the two joint distributions [1.16] and [1.21] the conditional distribution of $\underset{\sim}{\tilde{y}}$ given $\underset{\sim}{\tilde{z}}=\underset{\sim}{z}$ is unaffected by selection on $\underset{\sim}{z}$.

Let the mean vector and the variance covariance matrix of $\underset{\sim}{\tilde{y}}$ and $\underset{\sim}{\tilde{z}}$ after selection be given by

$$\underset{\sim}{\mu}_s = \begin{bmatrix} \underset{\sim}{\mu}_{ys} \\ \underset{\sim}{\mu}_{zs} \end{bmatrix} \quad \text{and } \underset{\sim}{\Sigma}_s = \begin{bmatrix} \underset{\sim}{\Sigma}_{yys} & \underset{\sim}{\Sigma}_{yzs} \\ \underset{\sim}{\Sigma}_{zys} & \underset{\sim}{\Sigma}_{zzs} \end{bmatrix}$$

We now prove Lawleys'[1943] version of the Pearson adjustment.Lawley used arguments based on characteristic functions.We will give an alternative proof to his theorem using simple results based on conditional expectations and variances.

THEOREM 1.2 (LAWLEY'S THEOREM)

Under model 1 defined in section [1.3.1] the parameters in the unselected population can be reconstructed from those in the selected population by the following relationships;

$$\underset{\sim}{\mu} = \begin{bmatrix} \underset{\sim}{\mu}_y \\ \underset{\sim}{\mu}_z \end{bmatrix} = \begin{bmatrix} \underset{\sim}{\mu}_{ys} + \underset{\sim}{H} (\underset{\sim}{\mu}_z - \underset{\sim}{\mu}_{zs}) \\ \underset{\sim}{\mu}_z \end{bmatrix}$$

$$\underset{\sim}{H} = \underset{\sim}{\Sigma}_{yzs} \underset{\sim}{\Sigma}_{zzs}^{-1}$$

and

$$\underset{\sim}{\Sigma} = \begin{bmatrix} \underset{\sim}{\Sigma}_{yy} & \underset{\sim}{\Sigma}_{yz} \\ \underset{\sim}{\Sigma}_{zy} & \underset{\sim}{\Sigma}_{zz} \end{bmatrix}$$

$$= \begin{bmatrix} \underset{\sim}{\Sigma}_{yys} + \underset{\sim}{H}(\underset{\sim}{\Sigma}_{zz} - \underset{\sim}{\Sigma}_{zzs})\underset{\sim}{H}^T & \underset{\sim}{H}\,\underset{\sim}{\Sigma}_{zz} \\ \underset{\sim}{\Sigma}_{zz}\underset{\sim}{H}^T & \underset{\sim}{\Sigma}_{zz} \end{bmatrix}$$

PROOF. (Alternative proof to Lawley's theorem)

From Eq [1.16] we defined the joint distribution for the unselected population as;

$$f(\underset{\sim}{y},\underset{\sim}{z};\underset{\sim}{\lambda},\underset{\sim}{\phi})=f(\underset{\sim}{z};\underset{\sim}{\phi})f(\underset{\sim}{y}|\underset{\sim}{z};\underset{\sim}{\lambda}).$$

After selection,using Eq [1.21] this becomes

$$f(\underset{\sim}{y},\underset{\sim}{z}|s;\underset{\sim}{\lambda},\underset{\sim}{\phi})=f(\underset{\sim}{z}|s;\underset{\sim}{\phi})f(\underset{\sim}{y}|\underset{\sim}{z};\underset{\sim}{\lambda}). \qquad [1.22]$$

We therefore see that the conditional distribution of $\tilde{\underset{\sim}{y}}$ given $\tilde{\underset{\sim}{z}}=\underset{\sim}{z}_i$ is unaffected by selection on $\underset{\sim}{z}$ hence;

$$E_1(\tilde{\underset{\sim}{y}}|\underset{\sim}{z}_i,s)= \underset{\sim}{m} +\underset{\sim}{H}\ \underset{\sim}{z}_i. \qquad [1.23]$$

Similarly

$$V_1(\tilde{\underset{\sim}{y}}|\underset{\sim}{z}_i,s)=\underset{\sim}{K}. \qquad [1.24]$$

Taking expectations and variances of the expression [1.23] and [1.24] over the model distribution of $\tilde{\underset{\sim}{z}}_i=\underset{\sim}{z}_i$ given the sample s we get

$$E_z\ E_1(\tilde{\underset{\sim}{y}}|\underset{\sim}{z}_i,s)= E_z(\underset{\sim}{m}|s)+\underset{\sim}{H}\ E_z(\underset{\sim}{z}_i|s)$$
$$= \underset{\sim}{m}+\underset{\sim}{H}\ E_z(\underset{\sim}{z}_i|s). \qquad [1.25]$$

Let $\underset{\sim}{\mu}_{ys}=E_z(E_1(\tilde{\underset{\sim}{y}}|\underset{\sim}{z}_i,s))$ and $\underset{\sim}{\mu}_{zs}=E_z(\underset{\sim}{z}_i|,s)$,

then [1.25] becomes

$$\underset{\sim}{\mu}_{ys} =\underset{\sim}{m} +\underset{\sim}{H}\ \underset{\sim}{\mu}_{zs}. \qquad [1.26]$$

Let

$$\underset{\sim}{\Sigma}_{yys} = V(\tilde{\underset{\sim}{y}}|s)$$
$$=V_z(E_1(\tilde{\underset{\sim}{y}}|\underset{\sim}{z}_i,s)|s)+E_z(V_1(\tilde{\underset{\sim}{y}}|\underset{\sim}{z}_i,s)|s)$$
$$=V_z((\underset{\sim}{m}+\underset{\sim}{H}\ \underset{\sim}{z}_i)|s)+E_z(\underset{\sim}{K}|s),$$

using [1.23] and [1.24],

$$=\underset{\sim}{H}\, V_z(\underset{\sim}{z}_i|s)\, \underset{\sim}{H}^T + \underset{\sim}{K}. \qquad [1.27]$$

Let

$V_z(\underset{\sim}{z}_i|s)=\underset{\sim}{\Sigma}_{zzs}$, then [1.27] becomes

$$\underset{\sim}{\Sigma}_{yys} = \underset{\sim}{H}\, \underset{\sim}{\Sigma}_{zzs}\, \underset{\sim}{H}^T + \underset{\sim}{K}, \qquad [1.28]$$

To evaluate $\underset{\sim}{H}$, we consider,

$$\underset{\sim}{\Sigma}_{yzs} = \mathrm{cov}((\underset{\sim}{\tilde{y}},\underset{\sim}{\tilde{z}})|s)$$

$$=E(\underset{\sim}{\tilde{y}}\underset{\sim}{\tilde{z}}-E(\underset{\sim}{\tilde{y}})E(\underset{\sim}{\tilde{z}}))|s)$$

$$=E_zE_1((\underset{\sim}{\tilde{y}}\underset{\sim}{\tilde{z}}-E_1(\underset{\sim}{\tilde{y}}|\underset{\sim}{z}_i)E_1(\underset{\sim}{\tilde{z}}|\underset{\sim}{z}_i)+E_1(\underset{\sim}{\tilde{y}}|\underset{\sim}{z}_i)E_1(\underset{\sim}{\tilde{z}}|\underset{\sim}{z}_i)-E(\underset{\sim}{\tilde{y}})E(\underset{\sim}{\tilde{z}}))|s)$$

$$=E_z(\mathrm{cov}_1(\underset{\sim}{\tilde{y}},\underset{\sim}{\tilde{z}}|\underset{\sim}{z}_i,s))+\mathrm{cov}_z(E_1(\underset{\sim}{\tilde{y}}|\underset{\sim}{z}_i,s),E_1(\underset{\sim}{\tilde{z}}|\underset{\sim}{z}_i,s))$$

$$=\mathrm{cov}_z(E_1(\underset{\sim}{\tilde{y}},\underset{\sim}{z}_i|\underset{\sim}{z}_i,s)|s),$$

since $E_1(\underset{\sim}{\tilde{z}}|\underset{\sim}{z}_i,s)=\underset{\sim}{z}_i$ and $\mathrm{cov}_1(\underset{\sim}{\tilde{y}},\underset{\sim}{\tilde{z}}|\underset{\sim}{z}_i,s)=0$

$$=\mathrm{cov}_z(\underset{\sim}{m}+\underset{\sim}{H}\underset{\sim}{z}_i,\underset{\sim}{z}_i|s),$$

using (1.23) and (1.24)

$$=\underset{\sim}{H}V_z(\underset{\sim}{z}_i|s).$$

$$=\underset{\sim}{H}\underset{\sim}{\Sigma}_{zzs}.$$

If $\underset{\sim}{\Sigma}_{zzs}$ is nonsingular then

$$\underset{\sim}{H}=\underset{\sim}{\Sigma}_{yzs}\,\underset{\sim}{\Sigma}_{zzs}^{-1}. \qquad [1.29]$$

Substituting [1.26] in [1.10],[1.28] in [1.12] and [1.29] in [1.13] we get the required result.

We note that in deriving Lawley's theorem,no distribution is assumed and the nature of the selection need not be known provided that it is ignorable i.e of the form $f(s|\underset{\sim}{z})$.

COROLLARY 1.1

Let $\underset{\sim}{x}=(z,z^1,z^2...z^k)$ where k is finite i.e $\underset{\sim}{x}$ is a nonlinear function of z, it can be shown that by conditioning on $\underset{\sim}{x}$ Lawley's theorem also holds.

We can therefore conclude from theorem [1.2] and corr[1.1] that Lawley's theorem holds even for curvilinear models.

By nonlinear here we mean nonlinear in the design variables $\underset{\sim}{z}_i$.Thus Lawley's results can be applied to general distributions with curvilinear regressions.Since multivariate normality implies linearity and homoscedasticity,Lawley's result is a relaxation of the multivariate normality assumptions.Pearson's main result was rediscovered by Anderson [1957] in the form of maximum likelihood estimators,Cohen[1951] considered the case of a single design variable and Smith[1978] extended the results to more than one design variable.Smith[1981] applied Lawley's theorem to situation where the Multivariate normality condition holds to get the maximum likelihood estimators.We will give an example of the application of Lawley's theorem when the population is multivariate normal.

EXAMPLE

After selecting a sample of size n using a known sampling scheme and if assumption [1.1] holds then the likelihood function given in [1.5] simplifies to the likelihood of $\underset{\sim}{\lambda}$ and $\underset{\sim}{\phi}$ given by

$$L(\underset{\sim}{\lambda},\underset{\sim}{\phi})\propto f(\underset{\sim}{Z}_U,\underset{\sim}{\phi})f(\underset{\sim}{y}_S|\underset{\sim}{Z}_U;\underset{\sim}{\lambda})$$

$$\propto L(\underset{\sim}{\phi})L(\underset{\sim}{\lambda}), \quad \text{see Anderson [1957].} \qquad [1.30]$$

Equation [1.30] implies that to get maximum likelihood estimators (MLE) of $\underset{\sim}{\lambda}$ and $\underset{\sim}{\phi}$ we can maximize separately the likelihood function of $\underset{\sim}{\lambda}$ and $\underset{\sim}{\phi}$ respectively.Using this property of the likelihood function Smith [1981] used Anderson's result and derived the MLE of $\underset{\sim}{\lambda}$ and $\underset{\sim}{\phi}$ under a noninformative sampling scheme.

This maximum likelihood estimator of the finite population mean is given by;

$$\hat{\underset{\sim}{\mu}}_y = \bar{\underset{\sim}{y}}_s + \underset{\sim}{B}_{yzs}(\bar{\underset{\sim}{z}}_U - \bar{\underset{\sim}{z}}_s), \qquad [1.31]$$

for scalars i.e p=1, and q=1 then [1.31] is the well known regression estimator of the finite population mean Cochran[1977],and the covariance matrix is given by;

$$\hat{\underset{\sim}{\Sigma}}_{yy} = \underset{\sim}{S}_{yys} + \underset{\sim}{B}_{yzs}(\underset{\sim}{S}_{zz} - \underset{\sim}{S}_{zzs})\underset{\sim}{B}^T_{yzs} , \qquad [1.32]$$

where

$$\underset{\sim}{S}_{yys} = n^{-1}\sum_s (\underset{\sim}{y}_i - \bar{\underset{\sim}{y}}_s)(\underset{\sim}{y}_i - \bar{\underset{\sim}{y}}_s)^T,$$

$$\underset{\sim}{S}_{zzs} = n^{-1}\sum_s (\underset{\sim}{z}_i - \bar{\underset{\sim}{z}}_s)(\underset{\sim}{z}_i - \bar{\underset{\sim}{z}}_s)^T ,$$

$$\underset{\sim}{S}_{yzs} = n^{-1}\sum_s (\underset{\sim}{y}_i - \bar{\underset{\sim}{y}}_s)(\underset{\sim}{z}_i - \bar{\underset{\sim}{z}}_s)^T ,$$

$$\underset{\sim}{S}_{zys} = n^{-1}\sum_s (\underset{\sim}{z}_i - \bar{\underset{\sim}{z}}_s)(\underset{\sim}{y}_i - \bar{\underset{\sim}{y}}_s)^T,$$

$$\underset{\sim}{B}_{yzs} = \underset{\sim}{S}_{yzs}\underset{\sim}{S}^{-1}_{zzs},$$

and

$\bar{\underset{\sim}{z}}_U$,$\bar{\underset{\sim}{z}}_s$ are the finite population mean and the sample mean of the design variables.

We see that expressions [1.31] and [1.32] are the Pearson type adjusted estimators of the mean and the population covariance matrix.

We now review randomization based approach to inference for both descriptive and analytical targets.

1.3.2 RANDOMIZATION BASED APPROACH TO INFERENCE

1.3.2.1 RANDOMIZATION BASED APPROACH TO INFERENCE FOR ANALYTICAL TARGETS.

In randomization inference to analytical targets the survey and design variables are regarded as fixed constants. The survey values are to be measured in the survey and the design values are assumed to be known. We define the degenerate case of the population model on which this inference is based as

$$f(\underset{\sim}{y}_i \mid \underset{\sim}{z}_i;\underset{\sim}{\lambda}) = \begin{cases} 1 & \text{if } \underset{\sim}{\tilde{y}}_i = \underset{\sim}{y}_i \quad i \varepsilon U, \\ 0 & \text{otherwise.} \end{cases} \qquad [1.33]$$

and

$$f(\underset{\sim}{z}_i;\underset{\sim}{\phi}) = \begin{cases} 1 & \text{if } \underset{\sim}{\tilde{z}} = \underset{\sim}{z}_i \quad i \varepsilon U, \\ 0 & \text{otherwise.} \end{cases} \qquad [1.34]$$

If we select a sample of size n using a known sampling scheme $f(s|\underset{\sim}{z})$, then using [1.33] and [1.34] the joint distribution of $\underset{\sim}{y}_S$ and $\underset{\sim}{z}$ is given by

$$f(\underset{\sim}{y}_i,\underset{\sim}{z}_i;\underset{\sim}{\lambda}\ \underset{\sim}{\phi}) = \begin{cases} 1 & \text{if } \underset{\sim}{\tilde{y}} = \underset{\sim}{y}_i \text{ and } \underset{\sim}{\tilde{z}} = \underset{\sim}{z}_i\ i \in s, \\ 0 & \text{otherwise.} \end{cases} \qquad [1.35]$$

From [1.5] under assumption [1.1] the likelihood function of $\underset{\sim}{\lambda}$ and $\underset{\sim}{\phi}$ is given by

$$L(\underset{\sim}{\lambda},\underset{\sim}{\phi}) \propto f(s|\ \underset{\sim}{z}) f(\underset{\sim}{y}_S,\underset{\sim}{z}_S;\underset{\sim}{\lambda},\underset{\sim}{\phi}).$$

Since s and $\underset{\sim}{z}$ are known $f(s|\ \underset{\sim}{z})$ is a constant, hence we get

$$L(\underset{\sim}{\lambda},\underset{\sim}{\phi}) \propto \begin{cases} \text{constant}, & \underset{\sim}{\tilde{y}} = \underset{\sim}{y}_i, \underset{\sim}{\tilde{z}} = \underset{\sim}{z}_i, \\ 0 & \text{otherwise.} \end{cases} \qquad [1.36]$$

We see from [1.36] that the likelihood is constant for populations that have generated the sample and is zero for those which could not generate the sample.Thus the likelihood is a step function and partitions the populations into two sets ,those with zero likelihood and those with constant nonzero likelihood.The constant likelihood given in [1.36] provides no means of choosing one member rather than the other since it lacks a unique maximum.We call such a likelihood a *noninformative likelihood.* Thus the likelihood function does not depend on the parameters $\underset{\sim}{\lambda}$ and $\underset{\sim}{\phi}$ and cannot be used for inference about them. Therefore analytical inference based on likelihood function is impossible under the degenerate model assumed in randomization approach.Another factor which has led the likelihood inference to randomization approach to its graveyard is the fact that since the likelihood function is constant for all the units in the sample there doesnot exist a unique maximum likelihood estimator of $\underset{\sim}{\theta}=(\underset{\sim}{\lambda},\underset{\sim}{\phi})$;see Godambe [1966].

However if the finite population is assumed to be very large i.e close to the superpopulation, and is a random sample from this population,we can regard it as approximating the superpopulation, and the finite population parameters are approximately equal to their superpopulation analogues.

Example.

Let the finite population be very large and assume that it is a random sample from the infinite population.Then if $\bar{y}_U$ is the finite population mean and μ_y its corresponding superpopulation analogue,then for arbitrarily small $\varepsilon \div 0$ by Chebyshev's weak law of large numbers we have

$$p[|\bar{y}_U-\mu_y|>\varepsilon]\le \frac{\sum_U var(y_i)}{N^2\ \varepsilon^2} .$$

Assume that $var(y_i)=\sigma^2$ for all $i\in U$,

Then we get

$$p[|\bar{y}_U-\mu_y|>\varepsilon]\le \frac{\sigma^2}{N\ \varepsilon^2} .$$

We see from this expression that if N is large then

$\lim_{N\to\infty} p[|\bar{y}_U-\mu_y|>\varepsilon]=0$, therefore $\bar{y}_U\to \mu_y$ in probability as $N\to\infty$,

That is the finite population mean converges to its superpopulation analogue for large N and so inference about the finite population mean is essentially the same as that about its superpopulation analogue, provided that the finite population is an i.i.d sample from the superpopulation.

A randomization estimator of $\bar{y}_U$ will be an estimator of μ_y with an error of $O_p(N^{-1/2})$.

1.3.2.2 RANDOMIZATION BASED APPROACH TO INFERENCE FOR DESCRIPTIVE TARGETS

In section 1.3 we defined I(s) the indicator function which identifies those units of the finite population U which are included in the sample s. Randomization inference generally requires the units to be selected by a probability rule, which is characterised by the following properties.

(i) The sample design is determined by the sampler before the survey is carried out. Thus the sample design cannot depend on the survey values to be measured in the survey. Such sample designs are called noninformative sample designs.

(ii) Every unit has a known positive probability of selection. Writing $\pi_i=E(I_i(s)|z)=p(I_i(s)=1|z)$, we require that $\pi_i>0$ for all i. In equal probability sample designs, such as the simple random sample design, this probability is same for all units.

Let S denote the set of all the $\binom{N}{n}$ possible samples of fixed size n which can be drawn from the finite population of size N. Since the only source of randomization is the sampling scheme $f(s|\underset{\sim}{z})$,the probability distribution on S determined by the sampling scheme $f(s|\underset{\sim}{z})$ is called the randomization distribution.The criterion of inference which has been adopted by those who advocate randomization inference like Kish and Frankel[1974],Cochran[1977] etc. is based on this distribution averaged over all the $\binom{N}{n}$ possible samples. This forms the basis of randomization inference under repeated sampling.

Example.

It may be possible to divide the population into strata within which units are relatively homogenous.The units may then be selected by simple random sampling within each strata.Let $N_1..N_H$ denote the H strata in the population such that $\sum_{h=1}^{H} N_h=N$.Select a sample of size n_h in each strata of size N_h such that $\sum_{h=1}^{H} n_h=n$.Let z_{ih} denote the priori information of the i^{th}unit in the h^{th}strata.Then the sample design is defined by

$$f(s|z)=\begin{cases} \prod_{h=1}^{H} 1/\binom{N_h}{n_h} & \text{if } \sum_{h=1}^{H} I_{ih} =n_h, \\ 0 & \text{othewise.} \end{cases}$$

where $\binom{N_h}{n_h}$ is the number of possible samples of size n_h which can be drawn from the population of size N_h.

If $\hat{M}$ is an estimator of M then we say that $\hat{M}$ is unbiased with respect to sample design $f(s|\underset{\sim}{z})$,i.e p-unbiased,if

$$E_p(\hat{M})=M,$$

where $E_p(.)$ denotes expectation with respect to the sample design $f(s|\underset{\sim}{z})$.

We say that $\hat{M}$ is asymptotically unbiased estimator of M under the sample design $f(s|\underset{\sim}{z})$ if

$$\lim_{\substack{N\Rightarrow\infty \\ n\Rightarrow\infty \\ n/N\Rightarrow c(\text{constant})}} E_p(\hat{M})=M \text{ (where M is a superpopulation parameter).}$$

Cleary if $\hat{M}$ is a p-unbiased esimator of M then it is also an asymptotically unbiased estimator of M, but not vice versa.e.g the ratio estimator of say finite population mean is biased but it is asymptotically unbiased.

Similarly if ε denotes a model then $\hat{M}$ is an asymptotically conditionally unbiased estimator of M with respect to model ε if

$$\lim_{\substack{N\Rightarrow\infty \\ n\Rightarrow\infty \\ n/N\Rightarrow c(\text{constant})}} E_\varepsilon(\hat{M}|z,s)=M \text{ (where M is a superpopulation parameter).}$$

Randomization based estimators are chosen using properties such as p-unbiasedness or asymptotic unbiasedness .

EXAMPLE

If D is some function of Y of interest,then to choose D_s as an estimator of D we establish conditions under which D_s is unbiased with respect to the design i.e p-unbiased.

Let $D=\sum_U h(y_i)$.

Consider estimators of D which are linear in the indicator variables $I_i(s)$ i.e

$$D_s = \sum_U w_i h(y_i)\ I_i(s).$$

Taking expectations with respect to the design we get

$$E_p(D_s)= \sum_U w_i h(y_i)\ \pi_i, \qquad [1.37]$$

since $E_p(I_i(s))=\pi_i$ the probability of including the i^{th} unit in the sample.

From [1.37] we see that the estimator D_s is p-unbiased if

$w_i=\pi_i^{-1}$, the inverse of the probability weight.

Hence

$$E_p(D_s)= \sum_U h(y_i) =D.$$

If $D=\bar{y}=N^{-1}\sum_U y_i$ denotes the finite population mean then using results above its estimator $D_s=\sum_s w_i y_i$ is p-unbiased where $w_i=\pi_i^{-1}$.This is the Horvitz-Thompson estimator of the finite population mean,hence the term inverse probability weighting. The randomization approach to inference has concentrated on the estimation of simple statistics like the means,totals etc.Kish and Frankel[1974] extended this approach to the estimation of more complex statistics such as the regression and correlation coefficients within the finite population framework. Their parameter of interest was the finite population analogue of the superpopulation parameter. In the case of regression analysis, the finite population regression coefficient is given by

$$\underset{\sim}{B}_N=(\underset{\sim}{Y}_2^T\underset{\sim}{Y}_2)^{-1}\ \underset{\sim}{Y}_2^T\ \underset{\sim}{y}_1, \qquad [1.38]$$

where $\underset{\sim}{Y}_2=[\underset{\sim}{y}_{21}\ldots\ldots\underset{\sim}{y}_{2N}]^T$ is a Nxp_2 matrix of explanatory or independent variables,and $\underset{\sim}{y}_1=[y_{11}\ldots\ldots y_{1N}]^T$ is a Nx1 vector of dependent variables.

Define the components of $\underset{\sim}{B}_N$ as the sums of products denoted by $\underset{\sim}{V}$ and $\underset{\sim}{Q}$ where

$$\underset{\sim}{V}= \underset{\sim}{Y}_2^T\underset{\sim}{Y}_2,$$

and

$$\underset{\sim}{Q}= \underset{\sim}{Y}_2^T\underset{\sim}{y}_1.$$

Using results from the above example the p-unbiased estimators of $\underset{\sim}{V}$ and $\underset{\sim}{Q}$ are given by

$$\underset{\sim}{V}_s=\underset{\sim}{Y}_{2s}^T\underset{\sim}{w}_s\ \underset{\sim}{Y}_{2s},$$

and

$$\underset{\sim}{Q}_s=\underset{\sim}{Y}_{2s}^T\underset{\sim}{w}_s\ \underset{\sim}{y}_{1s},$$

where $\underset{\sim}{Y}_{2s}=[\underset{\sim}{y}_{21}\ldots\ldots\underset{\sim}{y}_{2n}]^T$ is a nxp_2 matrix of explanatory or independent variables, $\underset{\sim}{y}_{1s}=[y_{11}\ldots\ldots y_{1n}]^T$ is a nx1 vector of dependent variables,and $\underset{\sim}{w}_s=\mathrm{diag}(w_1\ldots\ldots w_n)$ is a nxn matrix with $w_i=\pi_i^{-1}$. Since $\underset{\sim}{B}_N$ is a function of $\underset{\sim}{V}$ and $\underset{\sim}{Q}$, and $\underset{\sim}{V}_s$ and $\underset{\sim}{Q}_s$ are p-unbiased estimators of $\underset{\sim}{V}$ and $\underset{\sim}{Q}$ therefore we define $\underset{\sim}{B}_s$ an estimator of $\underset{\sim}{B}_N$ as

$$\underset{\sim}{B}_s=(\underset{\sim}{Y}_{2s}^T \underset{\sim}{w}_s \underset{\sim}{Y}_{2s})^{-1} \underset{\sim}{Y}_{2s}^T \underset{\sim}{w}_s \underset{\sim}{Y}_{1s}. \qquad [1.39]$$

$\underset{\sim}{B}_s$ is an asymptotically unbiased estimator of $\underset{\sim}{B}_N$ see Campbell [1977].

Kish and Frankel [1974] studied the properties of the estimator [1.39] in clustered stratified samples. They considered the case of one independent variable and equal weights in all strata. Thus their sample design was the clustered sample design, but since they considered only the case of equal weights i.e equal probability design, the estimator [1.39] reduces to the ordinary least square regression estimator.

Once an estimator has been selected its randomization variance is estimated. Kish and Frankel[1974] suggested three methods of estimating sampling variances for repeated samples using the complex designs. These are;

(a) Taylor series method,

(b) Balanced repeated replication,

and the

(c) Jackknife method.

Inferences are made by the use of the central limit theorem.

1.3.3 LIMITATIONS

The randomization approach appeals to most analysts because randomization protects the estimators derived from many biases for example, individual selection bias. Those who advocate randomization methods contend that these methods have always been used in Government surveys and the results have never caused a national outcry, so they say that randomization approach is good because it always 'works'. This was challenged by Smith[1983] who compared the actual result of the National election in Britain with the results of two of the main opinion polls results,the National opinion polls which uses random samples and Gallup polls which uses nonrandom samples.His results shown that the Gallup polls 'works' as well as the national opinion polls.

Since randomization inference is not based on any model assumptions it is completely robust to model mispecfications.The advocates of this inference claim that it does not depend on any assumptions.This is not always true.Smith[1979] have shown that it is possible to construct populations where the central limit theorem(CLT) does not hold even for large samples.Since randomization inference depends on this assumption it fails for those populations constructed by Smith[1979].

EXAMPLE

Suppose a survey is conducted to investigate the number of people having a rare disease.The rate of occurence of the disease,denoted by λ,is very small i.e one out of 500 people is found to have the disease.If $\tilde{d}$ denotes the number of people in the population having the disease ,d its realized value and n the size of the sample taken,then the Poisson approximation to the distribution of $\tilde{d}$ is

$$p(\tilde{d}=d)=(n\lambda)^{d}\exp(-n\lambda)/d^{\circ}$$

If we consider a sample of size n=1000, $\lambda=0.002$ then the probability of having no diseased people in the sample is

$$p(\tilde{d}=0)=e^{-2}=0.14$$

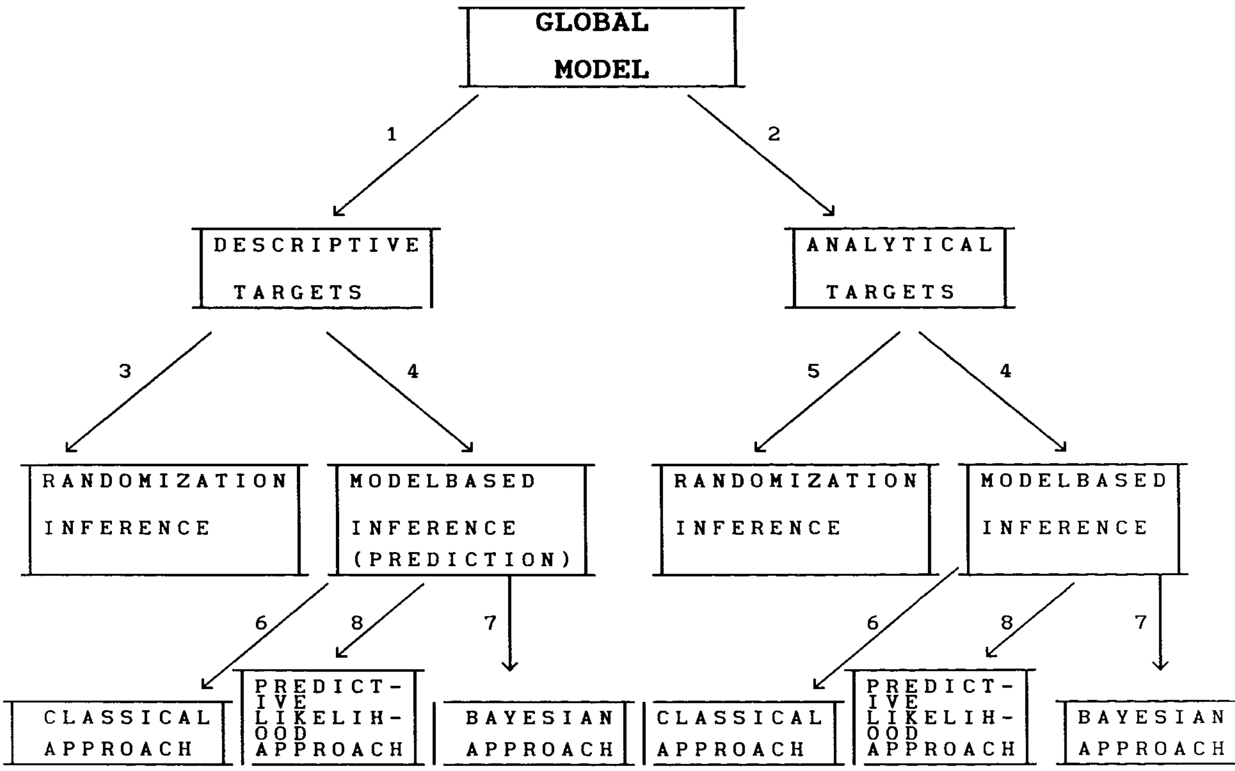
Thus 14% of the samples will contain people without the disease.If $\hat{d}$ is an estimate of the number of people with the disease then for these samples standard error (s.e) $\hat{d}=0$ hence,

$$z=\frac{\hat{d}-d}{s.e(\hat{d})}$$ is undefined for 14% of the samples.

Such a sampling distribution is not normal since it has a large chunk of probability corresponding to $\tilde{d}=0$.Therefore for this type of population the CLT does not hold.

Thus for the randomization inference to be feasible we need to make some assumptions to exclude those population where the CLT does not hold.Randomization inference also assumes a noninformative type of design.Suppose the sample selection is not independent of the values of the survey variables given those of the design variables,then the sample design becomes informative.This type of sample design which depends on the values of the survey variables as well as on those of design variables is no longer a random selection process.Examples of designs of this type are Quota sampling designs; see Smith [1983] and designs based on non response; see Little[1982].Since randomization inference is based on the finite population values which are taken to be constant,conditional inference is not possible, only unconditional inference to descriptive targets is possible. To be able to make conditional inference on analytical targets possible within this framework we need to specify a nondegenerate population model. Faced with this kind of situation the advocates of the randomization inference ignore these problems and proceed under the cover of randomization.We conclude this chapter by summarizing all the different approaches to statistical inference diagramatically.

1.4 SUMMARY OF THE STATISTICAL APPROACHES TO INFERENCE REVIEWED



where

1. Parameter of interest is a known function of $\underset{\sim}{Y}_U$.
2. Parameter of interest cannot be expressed as a function of $\underset{\sim}{Y}_U$.
3. The random variables $\underset{\sim}{\tilde{Y}}_U$ and $\underset{\sim}{\tilde{Z}}_U$ are fixed constants.
4. $\underset{\sim}{Y}_U$ and $\underset{\sim}{Z}_U$ are realized values of random Variables $\underset{\sim}{\tilde{Y}}_U$ and $\underset{\sim}{\tilde{Z}}_U$ respectively.
5. If the finite population is a random sample from the superpopulation such that N is very large. In that case the finite population parameters concide with those of superpopulation.
6. The parameter $\underset{\sim}{\lambda}$ is regarded as fixed and is estimated from the conditional distribution $f(\underset{\sim}{Y}_s|\underset{\sim}{Z}_U;\underset{\sim}{\lambda})$, which may not be known exactly. The estimated value of $\underset{\sim}{\lambda}$ is then used in the conditional distribution $f(\underset{\sim}{Y}_{\bar{s}}|\underset{\sim}{Z}_U;\underset{\sim}{\lambda})$ to

infer about the unobserved units.

7. Treats $\underset{\sim}{\lambda}$ as a realized value of the random variable $\underset{\sim}{\tilde{\lambda}}$ and assumes that the prior distribution of $\underset{\sim}{\tilde{\lambda}}$ is known. The unobserved values are obtained from the conditional distribution $f(\underset{\sim}{Y}_{\bar{S}}|\underset{\sim}{Z}_U;\underset{\sim}{Y}_S)$ implied by the posterior distribution$f(\underset{\sim}{\lambda}|\underset{\sim}{Y}_S,\underset{\sim}{Z}_U)$.

8. Assumes that the exact sample distribution $f(\underset{\sim}{Y}_S|\underset{\sim}{Z}_U;\underset{\sim}{\lambda})$ is exactly kno w n, but the priori distribution may be unknown, then the unknown value of $\underset{\sim}{\lambda}$ is regarded as equal to either the maximum likelihood estimate or a sufficient statistic for $\underset{\sim}{\lambda}$.

1.5 OUTLINE OF THE THESIS

Throughout this thesis we assume that the sample selection scheme is noninformative, that is it is of the form $f(s|z)$.We will also assume that the aggregate model is the appropriate model to fit for the whole set of data,that is in our regression problem,we fit only one regression equation to the whole set of data.Our interest is in the robust estimation of the superpopulation parameters,hence our focus is the analytical inference to these superpopulation parameters and their performance in different multivariate situations.We present the outline of the thesis in table 1.2.

Table 1.2 Summary of the work done in this thesis

CHAPTER	
two	Review of various methods of deriving parametric estimators of the covariance matrix and the mean.
three	Derived the asymptotic properties of the Fuller estimators when; (i) linear and homoscedastic model assumptions hold. (ii) linearity assumption is violated, but homoscedasticity assumption holds. (iii) Homoscedasticity assumption is violated but linearity assumption holds. We derived the conditional variances of the Fuller and maximum likelihood estimators under the linear homoscedastic model and investigated which of the two optimal estimators is better in terms of minimum variance.
four	We investigated empirically the asymptotic properties of the Fuller estimators and compared its empirical properties with the ordinary least squares ,maximum likelihood,probability weighted and probability weighted adjusted estimators when; (i) linear and homoscedastic assumptions are satisfied and under violation of, (ii) linearity assumption, (iii) Homoscedasticity assumption.

TABLE 1.2 CONT...

five	We derived nonparametric estimators of the mean and the covariance matrix. Chosing kernel estimators as one type of nonparametric estimators we proved that the kernel estimators of the mean and covariance matrix which we derived are consistent as sample,population size tend to infinity and the bandwidth parameter tends to zero.
six	We studied empirically the performance of the nonparametric estimators and compared their performance with the maximum likelihood,probability weighted adjusted and the unweighted Fuller estimators.
seven	Conclusion and recomendations.

CHAPTER 2

ALTERNATIVE PARAMETRIC REGRESSION ESTIMATORS

2.1 ROBUSTNESS

The term robustness was introduced by Box(1953) but the term does not lend itself to a clear statistical definition.According to Kendall and Buckland(1981) :-

"Many test procedures involving probability levels depend for their exactitude on assumptions concerning the generating mechanism e.g that the parent variation is normal.If the inferences are little affected by departures from these assumptions then,the inferences are said to be 'robust'.In a rather more general sense a statistical procedure is described as robust if it is not sensitive to departures from the assumptions on which it is based".

A robust statistical procedure can be thought of as one which performs well over a range of situations and is able to stand up to a certain amount of abuse without breaking down.Model based inferences are carried out by assuming a parametric model with a specified distributional form such as in linear models with normally distributed errors.The classical approach to inference in such a setting is to estimate the parameter of interest in some optimal way,such as least squares.There is a potential problem with this approach.In most cases our model is an approximation and not strictly true.Thus the estimate obtained may not be optimal for the correct model,thus to claim that the estimate obtained is optimal to the correct model we must rely on the principle of continuity that guarantees that a small change in the model results in a small change from the optimal.However many optimization procedures do not have this property.We consider robust estimators to be ones which have the property that they are continous in the neighbourhood of the idealized model,hence robustness can be thought as a protection for deviations from the underlying model distribution.

A lot of work has been done to devise robust statistical procedures.Quenouille(1956) proposed the jackknife techniques which permits the reduction of the bias and variance of most estimators regardless of the distribution underlying the data.The statistician is thus relieved from some frequently questionable distributional assumptions.Anscombe (1960) reviewed the principles underlying the rejection of outliers and stimulated theoretical and experimental research on how to account for observations appearing in the tails of the sample distributions.Huber(1964) introduced the M-estimators and proposed a most robust estimator among them for the estimation of the location when the underlying distribution is a contaminated normal.

There is a well developed theory in sample surveys e.g Godambe(1966),Kish and Frankel(1974) etc. for both estimation and testing that yield procedures which give up a certain percentage of efficiency when the model is correct and maintains a high efficiency in the neighbourhood when the model is not correct.These procedures are based on the randomization distribution.Randomization based procedures have been thought to offer robust alternative procedures to the model based procedures.However there are problems in this approach.There are two types of distributional framework ,the conditional and the unconditional framework.The model based procedures are based on the conditional distribution,conditioning on the sample and the values of the design variables whereas in randomization procedures the distribution of the survey variables is averaged over all possible samples.The question then is "How can these two procedures be compared when they are based on two different distributional frameworks?".Using the bias criterion as a measure of robustness all randomization based procedures are found to be robust.

We give the following example from Skinner,Holt and Smith (1989) to demonstrate how this robustness criterion is inappropriate for comparing estimators defined in the two frameworks. Consider a population with a natural stratification and suppose this is ignored and a simple random sample (srs) design is used.The only apparent effect of using the srs design and the corresponding srs estimator within randomization inference is a loss of efficiency relative to the stratified design see Kish and Frankel(1974).The bias due to misspecification under the model based framework becomes an increase in variance in the randomization framework.Thus as long as the estimator is unbiased relative to the chosen design,randomization inference is independent of the population structure and even the most inefficient sample estimator may be robust in the randomization framework.We therefore need another criterion for measuring robustness in the two distributional frameworks.We propose below a criterion for measuring robustness which does not give conflicting results in the two frameworks.This criterion is the unconditional mean square error of the estimator.If we denote by $\hat{\theta}$,the estimator of the parameter of interest θ, then we define the unconditional mean square error of $\hat{\theta}$, as;

$$MSE(\hat{\theta})=var(\hat{\theta})+(E(\hat{\theta})-\theta)^2,$$

where $var(\hat{\theta})$ denotes the unconditional variance of the estimator and $E(\hat{\theta})$ denotes the unconditional expectation of $\hat{\theta}$.

We see that this criterion accounts for both the variance and the bias of the estimator.We propose the use of this unconditional mean square error criterion for comparing these alternative estimators from both the randomization and the model based distributional frameworks.

Empirical studies by Nathan and Holt(1980) using the bias criterion as a measure of robustness found that the adjusted Pearson type estimators given in (1.31) and (1.32) were very sensitive to the violations of the linearity and homoscedasticity assumptions.Thus they are not *robust* to the mispecification of the model.Though the design based estimators are robust to any mispecification in the model,they were found to be very inefficient for unequal probability designs.In the search for a robust efficient estimator attempts have been made to incorporate the weights in randomization based estimators in model based estimators. Nathan and Holt (1980) proposed the weighted version of the estimators (1.31) and (1.32) which they considered to be a kind of a compromise estimator between the non robust efficient model based estimators and the robust inefficient design based estimators.Through some empirical studies Holmes(1987) found that the Nathan and Holt compromise estimator is conditionally biased when the linearity and homoscedasticity assumptions are violated and did not have any significant gain in efficiency over the probability weighted estimator.Rubin(1985) suggested replacing the design variables z in the model defined in eq [1.16],chapter 1 with the probability weights and shows that under certain conditions the probability weights π_i are an adequate summary of the information on z.He indicated that it might be easier to construct a conditional distribution of y given π_i than on y given z.Smith (1988) shows that if the probability weights π_i are measures of size,then Rubin's approach can be interpreted as size biased sampling.Heshows that if the exact form of eq [1.16] is not known ,then methods of moments lead to probability weighted estimators similar to those used in randomization inference.Thus probability weighted estimators are the most appropriate estimators for size biased sampling designs.Holmes(1987) among others have shown that these probability weighted estimators are severely conditionally biased and not very efficient.The search for a robust efficient estimator is the theme of this thesis and led us to look at Chambers (1986) paper which treated the problem of estimating the superpopulation parameters using the ideas of a loss function.

2.2 CHAMBERS' APPROACH

According to Chambers (1986), if all the values in the population are observed then the optimal estimator $\underset{\sim}{\theta}_N$ of $\underset{\sim}{\theta}$ will be that value which minimizes an appropriately chosen loss function. Denote this loss function by $L_N(\underset{\sim}{y}_U;\underset{\sim}{\theta})$ where $\underset{\sim}{y}_U$ is an Nx1 vector of survey variables, i.e $\underset{\sim}{y}_U=(y_1 \ldots y_N)^T$ then $\underset{\sim}{\theta}_N$ is chosen such that

$$L_N(\underset{\sim}{y}_U;\underset{\sim}{\theta}_N)=\text{Min}(L_N(\underset{\sim}{y}_U;\underset{\sim}{\theta}),\underset{\sim}{\theta}\in\Theta).$$

The choice of this population loss function will depend on how tightly the conditional distribution $f(\underset{\sim}{y}_U|\underset{\sim}{z}_U;\lambda)$ has been defined.

(i) If the conditional distribution $f(\underset{\sim}{y}_U|\underset{\sim}{z}_U;\lambda)$ is completely specified then the population loss function $L_N(\underset{\sim}{y}_U;\underset{\sim}{\theta})$ may correspond to the inverse of the likelihood function.

(ii) If only two moments are known as in most regression problems, then the loss function $L_N(\underset{\sim}{y}_U;\underset{\sim}{\theta})$ may be chosen to be least squares.

In practice it is usually impossible to observe all the survey values in the whole finite population. We therefore rely on a sample which is selected from the finite population. The process of selecting the sample may or maynot depend on the survey values $\underset{\sim}{Y}_U$. If we regard the sample selected as our whole population, then the loss function based on this sample denoted by $L_n(\underset{\sim}{y}_s;\underset{\sim}{\theta}_s)$ is given by;

$$L_n(\underset{\sim}{y}_s;\underset{\sim}{\theta}_s)=\text{Min}(L_n(\underset{\sim}{y}_s;\underset{\sim}{\theta}),\underset{\sim}{\theta}\in\Theta),$$

where $\underset{\sim}{\theta}_s$ denotes the estimator of $\underset{\sim}{\theta}$ based on the sample of size n. Based only on the observed values $\underset{\sim}{y}_s$ we have the problem;

'How do we choose an estimator $\underset{\sim}{\theta}_s$ for $\underset{\sim}{\theta}$ by minimizing the loss function defined by the conditional distribution of $\tilde{y}_U$ given z_U?'.

Like in section 1.3.1.2 ,the parameter of interest θ does not occur explicitly in this conditional distribution.

The conditional distribution $f(y_U|z_U;\lambda)$ factorises to;

$$f(y_U|z_U;\lambda)=f(y_s|z_U;\lambda)f(y_{\bar{s}}|y_s,z_U;\lambda).$$

Since λ, which indexes $f(y_{\bar{s}}|y_s,z_U;\lambda)$, is unknown,some estimated value of λ must be used.Chambers used the Expectation Mi n imi z ation(EM) algorithm described by Dempster,Laird and Rubin(1977) to motivate an approach to adjusting $L_N(y_U,\theta)$ for the missing observations $y_{\bar{s}}$.He suggested a one step Expectation Minimization algorithm as follows;

(i) Calculate θ_s from the sample data as an optimal estimate of λ given z.Replace the unknown parameter λ by $\theta_s$and take expectation of the loss function i.e

$$E(L_N(y_U;\theta)|y_s,z_U)=\int L_N(y_U;\theta)f(y_{\bar{s}}|y_s,z_U)\ dy_{\bar{s}}\ ,$$

then choose an estimator $\hat{\theta}_N$ which minimizes this expectation i.e

$$E(L_N(y_U;\theta)|y_s,z_U)=\text{Min}\int L_N(y_U;\theta)f(y_{\bar{s}}|y_s,z_U)\ dy_{\bar{s}}\ .$$

To evaluate this expectation Chambers considered a multivariate Normal case and obtained estimators identical to the adjusted Pearson estimators derived in [1.31] and [1.32].

Example.

Assume that the survey variables and the design variables have a joint i.i.d normal distribution with means

μ_y , μ_z and variances σ_y^2 and σ_z^2. Since the joint distribution here is completely known using chambers' result the most appropriate loss function is the inverse of the log likelihood given by,

$$L_N(\underset{\sim}{y}_U;\underset{\sim}{\theta})=\sum_U (y_i-\mu_y)^2\sigma_y{}^2 + N\log\sigma_y{}^2.$$

In this case the parameter of interest is $\theta=(\mu_y,\sigma_y^2)$. By analogue with the generalized EM algorithm, chambers minimized the above loss function to obtain the following expression for the estimator of the variance of y, denoted by $\hat{S}_C$ given by;

$$N\hat{S}_C = \sum_s y_i^2 + E_1(\sum_{\bar{s}} y_i^2|s,z;\lambda) - NE_1(\mu_y^2|s,z;\lambda),$$

where $E_1(.)$ denotes conditional expectation under model 1 defined in chapter1.
using Taylors'approximation we get,

$$N\hat{S}_C \cong \sum_s y_i^2 + E_1(\sum_{\bar{s}} y_i^2|s,z;\lambda) - N(E_1(\mu_y|s,z;\lambda))^2$$

$$=\sum_s y_i^2 + \sum_{\bar{s}} E_1(y_i^2|s,z;\lambda) - N(E_1(\mu_y|s,z;\lambda))^2$$

$$\cong nS_{yys} + n\bar{y}_s^2 + \sum_{\bar{s}}[V_1(y_i|s,z;\lambda) + (E_1(y_i|s,z;\lambda))^2]$$

$$-N(E_1(\bar{y}_U|s,z;\lambda))^2,$$

since $\mu_y = \bar{y}_U + O_p(N^{-1/2})$,
S_{yys} denotes the sample variance,
and $\bar{y}_U, \bar{y}_s$ denotes the finite population and sample mean respectively.

$$= nS_{yys} + n\bar{y}_s^2 + \sum_{\bar{s}} [S_{yys} - \hat{H}^2 S_{zzs} + (\bar{y}_s + (z_i - \bar{z}_s)\hat{H})^2] - N(\hat{\bar{y}}_U)^2$$

using model 1 assumptions.

$$=nS_{yys}+n\bar{y}_s^2+\sum_{\bar{s}} [S_{yys}-\hat{H}^2S_{zzs}+(\bar{y}_s+(z_i-\bar{z}_s)\hat{H})^2]-N[\bar{y}_s-\hat{H}(\bar{z}_U-\bar{z}_s)]^2,$$

since $\hat{\bar{y}}_U=\bar{y}_s-\hat{H}(\bar{z}_U-\bar{z}_s)$ from theorem 1.1.

simplifying this expression we get,

$\hat{S}_c=S_{yys}+\hat{H}^2(S_{zz}-S_{zzs})$,which is the Pearson adjusted estimator of the variance.

Therefore the use of loss functions does not give a new estimator,but is another vehicle of deriving the Pearson adjusted estimators of the covariance matrix.On this basis we saw no need of pursuing this approach further.

In the next section we will derive the estimators of the covariance matrix using quadratic predictors.

2.3 ESTIMATION OF THE COVARIANCE MATRIX USING QUADRATIC PREDICTORS

While prediction of linear functions of $\tilde{y}_i$, such as the finite population mean has received much attention,the prediction of quadratic functions of $\tilde{y}_i$,such as the finite population variance,has received less attention. Mukhopadhyay(1978) derived the predictor of the finite population variance under the normal model assumptions for a single survey variable with zero means. Skinner(1983) derived the minimum variance unbiased predictors of the finite population mean and covariance matrix under the assumption that the variable $\tilde{y}_i$ and $\tilde{z}_i$ are generated by a multivariate normal model.In this work we will relax the normality model assumptions made by Skinner (1983) and derive consistent estimators of the finite population covariance matrix using the predictors of the quadratic function of the survey variables in the case where the population is not clustered.

The finite population covariance matrix is given by

$$\underset{\sim}{S}_{yy} = N^{-1}\sum_{i=1}^{N}(\underset{\sim}{y}_i-\bar{\underset{\sim}{y}}_U)(\underset{\sim}{y}_i-\bar{\underset{\sim}{y}}_U)^T.$$

This can alternatively be written as,

$$S_{yy} = \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}\ y_i y_j^T \quad ,$$

where
$$a_{ij} = \begin{cases} -1/N^2 \text{ if } i\neq j, \\ 1/N(1-1/N) \text{ if } i=j. \end{cases}$$

Example 2.1

Let p=1, then the variance of $\tilde{y}$ is given by,

$$S_{yy}=N^{-1}\sum_U(y_i-\bar{y}_U)^2$$
$$=N^{-1}\sum_U(y_i -(\sum_U(y_i)/N))^2$$
$$= N^{-1}\sum_{i\in U} y_i^2 -N^{-2}(\sum_{i\in U} y_i^2 + \sum_{\substack{i\in U\\ i\neq j}}\sum_{j\in U} y_i y_j)$$
$$=N^{-1}(1-N^{-1})[\sum_{i\in U} y_i^2]-N^{-2}[\sum_{\substack{i\in U\\ i\neq j}}\sum_{j\in U} y_i y_j].$$

Thus

$$S_{yy} = \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}\ y_i y_j \ ,$$

where
$$a_{ij} = \begin{cases} -1/N^2 \text{ if } i\neq j, \\ 1/N(1-1/N) \text{ if } i=j. \end{cases}$$

Alternatively partitioning into the sets s and $\bar{s}$, we may write $\underset{\sim}{S}_{yy}$ as;

$$S_{yy} = \Big(\sum_{i\in s} a_{ii}\mathbf{y}_i\mathbf{y}_i^T + \sum_{i\in\bar{s}} a_{ii}\mathbf{y}_i\mathbf{y}_i^T\Big) + \Big(\sum_{\substack{i\in s\\ i\neq j}}\sum_{j\in s} a_{ij}\mathbf{y}_i\mathbf{y}_j^T + \sum_{\substack{i\in\bar{s}\\ i\neq j}}\sum_{j\in\bar{s}} a_{ij}\mathbf{y}_i\mathbf{y}_j^T + \sum_{i\in\bar{s}}\sum_{j\in s} a_{ij}\mathbf{y}_i\mathbf{y}_j^T + \sum_{i\in s}\sum_{j\in\bar{s}} a_{ij}\mathbf{y}_i\mathbf{y}_j^T\Big).$$

To predict S_{yy} we will consider two approaches;

(1) In the first approach we will predict the quadratic terms $\mathbf{y}_i\mathbf{y}_j^T$ in S_{yy} by $\hat{\mathbf{y}}_i\hat{\mathbf{y}}_j^T$, that is we predict each vector separately using the predictor defined below. This is a bilinear predictor for predicting quadratic terms which is a naive way of predicting S_{yy}, hence we call the predictor obtained the naive predictor and denote it by $\hat{S}_N$. We find,

$$\hat{S}_N = \Big(\sum_{i\in s} a_{ii}\hat{\mathbf{y}}_i\hat{\mathbf{y}}_i^T + \sum_{i\in\bar{s}} a_{ii}\hat{\mathbf{y}}_i\hat{\mathbf{y}}_i^T\Big) + \Big(\sum_{\substack{i\in s\\ i\neq j}}\sum_{j\in s} a_{ij}\hat{\mathbf{y}}_i\hat{\mathbf{y}}_j^T + \sum_{\substack{i\in\bar{s}\\ i\neq j}}\sum_{j\in\bar{s}} a_{ij}\hat{\mathbf{y}}_i\hat{\mathbf{y}}_j^T + \sum_{i\in\bar{s}}\sum_{j\in s} a_{ij}\hat{\mathbf{y}}_i\hat{\mathbf{y}}_j^T + \sum_{i\in s}\sum_{j\in\bar{s}} a_{ij}\hat{\mathbf{y}}_i\hat{\mathbf{y}}_j^T\Big) \qquad [2.1]$$

where

$$\hat{\mathbf{y}}_i = \begin{cases} E_1(\tilde{\mathbf{y}}_i \mid \mathbf{z}, s)\big|_{\lambda=\hat{\lambda}} & \text{for } i\in\bar{s},\\ \mathbf{y}_i & \text{for } i\in s.\end{cases}$$

(2) The second approach will be to predict the quadratic terms $\mathbf{y}_i\mathbf{y}_i^T$ using the quadratic predictor defined below. We denote this quadratic predictor by $\hat{S}_q$ and write,

$$\hat{S}_q = \Big(\sum_{i\in s} a_{ii}\widehat{\mathbf{y}_i\mathbf{y}_i^T} + \sum_{i\in\bar{s}} a_{ii}\widehat{\mathbf{y}_i\mathbf{y}_i^T}\Big) + \Big(\sum_{\substack{i\in s\\ i\neq j}}\sum_{j\in s} a_{ij}\widehat{\mathbf{y}_i\mathbf{y}_j^T} + \sum_{\substack{i\in\bar{s}\\ i\neq j}}\sum_{j\in\bar{s}} a_{ij}\widehat{\mathbf{y}_i\mathbf{y}_j^T} + \sum_{i\in\bar{s}}\sum_{j\in s} a_{ij}\widehat{\mathbf{y}_i\mathbf{y}_j^T} + \sum_{i\in s}\sum_{j\in\bar{s}} a_{ij}\widehat{\mathbf{y}_i\mathbf{y}_j^T}\Big), \qquad [2.2]$$

where $\hat{\underset{\sim}{y}_i\underset{\sim}{y}_j^T}$ is a quadratic predictor of $\underset{\sim}{y}_i\underset{\sim}{y}_j^T$.

We use the classical prediction approach to predict the values of the non sample units.The predictor of $\underset{\sim}{y}_i\underset{\sim}{y}_j^T$ is given by

$$\hat{\underset{\sim}{y}_i\underset{\sim}{y}_j^T} = \begin{cases} E(\tilde{\underset{\sim}{y}}_i\tilde{\underset{\sim}{y}}_i^T|\underset{\sim}{z},s;\underset{\sim}{\lambda})|_{\underset{\sim}{\lambda}=\hat{\underset{\sim}{\lambda}}} & \text{for } i=j, i\in\bar{s}, \\ \underset{\sim}{y}_i\underset{\sim}{y}_i^T & \text{for } i=j, i\in s, \\ \underset{\sim}{y}_i E(\tilde{\underset{\sim}{y}}_j^T|\underset{\sim}{z},s\ ;\underset{\sim}{\lambda})|_{\underset{\sim}{\lambda}=\hat{\underset{\sim}{\lambda}}} & \text{for } i\neq j, i\in s, j\in\bar{s}, \\ E(\tilde{\underset{\sim}{y}}_i|\underset{\sim}{z},s\ ;\underset{\sim}{\lambda})|_{\underset{\sim}{\lambda}=\hat{\underset{\sim}{\lambda}}}\ \underset{\sim}{y}_j^T & \text{for } i\neq j, i\in\bar{s}, j\in s, \\ E(\tilde{\underset{\sim}{y}}_i\tilde{\underset{\sim}{y}}_j^T|\underset{\sim}{z}\ ,s;\underset{\sim}{\lambda})|_{\underset{\sim}{\lambda}=\hat{\underset{\sim}{\lambda}}} & \text{for } i\neq j, i\in\bar{s}, j\in\bar{s}, \\ \underset{\sim}{y}_i\underset{\sim}{y}_j^T & \text{for } i\neq j, i\in s, j\in s\ . \end{cases} \qquad [2.3]$$

Example 2.2

Assume the following model

$$y_i=\mu+\varepsilon_i$$

Using the naive approach,the bilinear predictor for quadratic term $\underset{\sim}{y}_i\underset{\sim}{y}_j^T$ is given by,

$$\hat{y_i y_j} = \hat{y}_i\hat{y}_j = \begin{cases} \bar{y}_s^2 & ,i\in\bar{s}\ ,i=j, \\ \bar{y}_s^2\ , & i,j\in\bar{s}, i\neq j, \\ y_i^2\ , & i\in s, i=j, \\ y_i\bar{y}_s, & i\in s, j\in\bar{s}\ , i\neq j, \\ y_i y_j, & i,j\in s, i\neq j, \\ y_j\bar{y}_s, & i\in\bar{s}, j\in s\ , i\neq j. \end{cases} \quad \text{under model } \varepsilon.$$

Now the quadratic predictor is given by

$$\hat{y_i y_j} = E_\varepsilon(\tilde{y}_i\tilde{y}_j|z\ ,s;\lambda)|_{\lambda=\hat{\lambda}} = cov_\varepsilon(\tilde{y}_i,\tilde{y}_j|z\ ,s;\lambda)|_{\lambda=\hat{\lambda}} +$$

$$E_\varepsilon(\tilde{y}_i|z\ ,s;\lambda)|_{\lambda=\hat{\lambda}} E_\varepsilon(\tilde{y}_j|z\ ,s;\lambda)|_{\lambda=\hat{\lambda}}$$

The first approach assumes that all the units not in the sample are uncorrelated.The second approach assumes that different units not in the sample are uncorrelated,but similar units are correlated with a constant variance,i.e

Under model ε,

$$cov_\varepsilon(\tilde{y}_i,\tilde{y}_j|z\ ,s;\lambda)|_{\lambda=\hat{\lambda}} = \begin{cases} 0 & i,j\varepsilon\bar{s},\ i\neq j, \\ S_{yys} & i,j\varepsilon\bar{s},\ i=j. \end{cases}$$

Thus the quadratic predictor under model ε is given by,

$$\hat{y_i y_j} = \begin{cases} \bar{y}_s^2+S_{yys}, & i\in\bar{s}\ ,i=j, \\ \bar{y}_s^2\ , & i,j\varepsilon\bar{s},i\neq j, \\ y_i^2\ , & i\in s,i=j, \\ y_i\bar{y}_s, & i\in s,j\in\bar{s}\ ,i\neq j, \\ y_iy_j, & i,j\in s,i\neq j, \\ y_j\bar{y}_s, & i\in\bar{s},j\in s\ ,i\neq j. \end{cases} \quad \text{under model } \varepsilon.$$

We see from this example that the main difference between the two approaches is that the quadratic prediction approach assumes that similar units not selected into the sample have a constant variance,while the naive prediction approach assumes zero variance for these nonsample units.

THEOREM 2.1

Under model 1 the naive predictor of the covariance matrix $\underset{\sim}{S}_{yy}$,is given by

$$\hat{\underset{\sim}{S}}_N=n/N\underset{\sim}{S}_{yys}+\hat{\underset{\sim}{H}}[\underset{\sim}{S}_{zz}-n/N)\underset{\sim}{S}_{zzs}]\hat{\underset{\sim}{H}}^T$$

where S_{yys}, S_{zz}, S_{zzs} and $\hat{\underset{\sim}{H}}$ are as defined in theorem 1.1

PROOF

From[2.1] the naive predictor of the covariance matrix is given by

$$\hat{\underset{\sim}{S}}_N = \left(\sum_{i\in s} a_{ii} \hat{\underset{\sim}{y}}_i \hat{\underset{\sim}{y}}_i^T + \sum_{i\in \bar{s}} a_{ii} \hat{\underset{\sim}{y}}_i \hat{\underset{\sim}{y}}_i^T \right) +$$

$$\left(\sum_{\substack{i\in s\\ i\neq j}} \sum_{j\in s} a_{ij} \hat{\underset{\sim}{y}}_i \hat{\underset{\sim}{y}}_j^T + \sum_{\substack{i\in \bar{s}\\ i\neq j}} \sum_{j\in \bar{s}} a_{ij} \hat{\underset{\sim}{y}}_i \hat{\underset{\sim}{y}}_j^T \right.$$

$$\left. + \sum_{i\in \bar{s}} \sum_{j\in s} a_{ij} \hat{\underset{\sim}{y}}_i \hat{\underset{\sim}{y}}_j^T + \sum_{i\in s} \sum_{j\in \bar{s}} a_{ij} \hat{\underset{\sim}{y}}_i \hat{\underset{\sim}{y}}_j^T \right) . \qquad [2.4]$$

Under model 1 the predictor of $\underset{\sim}{y}_i$ is given by

$$\hat{\underset{\sim}{y}}_i = E_1(\tilde{\underset{\sim}{y}}_i | \underset{\sim}{z}, s)\Big|_{\lambda=\hat{\lambda}} = \begin{cases} \hat{\underset{\sim}{m}} + \hat{\underset{\sim}{H}}\underset{\sim}{z}_i, & i\in \bar{s}, \\ y_i, & i\in s . \end{cases} \qquad [2.5]$$

Substituting [2.5] and value of a_{ij} in [2.4] we get

$$\hat{S}_N = 1/N\left(\sum_{i\in s} \underset{\sim}{y}_i \underset{\sim}{y}_i^T + \sum_{i\in \bar{s}} ((\hat{\underset{\sim}{m}} + \hat{\underset{\sim}{H}}\underset{\sim}{z}_i)(\hat{\underset{\sim}{m}} + \hat{\underset{\sim}{H}}\underset{\sim}{z}_i)^T) \right) -$$

$$1/N^2\left(\sum_{i\in s} \underset{\sim}{y}_i \underset{\sim}{y}_i^T + \sum_{i\in \bar{s}} ((\hat{\underset{\sim}{m}} + \hat{\underset{\sim}{H}}\underset{\sim}{z}_i)(\hat{\underset{\sim}{m}} + \hat{\underset{\sim}{H}}\underset{\sim}{z}_i)^T) + \sum_{\substack{i\in s\\ i\neq j}} \sum_{j\in s} \underset{\sim}{y}_i \underset{\sim}{y}_j^T \right.$$

$$+ \sum_{\substack{i\in \bar{s}\\ i\neq j}} \sum_{j\in \bar{s}} (\hat{\underset{\sim}{m}} + \hat{\underset{\sim}{H}}\underset{\sim}{z}_i)(\hat{\underset{\sim}{m}} + \hat{\underset{\sim}{H}}\underset{\sim}{z}_j)^T + \sum_{i\in \bar{s}} \sum_{j\in s} (\hat{\underset{\sim}{m}} + \hat{\underset{\sim}{H}}\underset{\sim}{z}_i)\underset{\sim}{y}_j^T$$

$$\left. + \sum_{i\in s} \sum_{j\in \bar{s}} \underset{\sim}{y}_i (\hat{\underset{\sim}{m}} + \hat{\underset{\sim}{H}}\underset{\sim}{z}_j)^T \right) \qquad [2.6]$$

We will simplify each term of [2.6] individually. Now consider the first term of [2.6]

$$\sum_{i\in s} \underset{\sim}{y}_i \underset{\sim}{y}_i^T = n\underset{\sim}{S}_{yys} + n\bar{\underset{\sim}{y}}_s \bar{\underset{\sim}{y}}_s^T \qquad [2.7]$$

second term equals to

$$\sum_{i\in\bar{s}} ((\hat{\underset{\sim}{m}}+\hat{\underset{\sim}{H}}\underset{\sim}{z}_i)(\hat{\underset{\sim}{m}}+\hat{\underset{\sim}{H}}\underset{\sim}{z}_i)^T) = (N-n)\hat{\underset{\sim}{m}}\hat{\underset{\sim}{m}}^T + \hat{\underset{\sim}{m}}(N\bar{\underset{\sim}{z}}_U^T - n\bar{\underset{\sim}{z}}_s^T)\hat{\underset{\sim}{H}}^T + \hat{\underset{\sim}{H}}(N\bar{\underset{\sim}{z}}_U - n\bar{\underset{\sim}{z}}_s)\hat{\underset{\sim}{m}}^T$$
$$+\hat{\underset{\sim}{H}}[N\underset{\sim}{S}_{zz} - n\underset{\sim}{S}_{zzs} - N\bar{\underset{\sim}{z}}_U\bar{\underset{\sim}{z}}_U^T - n\bar{\underset{\sim}{z}}_s\bar{\underset{\sim}{z}}_s^T]\hat{\underset{\sim}{H}}^T \qquad [2.8]$$

fifth term reduces to

$$\sum_{i\in s}\sum_{\substack{j\in s\\ i\neq j}} \underset{\sim}{y}_i \underset{\sim}{y}_j^T = n^2\bar{\underset{\sim}{y}}_s\bar{\underset{\sim}{y}}_s^T - n\underset{\sim}{S}_{yys} + n\bar{\underset{\sim}{y}}_s\bar{\underset{\sim}{y}}_s^T \qquad [2.9]$$

sixth term reduces to

$$\sum_{i\in\bar{s}}\sum_{\substack{j\in\bar{s}\\ i\neq j}} (\hat{\underset{\sim}{m}}+\hat{\underset{\sim}{H}}\underset{\sim}{z}_i)(\hat{\underset{\sim}{m}}+\hat{\underset{\sim}{H}}\underset{\sim}{z}_j)^T = (N-n)^2\hat{\underset{\sim}{m}}\hat{\underset{\sim}{m}}^T + \hat{\underset{\sim}{m}}(N-n)(N\bar{\underset{\sim}{z}}_U - n\bar{\underset{\sim}{z}}_s)\hat{\underset{\sim}{H}}^T +$$
$$\hat{\underset{\sim}{H}}(N-n)(N\bar{\underset{\sim}{z}}_U - n\bar{\underset{\sim}{z}}_s)\hat{\underset{\sim}{m}}^T + \hat{\underset{\sim}{H}}(N\bar{\underset{\sim}{z}}_U - n\bar{\underset{\sim}{z}}_s)(N\bar{\underset{\sim}{z}}_U - n\bar{\underset{\sim}{z}}_s)^T\hat{\underset{\sim}{H}}^T$$
$$-[(N-n)\hat{\underset{\sim}{m}}\hat{\underset{\sim}{m}}^T + \hat{\underset{\sim}{m}}(N\bar{\underset{\sim}{z}}_U^T - n\bar{\underset{\sim}{z}}_s^T)\hat{\underset{\sim}{H}}^T + \hat{\underset{\sim}{H}}(N\bar{\underset{\sim}{z}}_U - n\bar{\underset{\sim}{z}}_s)\hat{\underset{\sim}{m}}^T$$
$$+\hat{\underset{\sim}{H}}[N\underset{\sim}{S}_{zz} - n\underset{\sim}{S}_{zzs} - N\bar{\underset{\sim}{z}}_U\bar{\underset{\sim}{z}}_U^T - n\bar{\underset{\sim}{z}}_s\bar{\underset{\sim}{z}}_s^T]\hat{\underset{\sim}{H}}^T]$$

$$[2.10]$$

seventh term reduces to

$$\sum_{i\in s}\sum_{j\in\bar{s}} \underset{\sim}{y}_i(\hat{\underset{\sim}{m}}+\hat{\underset{\sim}{H}}\underset{\sim}{z}_j)^T = n(N-n)\hat{\underset{\sim}{m}}\hat{\underset{\sim}{m}}^T + nN\hat{\underset{\sim}{m}}\bar{\underset{\sim}{z}}_U^T\hat{\underset{\sim}{H}}^T - n^2\hat{\underset{\sim}{m}}\bar{\underset{\sim}{z}}_s^T\hat{\underset{\sim}{H}}^T + n(N-n)\hat{\underset{\sim}{H}}\bar{\underset{\sim}{z}}_s\hat{\underset{\sim}{m}}^T$$

$$+nN\hat{\underset{\sim}{H}}\bar{\underset{\sim}{z}}_s\bar{\underset{\sim}{z}}_U^T\hat{\underset{\sim}{H}}^T - n^2\hat{\underset{\sim}{H}}\bar{\underset{\sim}{z}}_s\bar{\underset{\sim}{z}}_s^T\hat{\underset{\sim}{H}}^T \qquad [2.11]$$

and lastly

$$\sum_{i\in\bar{s}}\sum_{j\in s}(\hat{m}+\hat{H}z_i)y_j^T = n(N-n)\hat{m}\hat{m}^T + nN\hat{H}\bar{z}_U\hat{m}^T - n^2\hat{H}\bar{z}_s\hat{m}^T + n(N-n)\hat{m}\bar{z}_s^T\hat{H}^T$$

$$+nN\hat{H}\bar{z}_U\bar{z}_s^T\hat{H}^T - n^2\hat{H}\bar{z}_s\bar{z}_s^T\hat{H}^T \qquad [2.12]$$

Substituting eqns [2.7]-[2.12] in [2.6] we get the required result.

THEOREM 2.2

The conditional expectation of the *naive predictor* of the finite population covariance matrix under model1 derived in theorem 2.1 is given by

$$E_1(\hat{S}_N|s,z)=\alpha(S_{yy}-HS_{zz}H^T)+HS_{zz}H^T,$$

where $\alpha=((n-q-1)/N - Tr(S_{zz}S_{zzs}^{-1})/n)$.

Proof

From Theorem [2.1] the naive predictor is

$$\hat{S}_N = n/N(S_{yys}-\hat{H}S_{zzs}\hat{H}^T)+\hat{H}S_{zz}\hat{H}^T$$
$$= n/N\hat{K}+\hat{H}S_{zz}\hat{H}^T \qquad [2.13]$$

where $\hat{K}=S_{yys}-\hat{H}S_{zzs}\hat{H}^T$ is an estimator of K.

Taking conditional expectation of eq[2.13] w.r.t the model 1 given s and z we get

$$E_1(\hat{S}_N|s,z)=n/NE_1(\hat{K}|s,z)+E_1(\hat{H}S_{zz}\hat{H}^T|s,z)$$

[2.14]

From Skinner(1983) we have

$$E_1(\hat{\underset{\sim}{K}}|s,\underset{\sim}{z}) = (n-q-1)\underset{\sim}{K}/n, \qquad [2.15]$$

where $\underset{\sim}{K}=\underset{\sim}{S}_{yy}-\underset{\sim}{H}\underset{\sim}{S}_{zz}\underset{\sim}{H}^T$,

$$E_1(\hat{\bar{\underset{\sim}{y}}}_U|s,\underset{\sim}{z})=\bar{\underset{\sim}{y}}_U ,$$

and

$$E_1(\hat{\underset{\sim}{H}}\underset{\sim}{S}_{zz}\hat{\underset{\sim}{H}}^T|s,\underset{\sim}{z})=\underset{\sim}{H}\underset{\sim}{S}_{zz}\underset{\sim}{H}^T+Tr(\underset{\sim}{S}_{zz}\underset{\sim}{S}_{zzs}^{-1})\underset{\sim}{K}/n. \qquad [2.16]$$

Substituting eq [2.15] and [2.16] in [2.14] we get

$$E_1(\hat{\underset{\sim}{S}}_N|s,\underset{\sim}{z})=((n-q-1)/N -Tr(\underset{\sim}{S}_{zz}\underset{\sim}{S}_{zzs}^{-1})/n)\underset{\sim}{K}+\underset{\sim}{H}\underset{\sim}{S}_{zz}\underset{\sim}{H}^T.$$

Hence the result.

COROLLARY 2.1

The naive predictor $\hat{\underset{\sim}{S}}_N$ is not a consistent predictor of $\underset{\sim}{S}_{yy}$. That is,

$$\lim_{\substack{n\Rightarrow\infty \\ N\Rightarrow\infty}} E(\hat{\underset{\sim}{S}}_N)\neq\underset{\sim}{S}_{yy}.$$

Proof.

From theorem 2.2 we have

$$E_1(\hat{\underset{\sim}{S}}_N|s,\underset{\sim}{z})=\alpha(\underset{\sim}{S}_{yy}-\underset{\sim}{H}\underset{\sim}{S}_{zz}\underset{\sim}{H}^T)+\underset{\sim}{H}\underset{\sim}{S}_{zz}\underset{\sim}{H}^T,$$

where $\alpha=((n-q-1)/N -Tr(\underset{\sim}{S}_{zz}\underset{\sim}{S}_{zzs}^{-1})/n)$.

Averaging over all possible samples we get

$$E(\hat{\underset{\sim}{S}}_N)=((n-q-1)/N-E_p(tr(\underset{\sim}{S}_{zz}\underset{\sim}{S}_{zzs}^{-1})/n)(\underset{\sim}{S}_{yy}-\underset{\sim}{H}\underset{\sim}{S}_{zz}\underset{\sim}{H}^T)$$
$$+\underset{\sim}{H}\underset{\sim}{S}_{zz}\underset{\sim}{H}^T.$$

If we consider the following sequence of consistency i.e $n\Rightarrow\infty$, $N\Rightarrow\infty$ and $n/N\Rightarrow c$(a constant) then

$$\lim_{\substack{n\Rightarrow\infty \\ N\Rightarrow\infty}} E(\hat{\underset{\sim}{S}}_N)=\underset{\sim}{H}\underset{\sim}{S}_{zz}\underset{\sim}{H}^T \neq\underset{\sim}{S}_{yy}. \qquad [2.17]$$

Example 2.3

Assume the same model as in example 2.2.
For a single survey variable the variance of $\tilde{y}$ is given by,

$$S_{yy}=N^{-1}\sum_{U}(y_i-\bar{y}_U)^2.$$

From eqn [2.1] we have,

$$\hat{S}_N=\Big(\sum_{i\in s} a_{ii}\,\hat{y}_i^{\,2}+\sum_{i\in\bar{s}} a_{ii}\hat{y}_i^2\Big)+$$

$$\Big(\sum_{\substack{i\in s\, j\in s\\ i\neq j}}\sum a_{ij}\,\hat{y}_i\hat{y}_j+\sum_{\substack{i\in\bar{s}\, j\in\bar{s}\\ i\neq j}}\sum a_{ij}\hat{y}_i\hat{y}_j$$

$$+\sum_{i\in\bar{s}\, j\in s}\sum a_{ij}\hat{y}_i\hat{y}_j+\sum_{i\in s\, j\in\bar{s}}\sum a_{ij}\hat{y}_i\hat{y}_j\Big)$$

Using the predictor given in [2.1] and substituting a_{ij} we get,

$$\hat{S}_N=N^{-1}[1-N^{-1}][nS_{yys}+n\bar{y}_s^2+(N-n)\bar{y}_s^2]-N^{-2}[n^2\bar{y}_s^2-n(S_{yys}+\bar{y}_s^2)$$
$$+(N-n)^2\bar{y}_s^2-(N-n)\bar{y}_s^2+2n(N-n)\bar{y}_s^2]$$
$$=n/NS_{yys}$$

Taking conditional expectations of $\hat{S}_N$ and averaging over all possible samples,we see that $\hat{S}_N$ is a biased estimator of S_{yy}.

We see from [2.17] that the naive predictor is not a consistent predictor of the finite population covariance matrix under model 1.In order to get a consistent predictor of the finite population covariance matrix some modifications are required.

Since the predictor of the finite population covariance matrix is a quadratic function of the values of the survey variables we will adopt the second approach to predict quadratic functions of $\tilde{y}_i$ under model 1.Mukhopadhyay(1978) also followed an approach similar to the one we will follow but he assumed zero conditional means.

THEOREM 2.3

Under model 1 the quadratic predictor of the covariance matrix is given by

$$\hat{S}_q = S_{yys} + \hat{H}(S_{zz} - S_{zzs})\hat{H}^T,$$

where S_{yys}, S_{zz}, S_{zzs} and $\hat{H}$ are as defined in [1.15].

Proof

From eqn[2.2] the quadratic predictor of the covariance matrix is given by

$$\hat{S}_q = \Big(\sum_{i\in s} a_{ii}\, \widehat{y_i y_i^T} + \sum_{i\in \bar{s}} a_{ii} \widehat{y_i y_i^T}\Big)$$

$$+ \Big(\sum_{\substack{i\in s\, j\in s\\ i\neq j}} a_{ij}\, \widehat{y_i y_j^T} + \sum_{\substack{i\in \bar{s}\, j\in \bar{s}\\ i\neq j}} a_{ij} \widehat{y_i y_j^T}$$

$$+ \sum_{i\in \bar{s}}\sum_{j\in s} a_{ij} \widehat{y_i y_j^T} + \sum_{i\in s}\sum_{j\in \bar{s}} a_{ij} \widehat{y_i y_j^T}\Big) . \qquad [2.18]$$

Under model 1 the predictor of $y_i y_j^T$ given in [2.3] becomes

$$
\widehat{\underset{\sim}{y}_i\underset{\sim}{y}_j^T} = \begin{cases}
(\underset{\sim}{\hat{m}}+\underset{\sim}{\hat{H}}\underset{\sim}{z}_i)(\underset{\sim}{\hat{m}}+\underset{\sim}{\hat{H}}\underset{\sim}{z}_i)^T+\underset{\sim}{\hat{K}}) & \text{for } i=j, i\in\bar{s}, \\
\underset{\sim}{y}_i\underset{\sim}{y}_i^T & \text{for } i=j, i\in s, \\
\underset{\sim}{y}_i(\underset{\sim}{\hat{m}}+\underset{\sim}{\hat{H}}\underset{\sim}{z}_j)^T & \text{for } i\neq j, i\in s, j\in\bar{s}, \\
(\underset{\sim}{\hat{m}}+\underset{\sim}{\hat{H}}\underset{\sim}{z}_i)\ \underset{\sim}{y}_j^T & \text{for } j\neq i, i\in\bar{s}, j\in s, \\
(\underset{\sim}{\hat{m}}+\underset{\sim}{\hat{H}}\underset{\sim}{z}_i)(\underset{\sim}{\hat{m}}+\underset{\sim}{\hat{H}}\underset{\sim}{z}_j)^T & \text{for } i\neq j, i\in\bar{s}, j\in\bar{s}, \\
\underset{\sim}{y}_i\underset{\sim}{y}_j^T & \text{for } i\neq j, i\in s, j\in s.
\end{cases} \qquad [2.19]
$$

using model 1 assumptions.

Substituting [2.19] and the value of a_{ij} in [2.18] we get

$$
N\underset{\sim}{\hat{S}}_q = \Big(\sum_{i\in s}\underset{\sim}{y}_i\underset{\sim}{y}_i^T + \sum_{i\in\bar{s}}((\underset{\sim}{\hat{m}}+\underset{\sim}{\hat{H}}\underset{\sim}{z}_i)(\underset{\sim}{\hat{m}}+\underset{\sim}{\hat{H}}\underset{\sim}{z}_i)^T+\underset{\sim}{\hat{K}})\Big) -
$$

$$
1/N\Big(\sum_{i\in s}\underset{\sim}{y}_i\underset{\sim}{y}_i^T + \sum_{i\in\bar{s}}((\underset{\sim}{\hat{m}}+\underset{\sim}{\hat{H}}\underset{\sim}{z}_i)(\underset{\sim}{\hat{m}}+\underset{\sim}{\hat{H}}\underset{\sim}{z}_i)^T+\underset{\sim}{\hat{K}}) + \sum_{i\in s}\sum_{\substack{j\in s\\ i\neq j}}\underset{\sim}{y}_i\underset{\sim}{y}_j^T
$$

$$
+\sum_{i\in\bar{s}}\sum_{\substack{j\in\bar{s}\\ i\neq j}}(\underset{\sim}{\hat{m}}+\underset{\sim}{\hat{H}}\underset{\sim}{z}_i)(\underset{\sim}{\hat{m}}+\underset{\sim}{\hat{H}}\underset{\sim}{z}_j)^T + \sum_{i\in\bar{s}}\sum_{j\in s}(\underset{\sim}{\hat{m}}+\underset{\sim}{\hat{H}}\underset{\sim}{z}_i)\underset{\sim}{y}_j^T
$$

$$
+\sum_{i\in s}\sum_{j\in\bar{s}}\underset{\sim}{y}_i(\underset{\sim}{\hat{m}}+\underset{\sim}{\hat{H}}\underset{\sim}{z}_j)^T\Big). \qquad [2.20]
$$

We will simplify each term of [2.20] individually.We note that apart from the second term in eq[2.20] and the sixth term all the other terms are identical to those of eq[2.6] we will therefore simplify only these two term.Now the second term simplifies to

$$\sum_{i\in\bar{s}}((\hat{m}+\hat{H}z_i)(\hat{m}+\hat{H}z_i)^T+\hat{K}) = (N-n)\hat{m}\hat{m}^T+\hat{m}(N\bar{z}_U^T-n\bar{z}_s^T)\hat{H}^T+\hat{H}(N\bar{z}_U-n\bar{z}_s)\hat{m}^T$$
$$+\hat{H}[NS_{zz}-nS_{zzs}-N\bar{z}_U^2\bar{z}_U^T-n\bar{z}_s^2\bar{z}_s^T]\hat{H}^T$$
$$+(N-n)\hat{K}.$$

[2.21a].

$$\sum_{i\in\bar{s}}\sum_{\substack{j\in\bar{s}\\ i\neq j}}(\hat{m}+\hat{H}z_i)(\hat{m}+\hat{H}z_j)^T=(N-n)^2\hat{m}\hat{m}^T+\hat{m}(N-n)(N\bar{z}_U-n\bar{z}_s)\hat{H}^T+$$
$$\hat{H}(N-n)(N\bar{z}_U-n\bar{z}_s)\hat{m}^T+\hat{H}(N\bar{z}_U-n\bar{z}_s)(N\bar{z}_U-n\bar{z}_s)^T\hat{H}^T$$
$$-[(N-n)\hat{m}\hat{m}^T+\hat{m}(N\bar{z}_U^T-n\bar{z}_s^T)\hat{H}^T+\hat{H}(N\bar{z}_U-n\bar{z}_s)\hat{m}^T$$
$$+\hat{H}[NS_{zz}-nS_{zzs}-N\bar{z}_U\bar{z}_U^T-n\bar{z}_s\bar{z}_s^T]\hat{H}^T\ (N-n)+\hat{K}]$$

[2.21b]

Substituting eqns [2.7]-[2.11] and eqns [2.21a] and [2.21b] in [2.20] we get the required result.

THEOREM 2.4

The quadratic predictor $\hat{S}_q$ derived in theorem [2.3] is an asymptotically unbiased predictor of the finite pouplation covariance matrix S_{yy} that is,

$$\lim_{\substack{n\Rightarrow\infty\\ N\Rightarrow\infty}} E(\hat{S}_q)=S_{yy}.$$

provided that the design matrix S_{zzs} is nonsingular.

PROOF.

From theorem 2.3 the quadratic predictor of the covariance matrix is given by

$$\hat{S}_q=S_{yys}+\hat{H}(S_{zz}-S_{zzs})\hat{H}^T$$

$$= \hat{\underset{\sim}{K}} + \hat{\underset{\sim}{H}}\underset{\sim}{S}_{zz}\hat{\underset{\sim}{H}}^T$$

where $\hat{\underset{\sim}{K}} = \underset{\sim}{S}_{yys} - \hat{\underset{\sim}{H}}\underset{\sim}{S}_{zzs}\hat{\underset{\sim}{H}}^T$.

Taking conditional expectation under model 1 we get

$$E_1(\hat{\underset{\sim}{S}}_q|s,\underset{\sim}{z}) = [E_1(\hat{\underset{\sim}{K}}|s,\underset{\sim}{z}) + E_1(\hat{\underset{\sim}{H}}\underset{\sim}{S}_{zz}\hat{\underset{\sim}{H}}^T|s,\underset{\sim}{z})]$$

$$= [E_1(\hat{\underset{\sim}{K}}|s,\underset{\sim}{z}) + E_1(\hat{\underset{\sim}{H}}\underset{\sim}{S}_{zz}\hat{\underset{\sim}{H}}^T|s,\underset{\sim}{z})] \quad [2.22]$$

From Skinner(1983) we have

$$E_1(\hat{\underset{\sim}{K}}|s,\underset{\sim}{z}) = (n-q-1)\underset{\sim}{K}/n, \quad [2.23]$$

where $\underset{\sim}{K} = \underset{\sim}{S}_{yy} - \underset{\sim}{H}\underset{\sim}{S}_{zz}\underset{\sim}{H}^T$,

$$E_1(\hat{\underset{\sim}{H}}\underset{\sim}{S}_{zz}\hat{\underset{\sim}{H}}^T|s,\underset{\sim}{z}) = \underset{\sim}{H}\underset{\sim}{S}_{zz}\underset{\sim}{H}^T + Tr(\underset{\sim}{S}_{zz}\underset{\sim}{S}_{zzs}^{-1})\underset{\sim}{K}/n, \quad [2.24]$$

and

$$E_1(\hat{\bar{\underset{\sim}{y}}}_U|s,\underset{\sim}{z}) = \bar{\underset{\sim}{y}}_U. \quad [2.25]$$

Substituting eq [2.23]-[2.25] in [2.22] we get

$$E_1(\hat{\underset{\sim}{S}}_q|s,\underset{\sim}{z}) = [(n-q-1)\underset{\sim}{K}/n + \underset{\sim}{H}\underset{\sim}{S}_{zz}\underset{\sim}{H}^T + Tr(\underset{\sim}{S}_{zz}\underset{\sim}{S}_{zzs}^{-1})\underset{\sim}{K}/n)] \quad [2.26]$$

Averaging [2.26] over all possible samples we get

$$E(\hat{\underset{\sim}{S}}_q) = [(n-q-1)\underset{\sim}{K}/n + \underset{\sim}{H}\underset{\sim}{S}_{zz}\underset{\sim}{H}^T + E_p(tr(\underset{\sim}{S}_{zz}\underset{\sim}{S}_{zzs}^{-1})\underset{\sim}{K}/n)$$

Taking limits as $N \Rightarrow \infty$ $n \Rightarrow \infty$ and $n/N \Rightarrow c$(constant) we get

$$\lim_{\substack{n\Rightarrow\infty \\ N\Rightarrow\infty}} E(\hat{\underset{\sim}{S}}_q) = \underset{\sim}{S}_{yy},$$

hence the result.

Example 2.4

Repeating example 2.3 using the quadratic predictor instead of the bilinear predictor we get,

$$\hat{S}_q = N^{-1}[1-N^{-1}][nS_{yys}+n\bar{y}_s^2 + (N-n)(\bar{y}_s^2+S_{yys})] - N^{-2}[n^2\bar{y}_s^2 - n(S_{yys}+\bar{y}_s^2) + (N-n)^2\bar{y}_s^2 - (N-n)(\bar{y}_s^2+S_{yys}) + 2n(N-n)\bar{y}_s^2]$$

$$= S_{yys}.$$

Taking conditional expectation of $\hat{S}_q$ given s and z, averaging over the all possible samples and then taking limits as n,N tends to infinity we see that the predictor $\hat{S}_q$ is an asymptotically unbiased predictor of S_{yy}.

COROLLARY 2.2

For most designs we can also assume

$$\lim_{n\Rightarrow\infty, N\Rightarrow\infty} Var(\hat{\underset{\sim}{S}}_q)=0,$$

then using theorem 2.4 we can conclude that the quadratic predictor is a consistent estimator of the covariance matrix.

However in some pathological designs as n and N increases $\underset{\sim}{S}_{zzs}$ tend to a singular matrix making $\underset{\sim}{S}_q$ an asymptotically biased predictor of $\underset{\sim}{S}_{yy}$. For example if the selection variables $\underset{\sim}{z}_i$ are chosen such that the $\underset{\sim}{z}_i$ are equal for all i, then $\underset{\sim}{S}_{zzs}$ is a singular matrix. We see from theorem 2.4 that quadratic predictor $\hat{\underset{\sim}{S}}_q$ is identical to the Pearson adjusted estimator of the covariance matrix. Thus this is another way of deriving the Pearson adjusted estimator for the covariance matrix.

In chapter 1 we derived Pearson adjusted estimators through Lawleys' (1943) theorem by assuming a linear homoscedastic model. What type of estimators would we get if we generalize Lawleys' theorem to;

(i) Nonlinear Homoscedastic models and

(ii) Nonlinear Heteroscedastic Models.

2.4 GENERALIZATION OF LAWLEYS' THEOREM TO NONLINEAR POPULATIONS

We will generalize the linear model used by Lawley(1943) to another linear model,call it model 5,with a vector $\underset{\sim}{v}$ as an independent variable which has both the linear and nonlinear components.Thus under model 5 we assume that;

(i)The conditional expectation of $\underset{\sim}{\tilde{y}}$ given $\underset{\sim}{v}$ is a linear function of $\underset{\sim}{v}$

$$E_5(\underset{\sim}{\tilde{y}} \,|\, \underset{\sim}{v})=\underset{\sim}{m} +\underset{\sim}{D}\underset{\sim}{v}, \qquad [2.27]$$

where $\underset{\sim}{v} =(\underset{\sim}{z},g(\underset{\sim}{z}))^T$,

$\underset{\sim}{D}=(\underset{\sim}{H},\underset{\sim}{R})$,

$\underset{\sim}{\tilde{y}}=(\tilde{y}_1...\tilde{y}_p)$,$\underset{\sim}{v}=(v_1...v_q)$,$g(\underset{\sim}{z})$ is a known nonlinear continous function of $\underset{\sim}{z}$,$\underset{\sim}{m}$ is px1 dimensional constant vector and $\underset{\sim}{D}$ is a pxq dimensional matrix of constants.

(ii)The conditional variance covariance matrix of $\underset{\sim}{\tilde{y}}$ given $\underset{\sim}{v}$ is constant and does not depend on the value of $\underset{\sim}{v}$.

$$V_5(\underset{\sim}{\tilde{y}} \,|\, \underset{\sim}{v})=\underset{\sim}{K}, \qquad [2.28]$$

where $\underset{\sim}{K}$ is a p×p constant matrix.

Example 2.5

Let V be a vector of both the linear and quadratic functions of $\underset{\sim}{z}$,then for p=1,q=1,$V=[z,z^2]$,the conditional expectation in model 5 will take the form

$$E_5(\tilde{y} \,|\, v)=m +Hz+ Rz^2,$$

and the conditional variance remains the same.

THEOREM 2.5

Under model 5, the parameters in the unselected population are related to those in the selected population by the following relationships;

$$\underset{\sim}{\mu} = \begin{bmatrix} \underset{\sim}{\mu}_y \\ \underset{\sim}{\mu}_v \end{bmatrix} = \begin{bmatrix} \underset{\sim}{\mu}_{ys} + \underset{\sim}{D}(\underset{\sim}{\mu}_v - \underset{\sim}{\mu}_{vs}) \\ \underset{\sim}{\mu}_v \end{bmatrix},$$

where $\underset{\sim}{\mu}_v = \begin{bmatrix} \underset{\sim}{\mu}_z & \underset{\sim}{\mu}_g \end{bmatrix}$,

$$\underset{\sim}{D} = \begin{bmatrix} \underset{\sim}{H}, \underset{\sim}{R} \end{bmatrix} = \underset{\sim}{\Sigma}_{yvs} \underset{\sim}{\Sigma}_{vvs}^{-1},$$

$$= [\underset{\sim}{\Sigma}_{yzs}, \underset{\sim}{\Sigma}_{ygs}] \begin{bmatrix} \underset{\sim}{\Sigma}_{zzs} & \underset{\sim}{\Sigma}_{gzs} \\ \underset{\sim}{\Sigma}_{zgs} & \underset{\sim}{\Sigma}_{ggs} \end{bmatrix}^{-1}, \qquad [2.29]$$

$$\underset{\sim}{\Sigma}_{vvs} = \begin{bmatrix} \underset{\sim}{\Sigma}_{zzs} & \underset{\sim}{\Sigma}_{gzs} \\ \underset{\sim}{\Sigma}_{zgs} & \Sigma_{ggs} \end{bmatrix},$$

and

$$\underset{\sim}{\Sigma} = \begin{bmatrix} \underset{\sim}{\Sigma}_{yy} & \underset{\sim}{\Sigma}_{yv} \\ \underset{\sim}{\Sigma}_{vy} & \underset{\sim}{\Sigma}_{vv} \end{bmatrix},$$

$$= \begin{bmatrix} \underset{\sim}{\Sigma}_{yys} + \underset{\sim}{D}(\underset{\sim}{\Sigma}_{vv} - \underset{\sim}{\Sigma}_{vvs})\underset{\sim}{D}^T & \underset{\sim}{D}\underset{\sim}{\Sigma}_{vv} \\ \underset{\sim}{\Sigma}_{vv}\underset{\sim}{D}^T & \underset{\sim}{\Sigma}_{vv} \end{bmatrix}. \qquad [2.30]$$

PROOF

Since the formulation of model 5 is exactly similar to that of model 1 the proof of this theorem follows directly from that of Lawleys' theorem in chapter 1.

From theorem 2.5 an estimator of the covariance matrix under the quadratic model which we call Quadratic estimator of the covariance matrix and denoted by $\hat{\Sigma}_{yy,q1}$ is given by;

$$\hat{\Sigma}_{yy,q1}=\Sigma_{yys}+\hat{H}(\Sigma_{zz}-\Sigma_{zzs})\hat{H}^T+\hat{R}(\Sigma_{gg}-\Sigma_{ggs})\hat{R}^T+\hat{H}(\Sigma_{zg}-\Sigma_{zgs})\hat{R}^T$$
$$+\hat{R}(\Sigma_{gz}-\Sigma_{gzs})\hat{H}^T. \qquad [2.31]$$

Using [2.31] we propose the following estimator of the regression coefficient.

$$\hat{B}_{12,q1}= \hat{\Sigma}_{y1y2,q1}\ \hat{\Sigma}^{-1}_{y2y2,q1}. \qquad [2.39]$$

COROLLARY 2.3

Lawleys' theorem is a special case of our generalized theorem when $R=0$ in the matrix D.

In the next section we will derive another estimator of the Covariance matrix when the Linear Homoscedastic assumptions are violated.This estimator is obtained after generalization of Lawleys'theorem to a nonlinear heteroscedastic model.

2.5 GENERALIZATION OF LAWLEY'S THEOREM TO NONLINEAR HETEROSCEDASTIC POPULATIONS

In theorem (1.5) chapter1 we proved Lawley's theorem under the linear homoscedastic model assumptions and in corr(1.1) we shown that the theorem also holds for nonlinear populations. In this section we will generalize the result to populations which are nonlinear and heteroscedastic ,call this model, model 6.

Under model 6 we assume that;

(i)The conditional expectation of $\tilde{\underset{\sim}{y}}$ given $\tilde{\underset{\sim}{v}}=\underset{\sim}{v}$ is a linear function of $\tilde{\underset{\sim}{v}}=\underset{\sim}{v}$, i.e

$$E_6\,(\tilde{\underset{\sim}{y}}\,|\underset{\sim}{v})=\underset{\sim}{m}+\underset{\sim}{D}\underset{\sim}{v}, \qquad [2.33]$$

where $\underset{\sim}{v}=(\underset{\sim}{z},g(\underset{\sim}{z}))^T$,

$\underset{\sim}{D}=(\underset{\sim}{H},\underset{\sim}{R})$,

$\tilde{\underset{\sim}{y}}=(\tilde{y}_1\ldots\tilde{y}_p)$, $\underset{\sim}{v}=(v_1\ldots v_q)$, $g(\underset{\sim}{z})$ is a nonlinear continous known function of $\underset{\sim}{z}$, $\underset{\sim}{m}$ is px1 dimensional constant vector and $\underset{\sim}{D}$ is a pxq dimensional matrix of constants.

(ii)The conditional variance covariance matrix of $\tilde{\underset{\sim}{y}}$ given $\underset{\sim}{v}$ is given by

$$V_6(\tilde{\underset{\sim}{y}}|\underset{\sim}{v})=W(\underset{\sim}{v})=[W_{jk}(\underset{\sim}{v})], \qquad [2.34]$$

where $\underset{\sim}{m}$ is a px1 constant vector, $\underset{\sim}{H}$ is a pxp constant matrix, $\underset{\sim}{v}$ is a qx1 vector of the realized values of the random variables $\tilde{\underset{\sim}{v}}$ and $W(\underset{\sim}{v})$ is a pxp nonsingular matrix.

For example when q=1 and p=2 we have

$$V_3(\tilde{\underset{\sim}{y}}|v)=[W_{jk}(v_i)]=\begin{bmatrix} w_{11}(v_i) & w_{12}(v_i) \\ w_{21}(v_i) & w_{22}(v_i) \end{bmatrix}.$$

As a special case we could have,

$W_{11}(v_i)=A_1+B_1z_i+C_1z_i^2$, $W_{12}(v_i)=A_3+B_3z_i+C_3z_i^2$ and

$W_{22}(v_i)=A_2+B_2z_i+C_2z_i^2$ where $\underset{\sim}{A}=(A_1,A_2,A_3)$, $\underset{\sim}{B}=(B_1,B_2,B_3)$ and

$\underset{\sim}{C}=(C_1,C_2,C_3)$ are constants, then

$$V_3(\tilde{\underset{\sim}{y}}|\underset{\sim}{v})=\begin{bmatrix} A_1+B_1z_i+C_1z_i^2 & A_3+B_3z_i+C_3z_i^2 \\ A_3+B_3z_i+C_3z_i^2 & A_2+B_2z_i+C_2z_i^2 \end{bmatrix}.$$

THEOREM 2.6

Under model 6, the population parameters in the unselected population can be reconstructed from those in the selected population by the following relationships;

$$\underset{\sim}{\mu} = \begin{bmatrix} \underset{\sim}{\mu}_y \\ \underset{\sim}{\mu}_v \end{bmatrix} = \begin{bmatrix} \underset{\sim}{\mu}_{ys} + \underset{\sim}{D} (\underset{\sim}{\mu}_v - \underset{\sim}{\mu}_{vs}) \\ \underset{\sim}{\mu}_v \end{bmatrix},$$

where $\underset{\sim}{\mu}_v = \begin{bmatrix} \underset{\sim}{\mu}_z & \underset{\sim}{\mu}_g \end{bmatrix}$,

$$\underset{\sim}{D} = \begin{bmatrix} \underset{\sim}{H}, \underset{\sim}{R} \end{bmatrix} = \underset{\sim}{\Sigma}_{yvs} \underset{\sim}{\Sigma}_{vvs}^{-1},$$

$$= [\underset{\sim}{\Sigma}_{yzs}, \underset{\sim}{\Sigma}_{ygs}] \begin{bmatrix} \underset{\sim}{\Sigma}_{zzs} & \underset{\sim}{\Sigma}_{gzs} \\ \underset{\sim}{\Sigma}_{zgs} & \underset{\sim}{\Sigma}_{ggs} \end{bmatrix}^{-1},$$

$$\underset{\sim}{\Sigma}_{vvs} = \begin{bmatrix} \underset{\sim}{\Sigma}_{zzs} & \underset{\sim}{\Sigma}_{gzs} \\ \underset{\sim}{\Sigma}_{zgs} & \underset{\sim}{\Sigma}_{ggs} \end{bmatrix},$$

and

$$\underset{\sim}{\Sigma} = \begin{bmatrix} \underset{\sim}{\Sigma}_{yy} & \underset{\sim}{\Sigma}_{yv} \\ \underset{\sim}{\Sigma}_{vy} & \underset{\sim}{\Sigma}_{vv} \end{bmatrix},$$

$$= \begin{bmatrix} \underset{\sim}{\Sigma}_{yys} + \underset{\sim}{D}(\underset{\sim}{\Sigma}_{vv} - \underset{\sim}{\Sigma}_{vvs})\underset{\sim}{D}^T + (\underset{\sim}{\mu}_f - \underset{\sim}{\mu}_{fs}) & \underset{\sim}{D}\,\underset{\sim}{\Sigma}_{vv} \\ \underset{\sim}{\Sigma}_{vv}\underset{\sim}{D}^T & \underset{\sim}{\Sigma}_{vv} \end{bmatrix},$$

where

$$\underset{\sim}{\mu}_f = E_v(\underset{\sim}{w}(v)) \text{ and } \underset{\sim}{\mu}_{fs} = E_v(\underset{\sim}{w}(v)|s).$$

PROOF.

Taking expectations and variances of the expression [2.33] and [2.34] over the model distribution of $\underset{\sim}{\tilde{v}}$ we get

$$E_v\, E_6\, (\underset{\sim}{\tilde{y}}\, |\underset{\sim}{v}) = E_v(\underset{\sim}{m}) + \underset{\sim}{D}\, E_v(\underset{\sim}{v})$$

using [2.33] and [2.34]

$$= \underset{\sim}{m} + \underset{\sim}{D}\, E_v(\underset{\sim}{v}). \qquad [2.35]$$

Also let $\underset{\sim}{\mu}_y = E_v(E_6(\underset{\sim}{\tilde{y}}|\underset{\sim}{v}))$,

$\underset{\sim}{\mu}_v = E_v(\underset{\sim}{v})$,

Thus [2.35] becomes

$$\underset{\sim}{\mu}_y = \underset{\sim}{m} + \underset{\sim}{D}\, \underset{\sim}{\mu}_v. \qquad [2.36]$$

Now

$$\underset{\sim}{\Sigma}_{yy} = V(\underset{\sim}{\tilde{y}})$$

$$= V_v(E_6(\underset{\sim}{\tilde{y}}|\underset{\sim}{v})) + E_v(V_6(\underset{\sim}{\tilde{y}}|\underset{\sim}{v}))$$

$$= V_v(\underset{\sim}{m} + \underset{\sim}{D}\underset{\sim}{\mu}_v) + E_v(\underset{\sim}{w}(v)),$$

using [2.33] and [2.34]

$$=\underset{\sim}{D}V_v(\underset{\sim}{v})\underset{\sim}{D}^T + \underset{\sim}{\mu}_f \ , \qquad [2.37]$$

Let

$$V_v(\underset{\sim}{v}) = \underset{\sim}{\Sigma}_{vv},$$

Thus [2.37] becomes

$$\underset{\sim}{\Sigma}_{yy} = \underset{\sim}{D}\ \underset{\sim}{\Sigma}_{vv}\underset{\sim}{D}^T + \underset{\sim}{\mu}_f. \qquad [2.38]$$

Also let

$$\underset{\sim}{\Sigma}_{yv} = \mathrm{COV}(\underset{\sim}{\tilde{y}},\underset{\sim}{\tilde{v}})$$

$$=E(\underset{\sim}{\tilde{y}}\underset{\sim}{\tilde{v}}-E(\underset{\sim}{\tilde{y}})E(\underset{\sim}{\tilde{v}}))$$

$$=E_vE_6\ (\underset{\sim}{\tilde{y}}\underset{\sim}{\tilde{v}}-E_6(\underset{\sim}{\tilde{y}}|\underset{\sim}{v})E_6(\underset{\sim}{\tilde{v}}|\underset{\sim}{v})+E_6(\underset{\sim}{\tilde{y}}|\underset{\sim}{v})E_6(\underset{\sim}{\tilde{v}}|\underset{\sim}{v})-E(\underset{\sim}{\tilde{y}})E(\underset{\sim}{\tilde{v}})|\underset{\sim}{v})$$

$$=E_v(\mathrm{cov}_6(\underset{\sim}{\tilde{y}},\underset{\sim}{\tilde{v}}|\underset{\sim}{v}))+\mathrm{cov}_v(E_6(\underset{\sim}{\tilde{y}}|\underset{\sim}{v}),E_6(\underset{\sim}{\tilde{v}}|\underset{\sim}{v})).$$

Now since $E_6(\underset{\sim}{\tilde{v}}|\underset{\sim}{v})=\underset{\sim}{v}$ and $\mathrm{cov}_6(\underset{\sim}{\tilde{y}},\underset{\sim}{\tilde{v}}|\underset{\sim}{v})=0$,

then

$$\underset{\sim}{\Sigma}_{yv}=\mathrm{COV}_v(E_6(\underset{\sim}{y}|\underset{\sim}{v}),\underset{\sim}{v})$$

$$=\mathrm{COV}_v(\underset{\sim}{m}+\underset{\sim}{D}\underset{\sim}{v}\ ,\underset{\sim}{v}),$$ using [2.33] and [2.34]

$$= \underset{\sim}{D}\ V_v(\underset{\sim}{v})$$

$$= \underset{\sim}{D}\ \underset{\sim}{\Sigma}_{vv}.$$

If $\underset{\sim}{\Sigma}_{vv}$ is nonsingular then

$$\underset{\sim}{D}=\underset{\sim}{\Sigma}_{yv}\ \underset{\sim}{\Sigma}_{vv}^{-1}. \qquad [2.39]$$

From eq(1.16) in chapter1 we defined the joint distribution for the whole population as;

$$f(\underset{\sim}{y},\underset{\sim}{v};\underset{\sim}{\lambda},\underset{\sim}{\phi})=f(\underset{\sim}{v},\underset{\sim}{\phi})\ f(\underset{\sim}{y}|\underset{\sim}{v};\underset{\sim}{\lambda}). \qquad [2.40]$$

After selection this becomes

$$f(\underset{\sim}{y},\underset{\sim}{v}|s;\underset{\sim}{\lambda},\underset{\sim}{\phi})=f(\underset{\sim}{v}|s;\underset{\sim}{\phi})\ f(\underset{\sim}{y}|\underset{\sim}{v};\underset{\sim}{\lambda}). \qquad [2.41]$$

We therefore see that the conditional distribution of $\underset{\sim}{\tilde{y}}$ given $\tilde{\underset{\sim}{v}}=\underset{\sim}{v}$ is unaffected by selection on z hence

$$E_6(\underset{\sim}{\tilde{y}}|\underset{\sim}{v},s)= \underset{\sim}{m} +\underset{\sim}{D}\ \underset{\sim}{v}. \qquad [2.42]$$

Similarly

$$V_6(\underset{\sim}{\tilde{y}}\ |\underset{\sim}{v},s)=W(\underset{\sim}{v}). \qquad [2.43]$$

Taking expectations and variances of the expression [2.42] and [2.43] over the model distribution of $\tilde{\underset{\sim}{v}}$ given the sample s we get

$$E_vE_6(\underset{\sim}{\tilde{y}}\ |\underset{\sim}{v},s)= E_v(\underset{\sim}{m}|s)+\underset{\sim}{D}\ E_v(\underset{\sim}{v})|s)$$

$$=\underset{\sim}{m}\ +\underset{\sim}{D}E_v(\underset{\sim}{v}|s), \qquad [2.44]$$

Also let $\underset{\sim}{\mu}_{ys}=E_v(E_6(\underset{\sim}{\tilde{y}}|\underset{\sim}{v},s)),$

$$\underset{\sim}{\mu}_{vs}=E_v(\underset{\sim}{v}|s),$$

Thus [2.44] becomes

$$\underset{\sim}{\mu}_{ys}\ =\underset{\sim}{m}\ +\underset{\sim}{D}\ \underset{\sim}{\mu}_{vs}. \qquad [2.45]$$

Now

$$\underset{\sim}{\Sigma}_{yys} = V(\underset{\sim}{\tilde{y}}|s)$$

$$=V_v((E_6(\underset{\sim}{\tilde{y}}|\underset{\sim}{v},s)|s)+E_v(V_6(\underset{\sim}{\tilde{y}}|\underset{\sim}{v},s)|s)$$

$$=V_v((\underset{\sim}{m}+\underset{\sim}{D}\ \underset{\sim}{v})|s)+E_v(w(\underset{\sim}{v})|s)$$

using [2.42] and [2.43]

$$=\underset{\sim}{D}V_v(\underset{\sim}{v}|s)\underset{\sim}{D}^T\ +E_v(w(\underset{\sim}{v})|s), \qquad [2.46]$$

let

$$\underset{\sim}{\mu}_{fs}=E_v(w(\underset{\sim}{v})|s) \text{ and } \underset{\sim}{\sum}_{vvs}=V_v(\underset{\sim}{v}|s),$$

Thus [2.46] becomes

$$\underset{\sim}{\sum}_{yys}=\underset{\sim}{D}\underset{\sim}{\sum}_{vvs}\underset{\sim}{D}^T+\underset{\sim}{\mu}_{fs}. \qquad [2.47]$$

lastly

$$\underset{\sim}{\sum}_{yv}=\text{COV}((\underset{\sim}{\tilde{y}},\underset{\sim}{\tilde{v}})|s)$$

$$=E((\underset{\sim}{\tilde{y}}\underset{\sim}{\tilde{v}}-E(\underset{\sim}{\tilde{y}})E(\underset{\sim}{\tilde{v}}))|s)$$

$$=E_vE_6((\underset{\sim}{\tilde{y}}\underset{\sim}{\tilde{v}}-E_6(\underset{\sim}{\tilde{y}}|\underset{\sim}{v},s)E_6(\underset{\sim}{\tilde{v}}|\underset{\sim}{v},s)+E_6(\underset{\sim}{\tilde{y}}|\underset{\sim}{v},s)E_6(\underset{\sim}{\tilde{v}}|\underset{\sim}{v},s)$$

$$-E(\underset{\sim}{\tilde{y}})E(\underset{\sim}{\tilde{v}}))|\underset{\sim}{v},s)|s)$$

$$=E_v((\text{cov}_6(\underset{\sim}{\tilde{y}},\underset{\sim}{\tilde{v}}|\underset{\sim}{v},s)|s)+\text{cov}_v(E_6(\underset{\sim}{\tilde{y}}|\underset{\sim}{v},s),E_6(\underset{\sim}{\tilde{v}}|\underset{\sim}{v},s))|s).$$

Now since $E_6(\underset{\sim}{\tilde{v}}|\underset{\sim}{v},s)=\underset{\sim}{v}$ and $\text{cov}_6(\underset{\sim}{\tilde{y}},\underset{\sim}{\tilde{v}}|\underset{\sim}{v},s)=0$,

then

$$\underset{\sim}{\sum}_{yvs}=\text{COV}_v((E_6(\underset{\sim}{\tilde{y}}|\underset{\sim}{v},s),\underset{\sim}{v})|s)$$

$$=\text{COV}_v((\underset{\sim}{m}+\underset{\sim}{D}\underset{\sim}{v},\underset{\sim}{v})|s) \quad \text{Using [2.42] and [2.43]}$$

$$=\underset{\sim}{D}\,V_v(\underset{\sim}{v}|s)$$

$$=\underset{\sim}{D}\underset{\sim}{\sum}_{vvs}.$$

If $\underset{\sim}{\sum}_{vvs}$ is nonsingular then

$$\underset{\sim}{D}=\underset{\sim}{\sum}_{yvs}\underset{\sim}{\sum}_{vvs}^{-1}. \qquad [2.48]$$

Substituting [2.45] in [2.36],[2.46] in [2.39] then add and subtract $\underset{\sim}{\mu}_{fs}$ from [2.38] and substitute [2.47] in [2.38] we get the required result.

From theorem 2.6 we get the following estimator of the covariance matrix when the model is nonlinear and heteroscedastic, denoted by $\hat{\Sigma}_{yy,q2}$;

$$\hat{\Sigma}_{yy,q2}=\Sigma_{yys}+\hat{H}(\Sigma_{zz}-\Sigma_{zzs})\hat{H}^T+\hat{R}(\Sigma_{gg}-\Sigma_{ggs})\hat{R}^T+\hat{H}(\Sigma_{zg}-\Sigma_{zgs})\hat{R}^T$$
$$+\hat{R}(\Sigma_{gz}-\Sigma_{gzs})\hat{H}^T+(\mu_f-\mu_{fs}). \qquad [2.49]$$

Using [2.49] we propose the following estimator of the regression coefficient.

$$\hat{B}_{12,q2}=\hat{\Sigma}_{y1y2,q2}\hat{\Sigma}^{-1}_{y2y2,q2}.$$

COROLLARY 2.4

Lawleys' theorem is a special case of our generalized theorem when $R=0$ in the matrix D and $\mu_f=\mu_{fs}$, i.e for linear homoscedastic model.

2.6 CONCLUSION

In this chapter we have shown that the Pearson's adjusted estimators can be obtained via a loss function approach and also via the quadratic prediction approach. We have also obtained estimators which accounts for nonlinearity and heteroscedasticity by generalizing Lawleys' theorem. We also proposed a compromise criterion to use for comparing the performance of estimators irrespective of their distributional framework. By relaxing the normality distributional assumptions, we derived consistent predictors of the covariance matrix under the linear homoscedastic model. We will study the asymptotic properties of these estimators empirically in chapter 4.

CHAPTER 3

FULLER REGRESSION ESTIMATORS

3.1 INTRODUCTION

The search for a robust estimator led us to consider an estimator of the regression coefficient which was proposed by Fuller(1982). The basic idea in the construction of this estimator is to estimate separately each finite population total in the formulae for the finite population analogues of the intercept and slope by a regression type estimator.In his proposed estimator Fuller adjusted each of the sample estimators of the finite population total with a linear and quadratic function of the design variable z to guard against non linear relationship between the survey variables and the design variables. Holmes(1987) studied the Fuller estimator numerically with Horvitz-Thompson type of weights and found that there was no significant gain in efficiency using this estimator instead of the probability weighted estimator. Through some empirical studies he also found that the Nathan and Holt compromise estimator and the Fuller design consistent estimators are conditionally biased when the Linearity and Homoscedasticity assumptions are violated.

In this chapter we derived the theoretical asymptotic properties of the Fuller estimator under the linear homoscedastic model and also when these underlying model assumptions are violated.We will Extend Holmes(1987) work on the Fuller estimator with Horvitz-Thompson type of weights to the case of general weights.

3.2 FULLER ESTIMATORS OF THE REGRESSION COEFFICIENTS

In order to consider the case of regression analysis we partition the px1 dimensional random vector $\tilde{y}$ into two random vectors $\tilde{y}_1$ and $\tilde{y}_2$,such that $\tilde{y}_1$ is the dependent variable in the regression and $\tilde{y}_2$ is a p_2x1 vector of independent variables.For $P_2=1$,we assume the following model,

$$\tilde{y}_1 = \alpha_{12} + \beta_{12}\tilde{y}_2 + \varepsilon,$$

where $E(\varepsilon)=0$,

and

$$V(\varepsilon)=\sigma^2$$

E(.) and V(.) denotes expectation and variance,α_{12} and β_{12} denotes the superpopulation parameters to be estimated. Our objective is the point estimation of these superpopulation parameters given by,

$$\beta_{12}=\sigma_{12}/\sigma_2^2,$$

and

$$\alpha_{12}=\mu_1-\beta_{12}\mu_2,$$

where σ_{12},σ_2^2 and μ_1,μ_2 denotes the superpopulation covariances and means as defined in chapter 1.

If the finite population taken is large and is a random sample from the superpopulation then

$$B_{12}= \beta_{12}+O_p(N^{-1/2})$$
$$\cong \beta_{12},$$

and

$$A_{12}= \alpha_{12}+O_p(N^{-1/2})$$
$$\cong \alpha_{12},$$

where A_{12} and B_{12} denotes the finite population regression coefficients.Thus for large N,A_{12} and B_{12} are the least squares estimators of the superpopulation parameters α_{12} and β_{12}.

Now the finite population regression coefficients are given by

$$B_{12} = \frac{S_{12}}{S_{22}}$$

$$= \frac{N^{-1}\sum_U y_{1i}y_{2i} - \bar{y}_{1U}\bar{y}_{2U}}{N^{-1}\sum_U y_{2i}^2 - \bar{y}_{2U}^2},$$

and

$$A_{12} = \bar{y}_{1U} - B_{12}\bar{y}_{2U}.$$

B_{12} and A_{12} are the ordinary least squares estimators of the superpopulation parameters β_{12} and α_{12} when the finite population is taken to be a simple random sample from the superpopulation.

The above equations can be rewritten in matrix form as

$$\begin{bmatrix} A_{12} \\ B_{12} \end{bmatrix} = \begin{bmatrix} 1 & N^{-1}\sum_U y_{2i} \\ N^{-1}\sum_U y_{2i} & N^{-1}\sum_U y_{2i}^2 \end{bmatrix}^{-1} \begin{bmatrix} N^{-1}\sum_U y_{1i} \\ N^{-1}\sum_U y_{2i}y_{1i} \end{bmatrix}$$

$$= \underset{\sim}{M}_{22}^{-1}\ \underset{\sim}{M}_{21}$$

$$= h(\underset{\sim}{Q}), \text{say}, \qquad [3.1]$$

where

$$\underset{\sim}{M}_{22} = \begin{bmatrix} 1 & N^{-1}\sum_U y_{2i} \\ N^{-1}\sum_U y_{2i} & N^{-1}\sum_U y_{2i}^2 \end{bmatrix}, \qquad [3.2]$$

$$\underset{\sim}{M}_{21} = \begin{bmatrix} N^{-1}\sum_U y_{1i} \\ N^{-1}\sum_U y_{2i}y_{1i} \end{bmatrix}, \qquad [3.3]$$

and

$$\underset{\sim}{Q} = \left[\text{Vech}(\underset{\sim}{M}_{22})^T, \underset{\sim}{M}_{21}^T \right]^T,$$

where $\text{Vech}(\underset{\sim}{M}_{22})$ is the vector of distinct elements of the symmetric matrix $\underset{\sim}{M}_{22}$ that is

$$\text{Vech}(\underset{\sim}{M}_{22}) = \begin{bmatrix} 1 \\ N^{-1}\sum_U y_{2i} \\ N^{-1}\sum_U y_{2i}^2 \end{bmatrix}.$$

For example if $\underset{\sim}{A}$ denotes a symmetric matrix i.e

$$\underset{\sim}{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots\cdots & a_{1n} \\ & a_{22} & \cdots\cdots & a_{2n} \\ & & \cdots\cdots & \\ & & & a_{nn} \end{bmatrix},$$

then we define $\text{vech}(\underset{\sim}{A}) = [a_{11}, a_{12}, a_{22} \ldots\ldots a_{1n} \ldots a_{nn}]^T$.

Thus

$$\underset{\sim}{Q} = [\ 1,\ N^{-1}\sum_U y_{2i},\ N^{-1}\sum_U y_{2i}^2,\ N^{-1}\sum_U y_{1i},\ N^{-1}\sum_U y_{2i}y_{1i}]^T,$$

Let

$$\underset{\sim}{q}_i = [\ 1,\ y_{2i},\ y_{2i}^2,\ y_{1i},\ y_{2i}y_{1i}]^T,$$

then

$$\underset{\sim}{Q} = N^{-1}\sum_U \underset{\sim}{q}_i .$$

We consider estimating

(i) $\underset{\sim}{Q}$ by $\hat{\underset{\sim}{Q}}$,

and then

(ii) $h(\underset{\sim}{Q})$ by $h(\hat{\underset{\sim}{Q}})$.

One possible estimator of $\underset{\sim}{Q}$ is the weighted sample estimator $\hat{\underset{\sim}{Q}}_s$,

where

$$\hat{\underset{\sim}{Q}}_s=[\sum_s w_i,\ \sum_s w_i\ y_{2i},\ \sum_s w_i y_{2i}^2,\ \sum_s w_i y_{1i},\ \sum_s w_i y_{2i} y_{1i}]^T,$$

$$= \sum_s w_i \underset{\sim}{q}_i. \qquad [3.4]$$

We note that the estimator $\hat{\underset{\sim}{Q}}_s$ given in (3.4) is none other than the well known probability weighted estimator if the weights are taken to be the Horvitz-Thompson type of weights i.e $w_i=\pi_i^{-1}/N$ where π_i is the probability of inclusion of the i^{th} unit in the sample.

We now define the estimator of $h(\underset{\sim}{Q})$ of the form

$$\begin{bmatrix} \hat{A}_{12} \\ \hat{B}_{12} \end{bmatrix} = \hat{\underset{\sim}{M}}_{22}^{-1}\hat{\underset{\sim}{M}}_{21} = h(\hat{\underset{\sim}{Q}}_s), \qquad [3.5]$$

where

$$\hat{\underset{\sim}{M}}_{21} = \begin{bmatrix} \hat{Q}_{4s} \\ \hat{Q}_{5s} \end{bmatrix}, \text{ and } \hat{\underset{\sim}{M}}_{22} = \begin{bmatrix} 1 & \hat{Q}_{2s} \\ \hat{Q}_{2s} & \hat{Q}_{3s} \end{bmatrix}.$$

Now suppose we have some auxiliary information z_i on each unit then $\bar{z}_U=N^{-1}\sum_U z_i$ and $S_{zz}=N^{-1}\sum_U (z_i-\bar{z}_U)^2$ are the finite population mean and variance of this design variable z_i.The differences $(\bar{z}_s^*-\bar{z}_U)$ and $(S_{zzs}^*-S_{zz})$ where $\bar{z}_s^*=\sum_s w_i z_i$ and $S_{zzs}^*=\sum_s w_i (z_i-\bar{z}_s^*)^2$ tell us something about the sample drawn that can be used to improve the weighted estimator defined in (3.5).Fuller(1982) proposed an estimator which uses this differences to improve the weighted estimator.In his proposed estimator Fuller adjusted each of the sample estimators of the population total with a linear and quadratic function of the design variable to guard against nonlinear relationships between the survey variable and the design variables.His proposed estimator of $\underset{\sim}{Q}$,denoted by

$\hat{\tilde{Q}}_f$, is as defined below.

$$\hat{\tilde{Q}}_f = \sum_s w_i \tilde{q}_i - \sum_s w_i \tilde{q}_i (\tilde{t}_i - \bar{\tilde{t}}^*)^T \left[\sum_s w_i (\tilde{t}_i - \bar{\tilde{t}}^*)(\tilde{t}_i - \bar{\tilde{t}}^*)^T \right]^{-1} \bar{\tilde{t}}^*, \quad [3.6]$$

where

$$\tilde{t}_i = [z_i - \bar{z}_U, (z_i - \bar{z}_U)^2 - S_{zz}]^T,$$

$$\bar{\tilde{t}}^* = \sum_s^n w_i \tilde{t}_i,$$

Fuller chose the weights w_i such that $\sum_s w_i = 1$.

For example w_i may be equal to $\pi_i^{-1}/\sum_s \pi_i^{-1}$. We call these type of weights, the Horvitz-Thompson type with ratio adjustment.

We may write $\hat{\tilde{Q}}_f$ as

$$\hat{\tilde{Q}}_f = \hat{\tilde{Q}}_s - \tilde{S}^*_{qt} \tilde{S}^{*-1}_{tt} \tilde{m}^*_t, \quad [3.7]$$

where

$\hat{\tilde{Q}}_s$ is as defined in (3.4).

$$\tilde{S}^*_{qt} = \begin{bmatrix} \sum_s w_i (z_i - \bar{z}_U) & \sum_s w_i [(z_i - \bar{z}_U)^2 - \sum_s w_i (z_i - \bar{z}_U)^2] \\ \sum_s w_i y_{2i} (z_i - \bar{z}_U) & \sum_s w_i y_{2i} [(z_i - \bar{z}_U)^2 - \sum_s w_i (z_i - \bar{z}_U)^2] \\ \sum_s w_i y_{2i}^2 (z_i - \bar{z}_U) & \sum_s w_i y_{2i}^2 [(z_i - \bar{z}_U)^2 - \sum_s w_i (z_i - \bar{z}_U)^2] \\ \sum_s w_i y_{1i} (z_i - \bar{z}_U) & \sum_s w_i y_{1i} [(z_i - \bar{z}_U)^2 - \sum_s w_i (z_i - \bar{z}_U)^2] \\ \sum_s w_i y_{1i} y_{2i} (z_i - \bar{z}_U) & \sum_s w_i y_{1i} y_{2i} [(z_i - \bar{z}_U)^2 - \sum_s w_i (z_i - \bar{z}_U)^2] \end{bmatrix},$$

$$= [\tilde{S}_{q_1 t}, \tilde{S}_{q_2 t}, \tilde{S}_{q_3 t}, \tilde{S}_{q_4 t}, \tilde{S}_{q_5 t}]^T,$$

$$\hat{\tilde{Q}}_f = [\hat{Q}_{1f}, \hat{Q}_{2f}, \hat{Q}_{3f}, \hat{Q}_{4f}, \hat{Q}_{5f}],$$

$$\underset{\sim}{S}^{*}_{tt}=$$

$$\begin{bmatrix} \sum_s w_i \,[(z_i-\bar{z}^{*}_s)^2 , & \sum_s w_i \,[(z_i-\bar{z}_U)^2 - \sum_s w_i \,[(z_i-\bar{z}_U)^2](z_i-\bar{z}^{*}_s) \\ \sum_s w_i[(z_i-\bar{z}_U)^2 - \sum_s w_i[(z_i-\bar{z}_U)^2](z_i-\bar{z}^{*}_s), & \sum_s w_i[(z_i-\bar{z}_U)^2-\sum_s w_i[(z_i-\bar{z}_U)^2]^2 \end{bmatrix}$$

and

$$\underset{\sim}{m}^{*}_{t} = [\, \sum_s w_i \,[(z_i-\bar{z}_U) \,,\, \sum_s w_i \,[(z_i-\bar{z}_U)^2]-S_{zz}] \,]^{T}.$$

We see that the elements of $\hat{\underset{\sim}{Q}}_f$ are the usual weighted sample estimators corrected for the differences between the sample and the population characteristics of the design variable z. From (3.6) the estimators of the finite population means are given by

$$\hat{\bar{y}}_{2U}= \sum_s w_i y_{2i} - \underset{\sim}{S}^{*}_{q_{2t}} \; \underset{\sim}{S}^{*-1}_{tt} \; \underset{\sim}{m}^{*}_{t}, \qquad [3.8]$$

$$\hat{\bar{y}}_{1U}= \sum_s w_i y_{1i} - \underset{\sim}{S}^{*}_{q_{4t}} \; \underset{\sim}{S}^{*-1}_{tt} \; \underset{\sim}{m}^{*}_{t} \,, \qquad [3.9]$$

$$\hat{\bar{y}}_{12U}= \sum_s w_i y_{2i} y_{1i} - \underset{\sim}{S}^{*}_{q_{5t}} \; \underset{\sim}{S}^{*-1}_{tt} \; \underset{\sim}{m}^{*}_{t}, \qquad [3.10]$$

$$\hat{\bar{y}}_{22U}= \sum_s w_i y^2_{2i} - \underset{\sim}{S}^{*}_{q_{3t}} \; \underset{\sim}{S}^{*-1}_{tt} \; \underset{\sim}{m}^{*}_{t},$$

where

$\underset{\sim}{S}_{q_{2t}}$, $\underset{\sim}{S}_{q_{3t}}$, $\underset{\sim}{S}_{q_{4t}}$ and $\underset{\sim}{S}_{q_{5t}}$ are defined above.

We estimate $h(\underset{\sim}{Q})$ by $h(\hat{\underset{\sim}{Q}}_f)$, where

$$h(\underset{\sim}{\hat{Q}_f}) = \begin{bmatrix} \hat{A}_{12} \\ \hat{B}_{12} \end{bmatrix} = \underset{\sim}{\hat{M}}_{22}^{-1}\underset{\sim}{\hat{M}}_{21}, \qquad [3.11]$$

$$\underset{\sim}{\hat{M}}_{21} = \begin{bmatrix} \hat{Q}_{4f} \\ \hat{Q}_{5f} \end{bmatrix}, \qquad [3.12]$$

and

$$\underset{\sim}{\hat{M}}_{22} = \begin{bmatrix} 1 & \hat{Q}_{2f} \\ \hat{Q}_{2f} & \hat{Q}_{3f} \end{bmatrix}.$$

In the next section we derive the asymptotic properties of these Fuller estimators of the regression coefficients under the (i) linear homoscedastic model assumptions,

(ii) quadratic homoscedastic model assumptions

and

(iii)linear heteroscedastic model assumptions.

3.2.1 ASYMPTOTIC PROPERTIES OF THE FULLER ESTIMATOR OF THE REGRESSION COEFFICIENT UNDER THE LINEAR AND HOMOSCEDASTIC MODEL 1

Consider the linear homoscedastic model 1 given in chapter 1.In this chapter we will consider the case where p=2 and q=1.Under these conditions the vectors $\underset{\sim}{m}$ and $\underset{\sim}{H}$ and the matrix $\underset{\sim}{K}$ defined in chapter 1,are given by,

$$\underset{\sim}{m}=[m_1,m_2]^T, \underset{\sim}{H}=[H_1,H_2]^T \text{ and } \underset{\sim}{K}=\begin{vmatrix} K_{11} & K_{12} \\ K_{12} & K_{22} \end{vmatrix}$$

where $m_1=\mu_{y1}-H_1\mu_z, m_2=\mu_{y2}-H_2\mu_z$,

$$K_{11}=\Sigma_{y1y1}-H_1\,\Sigma_{zz}H_1, K_{12}=\Sigma_{y1y2}-H_1\,\Sigma_{zz}H_2$$

$$K_{22}=\Sigma_{y2y2}-H_2\,\Sigma_{zz}H_2, H_1=\Sigma_{y1z}\,\Sigma_{zz}^{-1} \text{ and } H_2=\Sigma_{y2z}\,\Sigma_{zz}^{-1}. \qquad [3.13]$$

From [3.6] the Fuller estimator of the regression coefficient is given by

$$\hat{\underset{\sim}{Q}}_f=\sum_s w_i \underset{\sim}{q}_i - \sum_s w_i \underset{\sim}{q}_i (\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T [\sum_s w_i (\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*,$$

$$=\sum_s \lambda_i \underset{\sim}{q}_i , \qquad [3.14]$$

where $\lambda_i=w_i - w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T[\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*$.

From [3.14] we see that the Fuller estimator is a linear function of the survey variables $\underset{\sim}{q}_i$,which may be linear or quadratic.We expect the asymptotic properties of $\hat{\underset{\sim}{Q}}_f$ to depend on the functional form of $\underset{\sim}{q}_i$,i.e whether it is linear or quadratic.Since $\underset{\sim}{q}_i$ has three linear components and two quadratic components,we will derive the asymptotic properties of only the second (linear) and thefifth (quadratic) component of $\hat{\underset{\sim}{Q}}_f$,i.e $\hat{Q}_{2f}$ and $\hat{Q}_{5f}$ and deduce the properties of the other linear and quadratic components accordingly.Throughout this chapter we will consider only those cases where $\sum_s w_i=1$.

We will look at the asymptotic properties of $\hat{\underset{\sim}{Q}}_f$ and then apply the standard result given in lemmas 3.1 and 3.2 below to deduce those of $h(\hat{\underset{\sim}{Q}}_f)$.

LEMMA 3.1

If $\underset{\sim}{Q}_f$ is a consistent estimator of $\underset{\sim}{Q}$ and h denotes a continous function ,then $h(\hat{\underset{\sim}{Q}}_f)$ is also a consistent estimator of $h(\underset{\sim}{Q})$ i.e

if $\hat{\underset{\sim}{Q}}_f - \underset{\sim}{Q} \xrightarrow{p} 0$ as $n,N \longrightarrow \infty$,

then

$h(\hat{\underset{\sim}{Q}}_f) - h(\underset{\sim}{Q}) \xrightarrow{p} 0$ as $n,N \longrightarrow \infty$.

Proof (see C.R.Rao 1973 p124).

LEMMA 3.2

Assuming that the weights w_i are of order n^{-1}, then under model ε, the unconditional variance of the Fuller estimators tend to zero for large n and N, i.e

$$\mathrm{Var}(\hat{\underset{\sim}{Q}}_f)=0, \text{ as } n,N\Rightarrow\infty.$$

Proof

We can write the Fuller estimator given in [3.14] alternatively as,

$$\hat{\underset{\sim}{Q}}_f=\sum w_i(1-g_i)\underset{\sim}{q}_i,$$

$$\text{where} \quad g_i=(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^*)^T[\sum_s w_i(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^*)(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^*)^T]^{-1}\bar{\underset{\sim}{t}}^*.$$

Taking conditional variances given s and z and also assuming that $\underset{\sim}{q}_i$ are mutually independent we get,

$$\mathrm{Var}_\varepsilon(\hat{\underset{\sim}{Q}}_f|s,z)=\sum w_i^2(1-g_i)^2\mathrm{Var}_\varepsilon(\underset{\sim}{q}_i|z).$$

Taking limits as $n,N\Rightarrow\infty$, we get

$$\lim_{n,N\Rightarrow\infty}\mathrm{Var}_\varepsilon(\hat{\underset{\sim}{Q}}_f|s,z)=0, \text{since } w_i=O(n^{-1}) \text{ and } \mathrm{Var}_\varepsilon(\underset{\sim}{q}_i|z)<\infty.$$

Now

$$\lim_{n,N\Rightarrow\infty}\mathrm{Var}(\hat{\underset{\sim}{Q}}_f)=\lim_{n,N\Rightarrow\infty}E_p(\mathrm{Var}_\varepsilon(\hat{\underset{\sim}{Q}}_f|s,z))+\lim_{n,N\Rightarrow\infty}\mathrm{Var}_p(E_\varepsilon(\hat{\underset{\sim}{Q}}_f|s,z))$$

$$=0.$$

Hence the result.

THEOREM 3.1

Under model 1 to order $O(n^{-1/2})$ approximation and assuming that the population size N is large and $\sum_s w_i=1$, then in general the Fuller estimator $h(\hat{\underset{\sim}{Q}}_f)$ is an asymptotically unbiased estimator of $h(\underset{\sim}{Q})$ i.e

$$E[h(\hat{\underset{\sim}{Q}}_f)]=h(\underset{\sim}{Q}) \quad \text{as } n,N \Rightarrow \infty,$$

where $h(\hat{\underset{\sim}{Q}}_f)$ denotes the Fuller estimators of the regression coefficients defined in [3.11].

Proof

We look at the asymptotic properties of the second and fifth components of $\hat{\underset{\sim}{Q}}_f$ separately.

The second component of [3.14] is given by

$$\hat{Q}_{2f}=\sum_s w_i y_{2i}-\sum_s w_i y_{2i}(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T[\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*.$$

Taking conditional expectation w.r.t the model 1 we have

$$E_1(\hat{Q}_{2f}|z,s)=\sum_s w_i E_1(y_{2i}|z)- \sum_s w_i E_1(y_{2i}|z,s)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T$$
$$[\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*. \qquad [3.15]$$

Consider the first term of (3.15)

$$\sum_s w_i E_1(y_{2i}|z)=\sum_s w_i[\mu_{y2}+H_2(z_i-\mu_z)$$

using model 1 assumptions and eqns[3.13] ,

$$=\mu_{y2}+H_2(\bar{z}_s^*-\mu_z). \qquad [3.16]$$

Evaluating the second term of **(3.15)** we have

$$\sum_s w_i E_1(y_{2i}|z) \ (\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T[\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*$$

$$= \sum_s w_i[\mu_{y2}+H_2(z_i-\mu_z)](\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^*)^T[\sum_s w_i(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^*)(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^*)^T]^{-1}\bar{\underset{\sim}{t}}^*$$

using model 1 assumptions and eqns[3.13],

$$=H_2\sum_s w_i(z_i-\bar{z}_U)(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^*)^T[\sum_s w_i(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^*)(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^*)^T]^{-1}\bar{\underset{\sim}{t}}^*$$

$$=(H_2,0)\sum_s w_i(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^*)(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^*)^T[\sum_s w_i(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^*)(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^*)^T]^{-1}\bar{\underset{\sim}{t}}^*$$

$$=H_2(\bar{z}_s^*-\bar{z}_U). \qquad [3.17]$$

Subtituting [3.16] and [3.17] in [3.15] we get

$$E_1(\hat{Q}_{2f}|z,s)=\mu_{y2}+H_2(\bar{z}_s^*-\mu_z)-H_2(\bar{z}_s^*-\bar{z}_U)$$

$$=\mu_{y2}+H_2(\bar{z}_U-\mu_z)$$

$$=\bar{y}_{2U}+O_p(N^{-1/2})$$

$$\cong Q_2, \text{ since N is large.} \qquad [3.18]$$

Averaging [3.18] over all possible samples we get

$$E_pE_1(\hat{Q}_{2f})\cong Q_2,$$

hence

$$E(\hat{Q}_{2f})\cong Q_2. \qquad [3.19]$$

From [3.18] and [3.19] we see that under model 1 $\hat{Q}_{2f}$ is approximately conditionally and unconditionally unbiased.

We now consider the fifth term of (3.14).Taking conditional expectation of $\hat{Q}_{5f}$ w.r.t model 1 we get

$$E_1(\hat{Q}_{5f}|z,s)=\sum_s w_iE_1(y_{1i}y_{2i}|z)-\sum_s w_iE_1(y_{1i}y_{2i}|z)(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^*)^T$$

$$[\sum_s w_i(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^*)(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^*)^T]^{-1}\bar{\underset{\sim}{t}}^*. \qquad [3.20]$$

To evaluate [3.20] we need to evalute the term

$$E_1(y_{1i}y_{2i}|z)= \text{cov}_1[(y_{1i}y_{2i}|z)+E_1(y_{1i}|z)E_1(y_{2i}|z)]$$

$$=[K_{12}+m_1m_2+H_2m_1z_i+H_1m_2z_i+H_1H_2z_i^2].$$

$$=\Sigma_{y1y2}-H_1H_2\Sigma_{zz}-H_2\mu_{y1}\mu_z-H_1\mu_{y2}\mu_z+H_1H_2\mu_z^2+\mu_{y1}\mu_{y2}$$

$$+H_2(\mu_{y1}-H_1\mu_z)z_i+H_1(\mu_{y2}-H_2\mu_z)z_i+H_1H_2\sum_{i=1}^{n}w_iz_i^2$$

using model 1 assumptions and eqns[3.13].

[3.21]

Substituting [3.21] in the first term of [3.20] we get

$$\sum_s w_iE_1(y_{1i}y_{2i}|z)=\Sigma_{y1y2}-H_1H_2\Sigma_{zz}-H_2\mu_{y1}\mu_z-H_1\mu_{y2}\mu_z+H_1H_2\mu_z^2$$

$$+\mu_{y1}\mu_{y2}+H_2(\mu_{y1}-H_1\mu_z)\bar{z}_s^*+H_1(\mu_{y2}-H_2\mu_z)\bar{z}_s^*+H_1H_2\sum_{i=1}^{n}w_iz_i^2$$

$$=\Sigma_{y1y2}+H_1H_2[S^*_{zzs}-\Sigma_{zz}]+H_2\mu_{y1}[\bar{z}_s^*-\mu_z]+$$

$$H_1\mu_{y2}[\bar{z}_s^*-\mu_z]+H_1H_2[\bar{z}_s^{*2}-\mu_z^2] \qquad [3.22]$$

Substituting eqn[3.21] in the second term of [3.20] we get

$$\sum_s w_iE_1(y_{1i}y_{2i}|z)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T[\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*$$

$$=\sum_s w_i(H_2(\mu_{y1}-H_1\mu_z)z_i+H_1(\mu_{y2}-H_2\mu_z)z_i+H_1H_2z_i^2)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T$$

$$[\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*$$

$$=[H_2\mu_{y1}\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T(z_i-\mu_z)+H_1\mu_{y2}\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T(z_i-\bar{z}_U)$$

$$+H_1H_2\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T((z_i-\mu_z)^2-S_{zz})][\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*$$

$$\cong[H_2\mu_{y1}\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T(z_i-\bar{z}_U)+H_1\mu_{y2}\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T(z_i-\bar{z}_U)$$

$$+H_1H_2\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T((z_i-\bar{z}_U)^2-S_{zz})][\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*$$

since for large N $\bar{z}_U\cong\mu_z$,

$$=[((H_2\mu_{y1},0)\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T)$$

$$+((H_1\mu_{y2},0)\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T)$$

$$+((0,H_1H_2)\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*$$

$$=(H_2\mu_{y1},0)+(H_1\mu_{y2},0)+(0,H_1H_2))\underset{\sim}{\bar{t}}^*$$

$$=H_2\mu_{y1}\bar{z}^*_s-H_2\mu_{y1}\bar{z}_U+H_1\mu_{y2}\bar{z}^*_s-H_1\mu_{y2}\bar{z}_U+H_1H_2\bar{z}_U^{\,2}+H_1H_2\sum_{i=1}^{n}w_iz_i^2$$

$$-2H_1H_2\bar{z}^*_s\bar{z}_U-H_1H_2S_{zz}. \qquad [3.23]$$

Substitute eqn[3.23] and [3.22] in [3.20] we get,

$$E_1(\hat{Q}_{5f}|z,s)=\Sigma_{y1y2}-H_1H_2\Sigma_{zz}-H_2\mu_{y1}\mu_z-H_1\mu_{y2}\mu_z+H_1H_2\mu_z^2$$

$$+H_2(\mu_{y1}-H_1\mu_z)\bar{z}^*_s+H_1(\mu_{y2}-H_2\mu_z)\bar{z}^*_s+H_1H_2\sum_{i=1}^{n}w_iz_i^2$$

$$+\mu_{y1}\mu_{y2}-H_2\mu_{y1}\bar{z}^*_s+H_2\mu_{y1}\bar{z}_U-H_1\mu_{y2}\bar{z}^*_s$$

$$+H_1\mu_{y2}\bar{z}_U-H_1H_2\bar{z}_U^2-H_1H_2\sum_{i=1}^{n}w_iz_i^2+2H_1H_2\bar{z}^*_s\bar{z}_U+H_1H_2S_{zz}$$

$$=\Sigma_{y1y2}+\mu_{y1}\mu_{y2}+H_1H_2(S_{zz}-\Sigma_{zz})+H_2\mu_{y1}(\bar{z}_U-\mu_z)-H_2H_1(\bar{z}_U^2-\mu^2_z)$$

$$+H_1\mu_{y2}(\bar{z}_U-\mu_z)+2H_2H_1\bar{z}^*_S(\bar{z}_U-\mu_z)$$

$$=\Sigma_{y1y2}+\mu_{y1}\mu_{y2}+O_p(N^{-1/2})$$

$$\cong Q_5,\quad \text{since N is large.} \qquad [3.24]$$

Averaging eqn [3.24] over all possible samples we get

$$E_pE_1(\hat{Q}_{5f})\cong Q_5.$$

$$\Rightarrow E(\hat{Q}_{5f})\cong Q_5. \qquad [3.25]$$

We see from eqn [3.24] and [3.25] that under the model 1 $\hat{Q}_{5f}$ is approximately conditionally and unconditionally unbiased estimator of $\hat{Q}_5$.

We see from eqn [3.18] ,[3.19] and [3.24],[3.25] that the linear and quadratic components of $\underset{\sim}{\hat{Q}_f}$ are both approximately conditionally and unconditionally unbiased under model 1.Thus,

$$E_1(\underset{\sim}{\hat{Q}_f}|z,s)=\underset{\sim}{Q}\ ,\ \text{as } n,N\Rightarrow\infty. \qquad [3.26]$$

and

$$E_p E_1(\hat{\underset{\sim}{Q}}_f) = \underset{\sim}{Q} \Rightarrow E(\hat{\underset{\sim}{Q}}_f) = \underset{\sim}{Q}, \quad \text{as } n, N \Rightarrow \infty. \qquad [3.27]$$

Applying lemmas 3.1 and 3.2 to eqn[3.27] we get the required result.

We will now study the asymptotic property of the weighted estimator $\hat{\underset{\sim}{Q}}_s$ with general weights under model 1,but with the restriction that $\sum_s w_i = 1$.

THEOREM 3.2

Under model 1 to $O(n^{-1/2})$ approximation and assuming N is large then if the weights w_i are Horvitz Thompson type with the ratio adjustment i.e $w_i = \pi_i^{-1} / \sum_s \pi_i^{-1}$ then the weighted estimator $h(\hat{\underset{\sim}{Q}}_s$ is asymptotically unbiased i.e,

$$E[h(\hat{\underset{\sim}{Q}}_s)] = h(\underset{\sim}{Q}) \text{ ,as } n, N \Rightarrow \infty.$$

where $\hat{\underset{\sim}{Q}}_s$ is as defined in [3.4].

Proof.

From [3.4] we have

$$\hat{\underset{\sim}{Q}}_s = \sum_s w_i \underset{\sim}{q}_i .$$

Taking conditional expectation of $\hat{\underset{\sim}{Q}}_s$ under model 1 we get

$$E_1(\hat{\underset{\sim}{Q}}_s | z, s) = \sum_s w_i E(\underset{\sim}{q}_i | z), \quad \text{as } n, N \Rightarrow \infty.$$

From [3.16] and [3.22] we deduce that

$$E_1(\hat{\underset{\sim}{Q}}_s | z, s) = \underset{\sim}{Q} \text{ ,as } n, N \Rightarrow \infty.$$

Averaging this conditional expectation over all possible samples we get

$$E_p E_1(\hat{\underset{\sim}{Q}}_s | z, s) = \underset{\sim}{Q} \text{ ,as } n, N \Rightarrow \infty.$$

$$\Rightarrow E(\hat{\underset{\sim}{Q}}_s) = \underset{\sim}{Q} \text{ ,as } n, N \Rightarrow \infty.$$

Applying Lemmas 3.1 and 3.2 to this result we get the required result.

Note that for general weights which does not satisfy $\sum_s w_i=1$ theorem 3.2 does not hold.

Skinner(1982),Holmes(1987) have proved that the maximum likelihood estimator is asymptotically unbiased estimator and the probability weighted estimator is conditionally biased under model 1. Theorem 3.1 results show that the Fuller estimator is also asymptotically unbiased estimator under model 1 for any set of weights w_i,provided that $\sum_s w_i=1$,hence is a serious contender to the maximum likelihood estimator and is preferable to the probability weighted estimator.

We will now study the asymptotic properties of the Fuller estimator when Model 1 assumptions are violated.We will treat the violations of the linearity and Homoscedasticity assumptions separately to get a clear picture.

3.2.2 ASYMPTOTIC PROPERTIES OF THE FULLER REGRESSION ESTIMATOR UNDER NONLINEAR HOMOSCEDASTIC MODEL.(MODEL 2)

We will derive the asymptotic properties of the Fuller estimators when the linearity assumption is violated.We call the resulting model,model 2.Under model 2 we assume that,

(i) The conditional expectation of $\tilde{\underset{\sim}{y}}$ given $\underset{\sim}{z}$ is a nonlinear function of $\underset{\sim}{z}$.

(ii) The conditional variance covariance matrix of $\tilde{\underset{\sim}{y}}$ given $\underset{\sim}{z}$ is constant.

(iii) $y_i \perp\!\!\!\perp y_j | z$, for $i\neq j$.

Thus the conditional moments of the survey variable $\underset{\sim}{y}$ given the design variable $\underset{\sim}{z}$ are assumed to obey

$$E_2(\tilde{\underset{\sim}{y}}|\underset{\sim}{z})=\underset{\sim}{m}+\underset{\sim}{H}\underset{\sim}{z}+\underset{\sim}{G}\underset{\sim}{x},$$

and

$$V_2(\tilde{y}|z)=K,$$

where m is a px1 constant vector, H and G are pxq and pxr constant matrices respectively. z and x are qx1 and rx1 vectors of the realized values of the random variables $\tilde{z}$ and $\tilde{x}$ respectively and x is a known function of z.

Under model 2 we have

$$m=\mu_y+H\mu_z+G\mu_x, \qquad [3.28]$$

and

$$K=\Sigma_{yy}-H\Sigma_{zz}H^T-G\Sigma_{xx}G^T-H\Sigma_{zx}G^T-G^T\Sigma_{xz}H, \qquad [3.29]$$

where $\mu_x=E(\tilde{x})$, $\Sigma_{xx}=V(\tilde{x})$, $\Sigma_{zx}=Cov(z,x)$,

and

$$[H,G]=[\Sigma_{yz},\Sigma_{yx}]\begin{bmatrix}\Sigma_{zz},\Sigma_{zx}\\ \Sigma_{yz},\Sigma_{yx}\end{bmatrix}^{-1}. \qquad [3.30]$$

In this study we will consider the case where $p=2$, $q=1$, $r=1$ and $x=z^2$.

THEOREM 3.3

Under model 2 to order $O(n^{-1/2})$ approximation and assuming that the population size N is large and $\sum_s w_i=1$, then in general the Fuller estimator $h(\hat{Q}_f)$ is an asymptotically biased estimator of $h(Q)$ i.e

$$E[h(\hat{Q}_f)]\neq h(Q), \text{ as } n,N\Rightarrow\infty.$$

where $h(\hat{Q}_f)$ denotes the Fuller estimators of the regression coefficients defined in [3.11].

Proof

We will derive the asymptotic properties for the second and fifth component of the Fuller estimators $\hat{\underset{\sim}{Q}}_f$ separately and then deduce those of the other components and then of $h(\hat{\underset{\sim}{Q}}_f)$.

Taking conditional expectation of the second term of eqn [3.14] w.r.t model 2 we have

$$E_2(\hat{Q}_{2f}|z,s)=\sum_s w_i E_2(y_{2i}|z)-\sum_s w_i E_2(y_{2i}|z)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T$$

$$[\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*. \qquad [3.31]$$

Consider the first term of eqn [3.31]

$$\sum_s w_i E_2(y_{2i}|z,s)= \sum_s w_i(m_2+H_2z_i+G_2x_i)$$

using model 2 assumptions.

$$= m_2+H_2\bar{z}_s^*+G_2\bar{x}_s^*$$

$$=\mu_{y2}+H_2(\bar{z}_s^*-\mu_z)+G_2(\bar{x}_s^* -\mu_x), \qquad [3.32]$$

using eqn [3.28].

Now the second term of [3.31] is

$$\sum_s w_i E_2(y_{2i}|z)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T[\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*$$

$$=\sum_s w_i(\mu_{y2}+H_2(z_i-\mu_z)+G_2(x_i -\mu_x))(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T$$

$$[\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*,$$

using model 2 assumptions and eqn[3.28]

$$= \sum_s w_i(H_2(z_i-\mu_z)+G_2(x_i -\mu_x))(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T$$

$$[\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*$$

$$= \sum_s w_i (H_2, G_2) \begin{bmatrix} z_i - \bar{z}_U \\ x_i - \bar{x}_U \end{bmatrix} (\underset{\sim}{t}_i - \underset{\sim}{\bar{t}}^*)^T [\sum_s w_i (\underset{\sim}{t}_i - \underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i - \underset{\sim}{\bar{t}}^*)^T]^{-1} \underset{\sim}{\bar{t}}^*,$$

[3.33]

where $\bar{x}_U = N^{-1} \sum_{i=1}^{N} x_i$.

Now consider

$$t_{2i} = (z_i - \bar{z}_U)^2 - S_{zz}$$

$$\Rightarrow \quad z_i^2 - \bar{x}_U = t_{2i} - \bar{x}_U - z_i^2 + \bar{z}_U^2 - 2\bar{z} z_i + S_{zz}$$

$$x_i - \bar{x}_U = t_{2i} - [\bar{x}_U + \bar{z}_U^2 - 2\bar{z}_U z_i - S_{zz}] \quad \text{since } x_i = z_i^2$$ [3.34]

substitute [3.34] in [3.33],[3.33] becomes

$$= \sum_s w_i (H_2, G_2) \begin{bmatrix} t_{1i} \\ t_{2i} - [\bar{x}_U + \bar{z}_U^2 - 2\bar{z}_U z_i - S_{zz}] \end{bmatrix} (\underset{\sim}{t}_i - \underset{\sim}{\bar{t}}^*)^T$$

$$[\sum_s w_i (\underset{\sim}{t}_i - \underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i - \underset{\sim}{\bar{t}}^*)^T]^{-1} \underset{\sim}{\bar{t}}^*$$

$$= \sum_s w_i (H_2, G_2)(\underset{\sim}{t}_i - \underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i - \underset{\sim}{\bar{t}}^*)^T [\sum_s w_i (\underset{\sim}{t}_i - \underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i - \underset{\sim}{\bar{t}}^*)^T]^{-1} \underset{\sim}{\bar{t}}^*$$

$$+ \sum_s w_i (H_2, G_2) \begin{bmatrix} 0 \\ -[\bar{x}_U + \bar{z}_U^2 - 2\bar{z}_U z_i - s_{zz}] \end{bmatrix} (\underset{\sim}{t}_i - \underset{\sim}{\bar{t}}^*)$$

$$[\sum_s w_i (\underset{\sim}{t}_i - \underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i - \underset{\sim}{\bar{t}}^*)^T]^{-1} \underset{\sim}{\bar{t}}^*$$

$$= (H_2, G_2) \underset{\sim}{\bar{t}}^* + (2\bar{z}_U G_2, 0) \sum_s w_i (\underset{\sim}{t}_i - \underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i - \underset{\sim}{\bar{t}}^*)^T$$

$$[\sum_s w_i (\underset{\sim}{t}_i - \underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i - \underset{\sim}{\bar{t}}^*)^T]^{-1} \underset{\sim}{\bar{t}}^*$$

$$= (H_2, G_2) \underset{\sim}{\bar{t}}^* + (2\bar{z}_U G_2, 0) \underset{\sim}{\bar{t}}^*$$

$$= (H_2 + 2\bar{z}_U G_2, G_2) [(\bar{z}_s^* - \bar{z}_U), \sum_s w_i [(z_i - \bar{z}_U)^2] - S_{zz}]^T$$

$$=(H_2 +2\bar{z}_U G_2)(\bar{z}_s^{*}-\bar{z}_U)+G_2 \ [\sum_s w_i \ [(z_i-\bar{z}_U)^2]-S_{zz}]. \qquad [3.35]$$

substitute [3.35] and [3.32] in [3.31] we get

$E_2(\hat{Q}_{2f}|z,s)$

$$=\mu_{y2}+H_2(\bar{z}_s^{*}-\mu_z)+G_2(\bar{x}_s^{*} -\mu_x)-((H_2+2\bar{z}_U G_2)(\bar{z}_s^{*}-\bar{z}_U)$$

$$+G_2 \sum_{i=1}^{n} w_i[(z_i-\bar{z}_U)^2]-S_{zz}]$$

$$= \mu_{y2}+H_2(\bar{z}_s^{*}-\mu_z)+G_2(\bar{x}_s^{*} -\mu_x)-((H_2 +2\bar{z}_U G_2)(\bar{z}_s^{*}-\bar{z}_U)-G_2(\bar{z}_s^{*}-\bar{z}_U)^2$$

$$-G_2(S_{zzs}^{*}-S_{zz})$$

$$=\mu_{y2}+G_2(\bar{x}_s^{*} -\mu_x)- 2\bar{z}_U G_2(\bar{z}_s^{*}-\bar{z}_U)+G_2(\bar{z}_s^{*}-\bar{z}_U)^2-G_2(S_{zzs}^{*}-S_{zz}). [3.36]$$

We see from [3.36] that under model 2 Q_{2f} is conditionally biased.
Averaging eqn [3.36] over all possible samples we get

$$E_p E_2(\underset{\sim}{\hat{Q}}_{2f}|z,s)=\mu_{y2}+G_2(E_p(\bar{x}_s^{*}) -\mu_x)- 2\bar{z}_U G_2(E_p(\bar{z}_s^{*})-\bar{z}_U)$$

$$+G_2(E_p[(\bar{z}_s^{*}-\bar{z}_U)^2] -g_2(E_p(S_{zzs}^{*})-S_{zz}). \qquad [3.37]$$

From [3.37] we see that since $E_p(\bar{x}_s^{*})\neq\mu_x$, $E_p(\bar{z}_s^{*})\neq\bar{z}_U$ and $E_p(S_{zzs}^{*})\neq S_{zz}$, then in general $\hat{Q}_{2f}$ is also unconditionally biased.

However if the weights are Horvitz-Thompson type with the ratio adjustment i.e then $w_i = \pi_i^{-1} / \sum_s \pi_i^{-1}$

$$E_p(\bar{x}_s^*) \cong \mu_x \text{ , } E_p(\bar{z}_s^*) \cong \bar{z}_U \text{ and } E_p(S_{zzs}^*) \cong S_{zz},$$

Then

$$E_p E_2(\hat{Q}_{2f} | z, s) \cong \mu_{y2}$$
$$= Q_2 .$$
$$\Rightarrow E(\hat{Q}_{2f}) \cong Q_2 . \qquad [3.38]$$

From [3.38] $\hat{Q}_{2f}$ is approximately unconditionally unbiased.

Taking conditional expectation of the fifth term $\hat{Q}_{5f}$ in [3.14] with respect to model 2 we have

$$E_2(\hat{Q}_{5f} | z, s) = \sum_s w_i E_2(y_{1i} y_{2i} | z) - \sum_s w_i E_2(y_{1i} y_{2i} | z)(\tilde{t}_i - \tilde{\bar{t}}^*)^T$$
$$[\sum_s w_i (\tilde{t}_i - \tilde{\bar{t}}^*)(\tilde{t}_i - \tilde{\bar{t}}^*)^T]^{-1} \tilde{\bar{t}}^* \qquad [3.39]$$

Consider the first term of [3.39] we get

$$\sum_s w_i E_2(y_{1i} y_{2i} | z)$$

$$= \sum_s w_i [cov_2(y_{2i}, y_{2i}) | z) + E_2(y_{1i} | z) E_2(y_{2i} | z)]$$

$$= \sum_s w_i [K_{12} + (m_1 + H_1 z_i + G_1 x_i)(m_2 + H_2 z_i + G_2 x_i) \quad \text{using model 2}$$

$$= \sum_s w_i [K_{12} + m_1 m_2 + m_2 (H_1 z_i + G_1 x_i) + m_1 (H_2 z_i + G_2 x_i)$$
$$+ H_1 H_2 z_i^2 + H_2 G_1 x_i z_i + G_2 H_1 x_i z_i + G_2 G_1 x_i^2]$$

$$= K_{12} + m_1 m_2 + \sum_s w_i [(\mu_{y2} - H_2 \mu_z - G_2 \mu_x)(H_1 z_i + G_1 x_i)$$
$$+ (\mu_{y1} - H_1 \mu_z - G_1 \mu_x)(H_2 z_i + G_2 x_i) + H_1 H_2 z_i^2 + H_2 G_1 x_i z_i + G_2 H_1 x_i z_i + G_2 G_1 x_i^2]$$

Substituting the value of K_{12}, m_1 and m_2 from [3.29] and [3.28]

$$\sum_s w_i E_2(y_{1i}y_{2i}|z)$$

$$=\Sigma_{y1y2} + H_1H_2(S^*_{zzs}-\Sigma_{zz}) + G_1G_2\ (S^*_{xxs}-\Sigma_{xx}) + G_2H_1(S^*_{xzs}-\Sigma_{xz})$$

$$+G_1H_2(S^*_{zxs}-\Sigma_{zx}) + H_2\mu_{y1}(\bar{z}^*_s-\mu_z) + G_2\mu_{y1}(\bar{x}^*_s-\mu_x)$$

$$+H_1\mu_{y2}(\bar{z}^*_s-\mu_z) + H_1H_2[\mu_z^2+\bar{z}^{*2}_s-2\mu_z\bar{z}^*_s)$$

$$+\mu_{y1}\mu_{y2}+G_1G_2[\mu_x^2+\bar{x}^{*2}_s-2\mu_x\bar{x}^*_s) + H_1G_2(\bar{x}^*_s\bar{z}^*_s-\mu_z\bar{x}^*_s)$$

$$+G_2H_1\mu_x(\mu_z-\bar{z}^*_s) + G_1H_2\mu_z(\mu_x-\bar{x}^*_s) + G_1\mu_{y2}(\bar{x}^*_s-\mu_x) + \ H_2G_1(\bar{x}^*_s\bar{z}^*_s-\bar{z}^*_s\mu_x).$$

[3.40]

Now consider the second term of [3.39]

$$\sum_s w_i E_2(y_{1i}y_{2i}|z)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T[\ \sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*$$

$$= \sum_s w_i[(H_1\mu_{y2}-2H_2H_1\mu_z-G_2H_1\mu_x+\mu_{y2}H_2-G_1H_2\mu_x)z_i$$

$$+(G_1\mu_{y2}-2G_2G_1\mu_z-G_1H_2\mu_z+\mu_{y1}G_2-G_2H_1\mu_z)x_i$$

$$+H_1H_2z_i^2+H_2G_1x_iz_i+G_2H_1x_iz_i+G_2G_1x_i^2](\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T$$

$$[\ \sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*]$$

$$= \sum_s w_i[Az_i+Bx_i](\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T\ [\ \sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*$$

$$+\ \sum_s w_i(H_1z_i+G_1x_i)(H_2z_i+G_2x_i)[\ \sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*,$$

[3.41]

where

$A=H_1\mu_{y2}-2H_2H_1\mu_z-G_2H_1\mu_x+\mu_{y2}H_2-G_1H_2\mu_x,$

and

$$B=G_1\mu_{y2}-2G_2G_1\mu_z-G_1H_2\mu_z+\mu_{y1}G_2-G_2H_1\mu_z.$$

$$=A[\bar{z}_s^*-\bar{z}_U]+2B\mu_z[\bar{z}_s^*-\bar{z}_U]+B[\bar{z}_s^*-\bar{z}_U]^2+B[S^*_{zzs}-S_{zz}]$$
$$+\sum_s w_i(H_1z_i+G_1x_i)(H_2z_i+G_2x_i)[\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*.$$

[3.42]

since the first component of [3.41] is similar to [3.33] Substituting [3.41] and [3.42] in [3.40] we get

$$E_2(\hat{Q}_{5f}|\underset{\sim}{z},s)$$
$$=\Sigma_{y1y2}+H_1H_2(S^*_{zzs}-\Sigma_{zz})+G_1G_2(S^*_{xxs}-\Sigma_{xx})+G_2H_1(S^*_{xzs}-\Sigma_{xz})$$
$$+G_1H_2(S^*_{zxs}-\Sigma_{zx})+H_2\mu_{y1}(\bar{z}_s^*-\mu_z)+G_2\mu_{y1}(\bar{x}_s^*-\mu_x)$$
$$+H_1\mu_{y2}(\bar{z}_s^*-\mu_z)+H_1H_2[\mu_z^2+\bar{z}_s^{*2}-2\mu_z\bar{z}_s^*)$$
$$+\mu_{y1}\mu_{y2}+G_1G_2[\mu_x^2+\bar{x}_s^{*2}-2\mu_x\bar{x}_s^*)+H_1G_2(\bar{x}_s^*\bar{z}_s^*-\mu_z\bar{x}_s^*)$$
$$+G_2H_1\mu_x(\mu_z-\bar{z}_s^*)+G_1H_2\mu_z(\mu_x-\bar{x}_s^*)+G_1\mu_{y2}(\bar{x}_s^*-\mu_x)+H_2G_1(\bar{x}_s^*\bar{z}_s^*-\bar{z}_s^*\mu_x)$$
$$-A[\bar{z}_s^*-\bar{z}_U]-2B\mu_z[\bar{z}_s^*-\bar{z}_U]-B[\bar{z}_s^*-\bar{z}_U]^2-B[S^*_{zzs}-S_{zz}]$$
$$-\sum_s w_i(H_1z_i+G_1x_i)(H_2z_i+G_2x_i)[\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*.$$

[3.43]

We see from eqn[3.43] that $\hat{Q}_{5f}$ is conditionally biased under the model 2.
Averaging eqn[3.43] over all possible samples we get

$$E_pE_2(\hat{Q}_{5f})$$
$$=\Sigma_{y1y2}+H_1H_2(E_p(S^*_{zzs})-\Sigma_{zz})+G_1G_2(E_p(S^*_{xxs})-\Sigma_{xx})$$
$$+G_2H_1(E_p(S^*_{xzs})-\Sigma_{xz})+G_1H_2(E_p(S^*_{zxs})-\Sigma_{zx})+H_2\mu_{y1}(E_p(\bar{z}_s^*)-\mu_z)$$
$$+G_2\mu_{y1}(E_p(\bar{x}_s^*)-\mu_z)+H_1\mu_{y2}(E_{p2}(\bar{z}_s^*)-\mu_z)+H_1H_2[\mu_z^2$$

$$+E_{p2}(\bar{z}_s^{*2})-2\mu_z E_{p2}(\bar{z}_s^*))+\mu_{y1}\mu_{y2}+G_1G_2[\mu_x^2+E_p(\bar{x}_s^{*2})$$

$$-2\mu_x E_p(\bar{x}_s^*))+H_1G_2(E_p(\bar{x}_s^*\bar{z}_s^*)-\mu_z E_p(\bar{x}_s^*))$$

$$+G_2H_1\mu_x(\mu_z-E_p(\bar{z}_s^*))+G_1H_2\mu_z(\mu_x-E_p(\bar{x}_s^*))$$

$$+G_1\mu_{y2}(E_p(\bar{x}_s^*)-\mu_x)+ H_2G_1(E_p(\bar{x}_s^*\bar{z}_s^*)-E_p(\bar{z}_s^*)\mu_x)$$

$$-A[E_p(\bar{z}_s^*)-\bar{z}_U]-2B\mu_z[E_p(\bar{z}_s^*)-\bar{z}_U]-BE_p[\bar{z}_s^*-\bar{z}_U]^2-B[E_p(S_{zzs}^*) -S_{zz}]$$

$$-E_p(\sum_s w_i(H_1z_i+G_1x_i)(H_2z_i+G_2x_i)[\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*). \quad [3.44]$$

From [3.43] and [3.44] we see that $\hat{Q}_{5f}$ is conditionally and unconditionally biased.However if the weights w_i are Horvitz Thompson type of weights with ratio adjustment then [3.44] becomes

$$E_pE_2(\hat{Q}_{5f})\cong\Sigma_{y1y2}+\mu_{y1}\mu_{y2}$$

$$\cong Q_5.$$

$$\Rightarrow E(\hat{Q}_{5f})\cong Q_5. \quad [3.45]$$

From equations [3.37],[3.38] and [3.43],[3.44] we see that the linear component and the quadratic component of $\underset{\sim}{\hat{Q}_f}$ are conditionally and unconditionally biased under model 2. We can therefore deduce that,

$$E_2(\underset{\sim}{\hat{Q}_f}|z,s)\neq\underset{\sim}{Q}, \quad \text{as } n,N\Rightarrow\infty, \quad [3.46]$$

and

$$E(\underset{\sim}{\hat{Q}_f})\neq\underset{\sim}{Q} \text{ , as } n,N\Rightarrow\infty. \quad [3.47]$$

Appling lemmas 3.1 and 3.3 to eqn [3.47] we get the required result.

The result of theorem 3.5 was a bit surprising,because we expected the Fuller estimator to be robust for the quadratic model 2.The reason why the estimator was found to be biased when the linearity assumption was violated may be because under model 1 we had only the linear and quadratic terms which were accounted for by the estimator,however under model 2 in addition to the linear and quadratic terms we also had quartic terms which were not accounted for in the construction of the estimator.

COROLLARY 3.2

If the weights are Horvitz-Thompson type with ratio adjustment i.e $w_i=\pi_i^{-1}/\sum_s \pi_i^{-1}$ then to $O(n^{-1/2})$ approximation and assuming that N is large,then the Fuller estimator is asymptotically unbiased,i.e

$$E(h(\hat{\underset{\sim}{Q}}_f))=h(\underset{\sim}{Q}) \text{ as } n,N\Rightarrow\infty.$$

Proof

Using eqn[3.38] and [3.45] we can deduce that

$$E(\hat{\underset{\sim}{Q}}_f)=\underset{\sim}{Q}. \text{ as } n,N\Rightarrow\infty$$

Applying lemmas 3.1 and 3.2 we get the required result. It follows from cor 3.2 that provided the weights w_i are of the Horvitz-Thompson type with ratio adjustment,then the Fuller estimator is asymptotically unbiased for any model. We now investigate the asymptotic properties of the Fuller estimator when only the homoscedastic assumption is violated.

3.2.3 ASYMPTOTIC PROPERTIES OF THE FULLER ESTIMATOR UNDER THE LINEAR HETEROSCEDASTIC MODEL(MODEL 3)

We will derive the asymptotic properties of the Fuller estimators when the Homoscedasticity assumption is violated.We call the resulting model,model 3.Under model 3 we assume that

(i) The conditional expectation of $\underset{\sim}{\tilde{y}}$ given $\underset{\sim}{z}$ is a linear function of $\underset{\sim}{z}$.

(ii) The conditional variance covariance matrix of $\tilde{\underset{\sim}{y}}$ given $\underset{\sim}{z}$ is a linear function of $\underset{\sim}{z}$.

(iii) $y_i \perp\!\!\!\perp y_j | z$ for $i \neq j$.

Thus the conditional moments of the survey variable $\underset{\sim}{y}$ given the design variable $\underset{\sim}{z}$ are assumed to obey

$$E_3(\tilde{\underset{\sim}{y}}|\underset{\sim}{z})=\underset{\sim}{m}+\underset{\sim}{H}\underset{\sim}{z},$$

and

$$V_3(\tilde{\underset{\sim}{y}}|\underset{\sim}{z})=W(\underset{\sim}{z})=[W_{jk}(z)],$$
$$=[\underset{\sim}{A}+\underset{\sim}{B}\underset{\sim}{z}].$$

where $\underset{\sim}{m}$ is a px1 constant vector, $\underset{\sim}{H}$ is a pxp constant matrix, $\underset{\sim}{z}$ is a qx1 vector of the realized values of the random variables $\tilde{\underset{\sim}{z}}$, $W(\underset{\sim}{z})$ is a pxp nonsingular matrix, $\underset{\sim}{A}=[A_1,A_2,A_3]$ and $\underset{\sim}{B}=[B_1,B_2,B_3]$ are vectors of constants.

Under model 3 the following relationships holds;

$$\underset{\sim}{m}=\underset{\sim}{\mu}_y-\underset{\sim}{H}\underset{\sim}{\mu}_z,$$

and

$$E_z(W_{jk}(\underset{\sim}{z}))=\underset{\sim}{\Sigma}_{yy}-\underset{\sim}{H}\underset{\sim}{\Sigma}_{zz}\underset{\sim}{H}^T.$$

Substitute $\underset{\sim}{\mu}_w=E_z(W_{jk}(\underset{\sim}{z}))$,

we therefore get

$$\underset{\sim}{\mu}_w=\underset{\sim}{\Sigma}_{yy}-\underset{\sim}{H}\underset{\sim}{\Sigma}_{zz}\underset{\sim}{H}^T. \qquad [3.48]$$

Also $\underset{\sim}{A}=\underset{\sim}{\mu}_w-\underset{\sim}{B}\underset{\sim}{\mu}_z$.

We will consider the case where q=1 and p=2.

Therefore

$$V_3(\tilde{\underset{\sim}{y}}|z)=[W_{jk}(z_i)]=\begin{bmatrix} w_{11}(z_i) & w_{12}(z_i) \\ w_{21}(z_i) & w_{22}(z_i) \end{bmatrix}.$$

$$=\begin{bmatrix} A_1+B_1z_i & A_3+B_3z_i \\ A_3+B_3z_i & A_2+B_2z_i \end{bmatrix}.$$

THEOREM 3.4

Under model 3 to order $O(n^{-1/2})$ approximation and assuming that the population size N is large and $\sum_s w_i=1$, then in general the Fuller estimator $h(\hat{\underset{\sim}{Q}}_f)$ is an asymptotically unbiased estimator of $h(\underset{\sim}{Q})$ i.e

$$E[h(\hat{\underset{\sim}{Q}}_f)]=h(\underset{\sim}{Q}), \text{ as } n,N \Rightarrow\infty.$$

where $h(\hat{\underset{\sim}{Q}}_f)$ denotes the Fuller estimators of the regression coefficients defined in [3.11].

Proof

Since model 3 has a linear structure as in model 1 the asymptotic properties of the linear components of $\hat{\underset{\sim}{Q}}_f$ are the same as those under model 1.We will therefore derive the asymptotic properties of only the quadratic component $\hat{Q}_{5f}$ under model 3.

We consider the fifth component of $\hat{\underset{\sim}{Q}}_f$.Taking conditional expectation of $\hat{Q}_{5f}$ with respect to the model 3 we get

$$E_3(\hat{Q}_{5f}|z,s)=\sum_s w_i E_3(y_{1i}y_{2i}|z)- \sum_s w_i E_3(y_{1i}y_{2i}|z)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T \left[\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T\right]^{-1}\underset{\sim}{\bar{t}}^* \qquad [3.49]$$

Consider the first term of [3.49] we have

$$\sum_s w_i E_3(y_{1i}y_{2i}|z)= \sum_s w_i(\text{cov}_3[(y_{1i}y_{2i}|z)+E_3(y_{1i}|z)E_3(y_{2i}|z)]$$

$$= \sum_s w_i[W_{12}(z_i)+m_1m_2+H_2m_1z_i+H_1m_2z_i+H_1H_2z_i^2]$$

using model 3 assumptions.

$$=W^{*}_{12S}+\sum_{S} w_i[m_1m_2+H_2m_1z_i+H_1m_2z_i+H_1H_2z_i^2] \qquad [3.50]$$

$$\text{where} \quad W^{*}_{12S}=\sum_{S} w_iW_{12}(z_i)$$

eqn[3.50] is identical to [3.21] but with K_{12} replaced by W^{*}_{12S} therefore using result [3.22] we get

$$\sum_{S} w_iE_3(y_{1i}y_{2i}|z) = W^{*}_{12S}+\mu_{w12}-\mu_{w12}-H_2\mu_{y1}\mu_z-H_1\mu_{y2}\mu_z+H_1H_2\mu_z^2$$

$$+H_2(\mu_{y1}-H_1\mu_z)\bar{z}^{*}_s+H_1(\mu_{y2}-H_2\mu_z)\bar{z}^{*}_s+H_1H_2\sum_{S} w_iz_i^2$$

$$+\mu_{y1}\mu_{y2}$$

$$=(W^{*}_{12S}-\mu_{w12})+\Sigma_{y1y2}-H_2\mu_{y1}\mu_z-H_1\mu_{y2}\mu_z+H_1H_2\mu_z^2$$

$$+H_2(\mu_{y1}-H_1\mu_z)\bar{z}^{*}_s+H_1(\mu_{y2}-H_2\mu_z)\bar{z}^{*}_s+H_1H_2\sum_{S} w_iz_i^2$$

$$+\mu_{y1}\mu_{y2}-H_1H_2\Sigma_{zz}, \qquad [3.51]$$

since $\mu_{w12}=\Sigma_{y1y2}-H_1H_2\Sigma_{zz}$ from [3.48].

We now evaluate the second term of eqn[3.49]

$$\sum_{S} w_iE_3(y_{1i}y_{2i}|z)(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^{*})^T[\sum_{S} w_i(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^{*})(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^{*})^T]^{-1}\bar{\underset{\sim}{t}}^{*}$$

$$=\sum_{S} w_i(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^{*})^T(W_{12}(z_i)+H_2(\mu_{y1}-H_1\mu_z)z_i+H_1(\mu_{y2}-H_2\mu_z)z_i+H_1H_2z_i^2)$$

$$[\sum_{S} w_i(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^{*})(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^{*})^T]^{-1}\bar{\underset{\sim}{t}}^{*}]$$

$$=\sum_{S} w_i(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^{*})^T(W_{12}(z_i)[\sum_{S} w_i(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^{*})(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^{*})^T]^{-1}\bar{\underset{\sim}{t}}^{*}]$$

$$+[H_2\mu_{y1}\sum_{S} w_i(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^{*})^T(z_i-\bar{z}_U)+H_1\mu_{y2}\sum_{S} w_i(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^{*})^T(z_i-\bar{z}_U)$$

$$+H_1H_2\sum_{S} w_i(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^{*})^T((z_i-\bar{z}_U)^2-S_{zz})][\sum_{S} w_i(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^{*})(\underset{\sim}{t}_i-\bar{\underset{\sim}{t}}^{*})^T]^{-1}\bar{\underset{\sim}{t}}^{*}.$$

[3.52]

The second term of eqn[3.52] is similar to second term of [3.20], therefore using result [3.23] we get

$$\sum_s w_i E_3(y_{1i}y_{2i}|z)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T[\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*$$

$$= \sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T(W_{12}(z_i)[\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*]$$

$$+H_2\mu_{y1}(\bar{z}_s^*-\bar{z}_U) +H_1\mu_{y2}(\bar{z}_s^*-\bar{z}_U) +H_1H_2\bar{z}_U^2 +H_1H_2\sum_s w_i z_i^2$$

$$-2H_1H_2\bar{z}_s^*\bar{z}_U- H_1H_2S_{zz}$$

$$= (B_3,0)\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T z_i[\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*]$$

$$+H_2\mu_{y1}(\bar{z}_s^*-\bar{z}_U) +H_1\mu_{y2}(\bar{z}_s^*-\bar{z}_U) +H_1H_2\bar{z}_U^2 +H_1H_2\sum_{i=1}^{n} w_i z_i^2$$

$$-2H_1H_2\bar{z}_s^*\bar{z}_U- H_1H_2S_{zz}$$

since $W_{12}(z_i)=A_3+B_3z_i$

$$=B_3(\bar{z}_S^*-\bar{z}_U)+H_2\mu_{y1}(\bar{z}_S^*-\bar{z}_U)+H_1\mu_{y2}(\bar{z}_S^*-\bar{z}_U)+H_1H_2\sum_s w_i(z_i-\bar{z}_U)^2-H_1H_2S_{zz}.$$

[3.53]

Substitute eqn[3.53] and [3.51] in [3.49] we get

$$E_3(\hat{Q}_{5f}|z,s\quad) =(W_{12s}^*-\mu_{w12})+\Sigma_{y1y2} -H_2\mu_{y1}\mu_z-H_1\mu_{y2}\mu_z +H_1H_2\mu_z^2$$

$$+H_2(\mu_{y1}-H_1\mu_z)\bar{z}_s^* +H_1(\mu_{y2}-H_2\mu_z)\bar{z}_s^* +H_1H_2\sum_s w_i z_i^2$$

$$+\mu_{y1}\mu_{y2}-B_3(\bar{z}_S^*-\bar{z}_U)-H_2\mu_{y1}(\bar{z}_S^*-\bar{z}_U)-H_1\mu_{y2}(\bar{z}_S^*-\bar{z}_U)$$

$$-\quad H_1H_2\sum_{i=1}^{n} w_i(z_i-\bar{z}_U)^2+H_1H_2S_{zz}.$$

$$\cong(W_{12s}^*-\mu_{w12})+\Sigma_{y1y2}+\mu_{y1}\mu_{y2} +H_1H_2(S_{zz}-\Sigma_{zz})$$

$$+H_2\mu_{y1}(\bar{z}_U-\mu_z)+H_2H_1(\bar{z}_U^2-\mu^2_z)-B_3(\bar{z}_S^*-\bar{z}_U)$$

$$+H_1\mu_{y2}(\bar{z}_U-\mu_z)+2H_2H_1\bar{z}_S^*(\bar{z}_U-\mu_z)$$

$$=\Sigma_{y1y2}+\mu_{y1}\mu_{y2} +H_1H_2(S_{zz}-\Sigma_{zz})$$

$$+H_2\mu_{y1}(\bar{z}_U-\mu_z)+H_2H_1(\bar{z}_U^2-\mu_z^2)-B_3(\bar{z}_U-\mu_z)$$

$$+H_1\mu_{y2}(\bar{z}_U-\mu_z)+2H_2H_1\bar{z}_s^*(\bar{z}_U-\mu_z)$$

$$\text{since } W_{12s}^*=\mu_{w12}+B_3(\bar{z}_s^*-\mu_z)$$

$$=\Sigma_{y1y2}+\mu_{y1}\mu_{y2}+O_p(N^{-1/2})$$

$$\cong Q_5, \quad \text{since N is large.} \qquad [3.54]$$

From [3.54] we see that $\hat{Q}_{5f}$ is approximately conditionally unbiased and hence approximately unconditionally unbiased estimator of Q_5.

We can deduce from eqns [3.18] and [3.54],that the linear and quadratic components of $\hat{\underset{\sim}{Q}}_f$ are approximately conditionally and unconditionally unbiased under model 3,i.e

$$E_3(\hat{\underset{\sim}{Q}}_f|z,s)=\underset{\sim}{Q}, \quad \text{as } n,N\Rightarrow\infty. \qquad [3.55]$$

Averaging over all possible samples we get

$$E_pE_3(\hat{\underset{\sim}{Q}}_f)=\underset{\sim}{Q} \quad \Rightarrow E(\hat{\underset{\sim}{Q}}_f)=\underset{\sim}{Q}, \quad \text{as } n,N\Rightarrow\infty. \qquad [3.56]$$

Applying lemmas 3.1 and 3.2 to eqn [3.56] we get the required result.

This theorem shows that unlike the maximum likelihood estimator which was found to be asymptotically biased when the homoscedasticity assumption was violated,the Fuller estimator retained its asymptotic unbiasedness properties under model 3. Thus the Fuller estimator is robust to the violations of the Homoscedastic assumptions.

3.3 COMPARISION OF VARIANCE OF THE FULLER AND MAXIMUM LIKELIHOOD REGRESSION ESTIMATORS

In theorem 3.1 we proved that the Fuller estimator is asymptotically unbiased estimator under model 1.Since the maximum likelihood estimator is also asymptotically unbiased under model 1 we need another criterion to choose between the two estimators.In the next section we will derive the conditional variances of the two estimators and base our choice of the prefered estimator on its variance.

3.3.1 VARIANCES OF THE FULLER ESTIMATORS

In this section we will derive the variances of the Fuller estimator studied in the previous sections.

LEMMA 3.3

To $O_p(n^{-1})$ approximation

$$h(\hat{\underset{\sim}{Q}}_f)-h(\underset{\sim}{Q})=\underset{\sim}{M}_{22}^{-1}\ \hat{\underset{\sim}{D}}_f,$$

where $\hat{\underset{\sim}{D}}_f=\hat{\underset{\sim}{M}}_{21}-\hat{\underset{\sim}{M}}_{22}h(\underset{\sim}{Q})$, [3.57]

$h(\hat{\underset{\sim}{Q}}_f),\hat{\underset{\sim}{M}}_{21},\hat{\underset{\sim}{M}}_{22}$ and $h(\underset{\sim}{Q}),\underset{\sim}{M}_{21},\underset{\sim}{M}_{22}$ are defined in [3.11],[3.12],[3.13] and [3.1],[3.2],[3.3] respectively.

Proof

Now $\hat{\underset{\sim}{D}}_f$ is an estimator of $\underset{\sim}{D}$ where $\underset{\sim}{D}$ is

$$\underset{\sim}{D}=\underset{\sim}{M}_{21}-\underset{\sim}{M}_{22}h(\underset{\sim}{Q})$$

$$=\underset{\sim}{M}_{22}(\underset{\sim}{M}_{22}^{-1}\underset{\sim}{M}_{21}-h(\underset{\sim}{Q}))$$

$$=\underset{\sim}{0},\qquad \text{since } h(\underset{\sim}{Q})=\underset{\sim}{M}_{22}^{-1}\underset{\sim}{M}_{21} \text{ from [3.3].} \qquad [3.58]$$

Now

$$\hat{\underset{\sim}{D}}_f = \underset{\sim}{D} + O_p(n^{-1/2})$$

$$= O_p(n^{-1/2}), \quad \text{using [3.58].} \qquad [3.59]$$

Noting that

$$\hat{\underset{\sim}{M}}_{22}^{-1} = \underset{\sim}{M}_{22}^{-1} + O_p(n^{-1/2}), \qquad [3.60]$$

and

$$h(\hat{\underset{\sim}{Q}}_f) = \hat{\underset{\sim}{M}}_{22}^{-1}\hat{\underset{\sim}{M}}_{21}. \qquad [3.61]$$

We have

$$h(\hat{\underset{\sim}{Q}}_f) - h(\underset{\sim}{Q}) = \hat{\underset{\sim}{M}}_{22}^{-1}(\hat{\underset{\sim}{M}}_{21} - \hat{\underset{\sim}{M}}_{22}h(\underset{\sim}{Q})), \quad \text{using [3.61]}$$

$$= \hat{\underset{\sim}{M}}_{22}^{-1}\hat{\underset{\sim}{D}}_f \quad \text{since } \hat{\underset{\sim}{D}}_f = \hat{\underset{\sim}{M}}_{21} - \hat{\underset{\sim}{M}}_{22}h(\underset{\sim}{Q})$$

$$= [\underset{\sim}{M}_{22}^{-1} + O_p(n^{-1/2})]\hat{\underset{\sim}{D}}_f \quad \text{using [3.60]}$$

$$= \underset{\sim}{M}_{22}^{-1}\hat{\underset{\sim}{D}}_f + O_p(n^{-1/2})O_p(n^{-1/2})$$

$$= \underset{\sim}{M}_{22}^{-1}\hat{\underset{\sim}{D}}_f + O_p(n^{-1}), \qquad [3.62]$$

which is the required result.

LEMMA 3.4

For $\hat{\underset{\sim}{D}}_f$ defined in [3.57] we may write

$$\hat{\underset{\sim}{D}}_f = \underset{\sim}{A}\hat{\underset{\sim}{Q}}_f,$$

$$\text{where } \underset{\sim}{A} = \begin{bmatrix} -A_{12} & -B_{12} & 0 & 1 & 0 \\ 0 & -A_{12} & -B_{12} & 0 & 1 \end{bmatrix},$$

and $\hat{\underset{\sim}{Q}}_f$ is as defined in [3.7].

Proof

From [3.57] we have

$$\hat{\underset{\sim}{D}}_f = \hat{\underset{\sim}{M}}_{21} - \hat{\underset{\sim}{M}}_{22}h(\underset{\sim}{Q})$$

$$=\begin{bmatrix} \hat{Q}_{4,f} \\ \hat{Q}_{5,f} \end{bmatrix} - \begin{bmatrix} \hat{Q}_{1,f} & \hat{Q}_{2,f} \\ \hat{Q}_{2,f} & \hat{Q}_{3,f} \end{bmatrix} \begin{bmatrix} A_{12} \\ B_{12} \end{bmatrix}$$

Using [3.12],[3.13] and [3.1]

$$=\begin{bmatrix} \hat{Q}_{4,f} - \hat{Q}_{1,f} A_{12} - \hat{Q}_{2,f} B_{12} \\ \hat{Q}_{5,f} - \hat{Q}_{2,f} A_{12} - \hat{Q}_{3,f} B_{12} \end{bmatrix}$$

$$=\begin{bmatrix} -A_{12} & -B_{12} & 0 & 1 & 0 \\ 0 & -A_{12} & -B_{12} & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{Q}_{1,f} \\ \hat{Q}_{2,f} \\ \hat{Q}_{3,f} \\ \hat{Q}_{4,f} \\ \hat{Q}_{5,f} \end{bmatrix}$$

$=\underset{\sim}{A}\hat{\underset{\sim}{Q}}_f$ as required.

LEMMA 3.5

Let $\underset{\sim}{d}_i=\underset{\sim}{A}\underset{\sim}{q}_i$ where $\underset{\sim}{A}$ is as defined in lemma 3.4 and q_i is defined in [3.4],then we may write

$$\hat{\underset{\sim}{D}}_f= \sum_s \lambda_i \underset{\sim}{d}_i,$$

where $\lambda_i=w_i- w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T[\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*$,

and $\underset{\sim}{\bar{t}}_i,\underset{\sim}{\bar{t}}^*$ are as defined in [3.6].

Proof

From [3.6] we have

$$\hat{\underset{\sim}{Q}}_f=\sum_s w_i \underset{\sim}{q}_i - \sum_s w_i \underset{\sim}{q}_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T[\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*$$

$$= \sum_s \lambda_i \underset{\sim}{q}_i . \qquad [3.63]$$

premultiply both sides of [3.63] by matrix $\underset{\sim}{A}$ defined in lemma 3.4 we get

$$\underset{\sim}{A}\hat{\underset{\sim}{Q}}_f = \sum_s \lambda_i \underset{\sim}{A}\underset{\sim}{q}_i , \qquad [3.64]$$

since $\underset{\sim}{d}_i = Aq_i$ and $\hat{\underset{\sim}{D}}_f = \underset{\sim}{A}\hat{\underset{\sim}{Q}}_f$ from Lemma 3.4.

[3.64] becomes

$$\hat{\underset{\sim}{D}}_f = \sum_s \lambda_i \underset{\sim}{d}_i .$$

which is the required result.

THEOREM 3.5

The conditional variance of the Fuller regression estimator $h(\hat{\underset{\sim}{Q}}_f)$ under model ε to $O(n^{-1})$ approximation and assuming N is large is given by

$$V_\varepsilon[h(\hat{\underset{\sim}{Q}}_f)|z,s] = \underset{\sim}{N}_{22}^{-1} \sum_s \lambda_i^2 \, V_\varepsilon(\underset{\sim}{d}_i|z)\underset{\sim}{N}_{22}^{-1} ,$$

where d_i and λ_i are defined in Lemma 3.4 and 3.5 respectively and

$$\underset{\sim}{N}_{22} = \begin{bmatrix} 1 & \mu_{y2} \\ \mu_{y2} & \mu_{y2}^2 + \Sigma_{y2y2} \end{bmatrix} ,$$

$V_\varepsilon(.)$ denotes the conditional variance under model ε.

Proof

From Lemma 3.3 we have

$$h(\hat{\underset{\sim}{Q}}_f) - h(\underset{\sim}{Q}) = \underset{\sim}{M}_{22}^{-1} \hat{\underset{\sim}{D}}_f + o_p(n^{-1}), \qquad [3.65]$$

Now $\underset{\sim}{M}_{22} = \underset{\sim}{N}_{22} + O_p(N^{-1/2})$,

Therefore [3.65] becomes

$$h(\hat{\mathbf{Q}}_f)-h(\mathbf{Q})=\mathbf{N}_{22}^{-1}\,\hat{\mathbf{D}}_f + O_p(n^{-1}),\ n/N\Rightarrow\text{constant}. \qquad [3.66]$$

Taking conditional variance of [3.66] we get the asymptotic variance of $h(\hat{\mathbf{Q}}_f)$ as

$$V_\varepsilon[h(\hat{\mathbf{Q}}_f)\,|\,z,s]=\mathbf{N}_{22}^{-1}\ V_\varepsilon(\hat{\mathbf{D}}_f\,|\,z,s)\mathbf{N}_{22}^{-1}$$

$$=\mathbf{N}_{22}^{-1}V_\varepsilon[\ \textstyle\sum_s\lambda_i\ \mathbf{d}_i\,|\,z]\mathbf{N}_{22}^{-1}$$

using Lemma 3.5

$$=\mathbf{N}_{22}\ \textstyle\sum_s\lambda_i^2\ V_\varepsilon(\mathbf{d}_i\,|\,z)\mathbf{N}_{22}^{-1} \qquad [3.67]$$

since $\mathbf{d}_i$ are mutually independent.

Hence the result.

Now model 1 defined in chapter1 can be reformulated as follows;

$$\left.\begin{aligned} y_{1i}&=\mu_{y1}+\beta_{1z}(z_i-\mu_z)+e_{1i},\\ y_{2i}&=\mu_{y2}+\beta_{2z}(z_i-\mu_z)+e_{2i}\ , \end{aligned}\right\} \qquad [3.68]$$

$$e_{1i}=\beta_{12.z}e_{2i}+\eta_{1i},$$

where

$$E_1(e_{2i}|z_i)=E_1(\eta_{1i}|y_{2i},z_i)=0,$$
$$E_1(e_{2i}^2|z_i)=K_2\ ,\quad E_1(\eta_{1i}^2|y_{2i},z_i)=\sigma_{1.2z}^2.$$

we also assume that the residuals pertaining to different units are conditionally independent.

Reparameterizing [3.68] we get

$$y_{1i}=\alpha_{1.2z}+\beta_{1z.2}(z_i-\mu_z)+\beta_{12.z}y_{2i}+\eta_{1i}, \qquad [3.69]$$

where

$$\alpha_{1.2z}=\mu_{y1}-\beta_{12.z}\mu_z\ ,$$

and

$$\beta_{1z.2}=\beta_{1z}-\beta_{12.z}\beta_{2z},$$

In order to be able to evaluate [3.67] we assume in addition to model 1 assumptions call this model 4 that

(i) $y_{1i} \perp\!\!\!\perp y_{2i} | z.$

(ii) The conditional third and fourth moments of y_{2i} given z_i are independent of z.i.e Under model 4 we have

$$\left.\begin{aligned} E_4(e_{2i}^3|z_i)&=\mu_{3z} ,\\ E_4(e_{2i}^4|z_i)&=\mu_{4z} ,\end{aligned}\right\} \qquad [3.70]$$

where μ_{3z} and μ_{4z} are assumed constant.

CORR 3.1

Under model 4 defined in [3.70] to $O(n^{-1})$ approximation and assuming that N is large then the conditional variance of the Fuller regression estimator is given by;

$$V_4(h(\hat{\underset{\sim}{Q}}_f)|z,s)=N_{22}^{-1} \sum_s \lambda_i^2 \begin{bmatrix} 1 & \\ m_2+H_2z_i & K_2+(m_2+H_2z_i)^2 \end{bmatrix} K_1N_{22}^{-1}$$

where N_{22}, λ_i are defined in theorem 3.7 and K_1, K_2, m_2, H_2 are as defined in model 1.

Proof

From Lemma 3.4 we have

$$\underset{\sim}{d}_i=\begin{bmatrix} y_{1i}-A_{12}-B_{12}y_{2i} \\ y_{2i}(y_{1i}-A_{12}-B_{12}y_{2i}) \end{bmatrix} = \begin{bmatrix} e_{1i} \\ y_{2i}e_{1i} \end{bmatrix}.$$

Using this value of $\underset{\sim}{d}_i$ in theorem 3.7 we get

$$V_4(h(\hat{Q}_f)|z,s)=N_{22}^{-1}\sum_s \lambda_i^2 \begin{bmatrix} V_4(e_{1i}|z) & \\ Cov_4(e_{1i},y_{2i}e_{1i}|z) & V_4(y_{2i}e_{1i}|z) \end{bmatrix} N_{22}^{-1}$$

[3.71]

Now $V_4(e_{1i}|z)=V_4(E_4(e_{1i}|z,y_{2i}))|z)+E_4(V_4(e_{1i}|z,y_{2i}))|z)$

$=E_4(V_4(e_{1i}|z,y_{2i}))|z)$, since from model 4

$E_4((e_{1i}|z,y_{2i})|z)=0.$

$=E_4(K_1|z)=K_1.$ [3.72]

$E_4(y_{2i}e_{1i}|z)=E_4(y_{2i}E_4(e_{1i}|z,y_{2i}))|z)$

$=0.$ [3.73]

$V_4(y_{2i}e_{1i}|z)=V_4(y_{2i}E_4(e_{1i}|z,y_{2i}))|z)+E_4(y_{2i}^2V_4(e_{1i}|z,y_{2i}))|z)$

$=E_4(y_{2i}^2K_1)|z)$ using model 4

$=K_1E_4(y_{2i}^2|z)$

$=K_1[K_2+(m_2+H_2z_i)^2].$ [3.74]

and lastly

$Cov_4(y_{2i}e_{1i},e_{1i}|z)=E_4(Cov_4(y_{2i}e_{1i},e_{1i}|z,y_{2i}))|z)$

$+Cov_4(E_4(y_{2i}e_{1i}|z,y_{2i}),E_4(e_{1i}|z,y_{2i}))|z))$

$=E_4(y_{2i}K_1|z)$

$=K_1(m_2+H_2z_i).$ [3.75]

substituting [3.75],[3.74],[3.73] and [3.72] in [3.71] we get the required result.

COROLLARY 3.2

Under model 4 to $O(n^{-1})$ approximation and assuming N is large the conditional variance of the Fuller regression estimator of the slope given z and s is given by;

$$V_4(\hat{B}_{12,f}|z,s)=\frac{K_1}{\Sigma^2_{y2y2}}[(K_2+m_2^2+\mu_{y2}^2-2m_2\mu_{y2})P_0+2H_2(m_2-\mu_{y2})P_1+H_2^2P_2],$$

where

$$P_j=\bar{z}_s^{*(j)}-2\underset{\sim}{S}_{tz}^{*(j)}\underset{\sim}{S}_{tt}^{*-1}\underset{\sim}{\bar{t}}^{*}+\underset{\sim}{\bar{t}}^{*T}[\underset{\sim}{S}_{tt}^{*-1}]^T\underset{\sim}{S}_{ttz}^{*(j)}\underset{\sim}{S}_{tt}^{*-1}\underset{\sim}{\bar{t}}^{*},$$

$$\bar{z}_s^{*(j)}=\sum_s w_i^2 z_i^j,$$

$$\underset{\sim}{S}_{tz}^{*(j)}=\sum_s w_i^2(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)z_i^j,$$

and

$$\underset{\sim}{S}_{ttz}^{*(j)}=\sum_s w_i^2(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T z_i^j.$$

where $j=0,1,2$

Proof

From theorem 3.6 under model 4 we have

$$V_4(h(\underset{\sim}{\hat{Q}}_f|z,s)\cong K_1\sum_s\lambda_i^2\begin{bmatrix}1 & \\ \mu_{y2} & T\end{bmatrix}^{-1}\begin{bmatrix}1 & \\ A_2 & A_3\end{bmatrix}\begin{bmatrix}1 & \\ \mu_{y2} & T\end{bmatrix}^{-1},$$

where

$T=\mu_{y2}^2+\Sigma_{y2y2}$,

$A_2=m_2+H_2z_i$,

$A_3=K_2+(m_2+H_2z_i)^2$,

$$=\frac{K_1}{\Sigma^2_{y2y2}}\sum_s\lambda_i^2\begin{bmatrix}T & \\ -\mu_{y2} & 1\end{bmatrix}\begin{bmatrix}1 & \\ A_2 & A_3\end{bmatrix}\begin{bmatrix}T & \\ -\mu_{y2} & 1\end{bmatrix}$$

$$=\frac{K_1}{\Sigma_{y2y2}^2}\sum_s \lambda_i^2\begin{bmatrix} T(T-A_2\mu_{y2}-\mu_{y2}(A_2T-\mu_{y2}A_3) & \\ (A_2T-\mu_{y2}A_3)-\mu_{y2}(T-A_2\mu_{y2}) & (A_3-A_2\mu_{y2})-\mu_{y2}(A_2-\mu_{y2})\end{bmatrix}. \quad [3.76]$$

From [3.76] we have

$$V_4(\hat{B}_{12,f}|z,s)\cong\frac{K_1}{\Sigma_{y2y2}^2}\sum_s \lambda_i^2\ [A_3+\mu_{y2}-2A_2\mu_{y2}] \quad [3.77]$$

Substitute values of A_2 and A_3 in [3.77] we get

$$V_4(\hat{B}_{12,f}|z,s)=\frac{K_1}{\Sigma_{y2y2}^2}\sum_s \lambda_i^2[K_2+(m_2+H_2z_i)^2+\mu_{y2}^2-2(m_2+H_2z_i)\mu_{y2}]$$

$$=\frac{K_1}{\Sigma_{y2y2}^2}[\sum_s \lambda_i^2(K_2+m_2^2+\mu_{y2}-2m_2\mu_{y2})+\sum_s \lambda_i^2 z_i 2H_2(m_2-\mu_{y2})$$

$$+H_2^2\sum_s \lambda_i^2 z_i^2]. \quad [3.78]$$

Now from lemma 3.5

$$\lambda_i=w_i-w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T[\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*$$

$$=w_i-w_i g_i, \quad [3.79]$$

where $\underset{\sim}{g}_i=(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T[\sum_s w_i(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)(\underset{\sim}{t}_i-\underset{\sim}{\bar{t}}^*)^T]^{-1}\underset{\sim}{\bar{t}}^*.$

Therefore substituting the value of λ_i below we get;

$$\sum_s \lambda_i^2 z_i^j=\sum_s (w_i-w_i\underset{\sim}{g}_i)(w_i-w_i\underset{\sim}{g}_i)^T z_i^j, \quad \text{where } j=0,1,2.$$

$$=\sum_s w_i^2 z_i^j-2\sum_s w_i^2\underset{\sim}{g}_i z_i^j+\sum_s w_i^2\underset{\sim}{g}_i\underset{\sim}{g}_i^T z_i^j. \quad [3.80]$$

From [3.80] consider the first term

$$\sum_s w_i^2 z_i^j=\bar{z}^*(j). \qquad j=0,1,2 \quad [3.81]$$

$$\sum_s w_i^2 g_i z_i^j = S_{tz}^{*(j)} S_{tt}^{*-1} \bar{t}^*, \qquad [3.82]$$

where

$$S_{tz}^{*(j)} = \sum_s w_i^2 (t_i - \bar{t}^*) z_i^j \; .$$

and lastly

$$\sum_s w_i^2 g_i g_i^T z_i^j = \bar{t}^{*T} [S_{tt}^{*-1}]^T S_{ttz}^{*(j)} S_{tt}^{*-1} \bar{t}^*, \qquad [3.83]$$

where

$$S_{ttz}^{*(j)} = \sum_s w_i^2 (t_i - \bar{t}^*)(t_i - \bar{t}^*)^T z_i^j \; .$$

Substituting [3.83],[3.82] and [3.81] in [3.80] and substituting the result in [3.78] we get the required result. In order to study the efficiency properties of the Fuller estimator in comparison with the maximum likelihood estimator we derive below the conditional variance of the maximum likelihood estimator under model 4.

THEOREM 3.6

Under model 4 to the $O(n^{-1})$ approximation for large N the conditional variance of the maximum likelihood estimator is given by;

$$V_4(\hat{B}_{12,ml}|z,s)=$$

$$K_1\left[\frac{(K_2+H_2^2S_{zzs})+2D_zH_2^2\,S_{zzs}+D_z^2H_2^2\,S_{zzs}}{n(K_2+H_2^2S_{zzs})[2D_zH_2^2\,S_{zzs}+(K_2+H_2^2S_{zzs})]}\right]+$$

$$\beta_{1z.2}^2\left[\frac{H_2S_{zz}}{(K_2+H_2^2S_{zzs})\,[K_2+H_2^2S_{zzs}+2(S_{zz}-S_{zzs})H_2^2]}\right.$$

$$\left.-\frac{H_2^2S_{zz}^2}{S_{y2y2}^2}\right],$$

where

$$D_z=S_{zz}/S_{zzs}-1\;,$$
$$S_{y2zs}=n^{-1}\sum_{i\in s}(y_{2i}-\bar{y}_{2s})(z_i-\bar{z}_s),$$

$$S_{y2y2s}=n^{-1}\sum_{i\in s}(y_{2i}-\bar{y}_{2s})^2,$$
$$S_{zzs}=n^{-1}\sum_{i\in s}(z_i-\bar{z}_s)^2 \quad\text{and}\quad S_{zz}=N^{-1}\sum_{i\in U}(z_i-\bar{z}_s)^2.$$

Proof

Holmes(1987) proved that under model 1 the conditional variance of the Maximum likelihood estimator is given by

$$V_1(\hat{B}_{12,ml}|z,s)=K_1E_1\left[\frac{S_{y2y2s}+\dfrac{2S_{y2zs}^2D_z}{S_{zzs}}+\dfrac{S_{y2zs}^2D_z^2}{S_{zzs}}}{n\left[S_{y2y2s}+\dfrac{2S_{y2zs}^2D_z}{S_{zzs}}\right]^2}\;\middle|\; z\right]$$

$$+V_1[(\beta_{12.z}+\beta_{1z.2}\beta_{z2,ml})|z), \qquad [3.84]$$

where

$$D_z=S_{zz}/S_{zzs}-1,$$

$$\tilde{S}_{y2zs}=n^{-1}\sum_{i\in s}(y_{2i}-\bar{y}_{2s})(z_i-\bar{z}_s),$$

$$S_{y2y2s}=n^{-1}\sum_{i\in s}(y_{2i}-\bar{y}_{2s})^2,$$

$$S_{zzs}=n^{-1}\sum_{i\in s}(z_i-\bar{z}_s)^2,$$

and

$$\beta_{z2,ml} = \frac{\dfrac{S_{zz}}{S_{zzs}}\, S_{y2zs}}{S_{y2y2s} + \dfrac{[S_{zz}-S_{zzs}]}{S_{zzs}^2}\, S_{y2zs}^2}$$

We will go a step further than Holmes(1987) and evaluate this expression [3.84] under a more restrictive model 4. Now consider the first term of [3.84] we have

$$E_4\left[\left.\frac{S_{y2y2s}+\dfrac{2S_{y2zs}^2 D_z}{S_{zzs}} + \dfrac{S_{y2zs}^2}{S_{zzs}} D_z^2}{n\left[S_{y2y2s}+\dfrac{2S_{y2zs}^2 D_z}{S_{zzs}}\right]^2}\,\right|\, z\right] \cong$$

$$\left[\frac{E_4(S_{y2y2s}|z) + \dfrac{2E_4(S_{y2zs}^2|z)}{S_{zzs}}\, D_z + \dfrac{E_4(S_{y2zs}^2|z)D_z^2}{S_{zzs}}}{n\left[E_4(S_{y2y2s}^2|z)+\dfrac{2D_z E_4(S_{y2zs}S_{y2y2s}|z)}{S_{zzs}} + \dfrac{E_4(S_{y2zs}^4|z)D_z^2}{S_{zzs}^2}\right]}\right]$$

[3.85]

Evaluating each term in [3.85] separately we have

$$E_4(S_{y2zs}|z)=n^{-1}E_4[\sum_s (z_i-\bar{z}_s)(y_{2i}-\bar{y}_{2s})|z]$$

$$=H_2S_{zzs}, \text{ using model 4.} \qquad [3.86]$$

$$E_4(S^2_{y2zs}|z)=n^{-2}\sum_{i\in s}(z_i-\bar{z}_s)^2E_4((y_{2i}-\bar{y}_{2s})^2|z)$$

$$+n^{-2}\sum_{i=j\in s}\sum(z_i-\bar{z}_s)(z_j-\bar{z}_s)E_4((y_{2i}-\bar{y}_{2s})(y_{2j}-\bar{y}_{2s})|z)$$

$$= n^{-2}\sum_{i\in s}(z_i-\bar{z}_s)^2E_4((y_{2i}-\bar{y}_{2s})^2|z)$$

$$+n^{-2}\sum_{i\in s}(z_i-\bar{z}_s)E_4((y_{2i}-\bar{y}_{2s}))|z)\sum_{j\in s}(z_j-\bar{z}_s)E_4((y_{2j}-\bar{y}_{2s}))|z)$$

since $y_{2i}\perp\!\!\!\perp y_{2j}|z$ for $i\neq j$

$$= n^{-2}\sum_{i\in s}(z_i-\bar{z}_s)^2[E_4((y_{2i}-\bar{y}_{2s})|z)^2 +V_4((y_{2i}-\bar{y}_{2s})|z)]$$

$$+n^{-2}\sum_{i\in s}(z_i-\bar{z}_s)E_4((y_{2i}-\bar{y}_{2s}))|z)\sum_{j\in s}(z_j-\bar{z}_s)E_4((y_{2j}-\bar{y}_{2s}))|z)$$

using model 4 assumptions we get

$$E_4(S^2_{y2zs}|z)=H_2^2\mu_{4z}/n +K_2S_{zzs}(1-1/n)/n +H_2^2S^2_{zzs} \qquad [3.87]$$

where μ_{4z} is the fourth moment of z.

$\cong H_2^2S^2_{zzs}$,ignoring terms to the $O(n^{-1})$.

$$E_4(S_{y2y2s}|z)=n^{-1}E_4[\sum_s(y_{2i}-\bar{y}_{2s})^2|z]$$

$\cong K_2+H_2^2S_{zzs}$,using model4 assumptions and ignoring terms to $O(n^{-1})$. [3.88]

$$E_4(S^2_{y2y2s}|z)=n^{-2}\sum_{i\in s} E_4((y_{2i}-\bar{y}_{2s})^4|z)$$

$$+n^{-2}\sum_{i\in s} E_4((y_{2i}-\bar{y}_{2s})^2)|z)\sum_{j\in s} E_4((y_{2j}-\bar{y}_{2s})^2)|z)$$

since $y_{2i}⫫y_{2j}|z$ for $i\neq j$

$\cong(K_2+H_2^2S_{zzs})^2$, using model 4 assumptions and ignoring terms to $O(n^{-1})$. [3.89]

$$E_4(S_{y2y2s}S^2_{y2zs}|z)= E_4(S_{y2y2s}|z)\ E_4(S^2_{y2zs}|z)$$

$$=(K_2+H_2^2S_{zzs})H_2^2S^2_{zzs},$$

using [3.87] and [3.89].

and lastly

$$E_4(S^4_{y2zs}|z)=n^{-4}\sum_{i\in s} (z_i-\bar{z}_s)^4E_4((y_{2i}-\bar{y}_{2s})^4|z)$$

$$+n^{-4}\sum_{i\in s} (z_i-\bar{z}_s)^2E_4((y_{2i}-\bar{y}_{2s})^2)|z)\sum_{j\in s} (z_j-\bar{z}_s)^2E_4((y_{2j}-\bar{y}_{2s})^2)|z).$$

since $y_{2i}⫫y_{2j}|z$, for $i\neq j$.

Evaluating this under model 4 and ignoring terms to $O(n^{-1})$ we get

$E_4(S^4_{y2zs}|z)\cong 0$. [3.90]

substituting [3.87]-[3.90] in [3.85] we get

$$\frac{\left[E_4(S_{y2y2s}|z)+\frac{2E_4(S^2_{y2zs}|z)}{S_{zzs}}\frac{D_z}{} + \frac{E_4(S^2_{y2zs}|z)D^2_z}{S_{zzs}}\right]}{n\left[E_4(S^2_{y2y2s}|z)+\frac{2D_z E_4(S^2_{y2zs}S_{y2y2s}|z)}{S_{zzs}} + \frac{E_4(S^4_{y2zs}|z)D^2_z}{S^2_{zzs}}\right]}$$

$$= \frac{(K_2+H^2_2 S_{zzs})+2D_z H^2_2\ S_{zzs} + D^2_z H^2_2\ S_{zzs}}{n(K_2+H^2_2 S_{zzs})[2D_z H^2_2\ S_{zzs} + (K_2+H^2_2 S_{zzs})]} \quad . \qquad [3.91]$$

Now consider the second term of [3.85] we have

$$V_4[(\beta_{12.z}+\beta_{1z.2}\beta_{z2,ml})|z)=\beta^2_{1z.2}V_4(\beta_{z2,ml}|z)$$

$$=\beta^2_{1z.2}[E_4(\beta^2_{z2,ml}|z)-E^2_4(\beta_{z2,ml}|z)]. \qquad [3.92]$$

Consider the first term of [3.92] we have

$$E_4(\beta^2_{z2,ml}|z)\cong \frac{\frac{S^2_{zz}}{S^2_{zzs}}E_4(S^2_{y2zs}|z)}{E_4(S^2_{y2y2s}|z)+2\frac{[S_{zz}-S_{zzs}]}{S^2_{zzs}}E_4(S_{y2y2s}S^2_{y2zs}|z) - \frac{[S_{zz}-S_{zzs}]^2}{S^4_{zzs}}E_4(S^4_{y2zs}|z)}. \qquad [3.93]$$

Substituting [3.87]-[3.90] in [3.93] we get

$$E_4(\beta^2_{z2,ml}|z) = \frac{H_2^2 S_{zz}^2}{(K_2+H_2^2 S_{zzs})[K_2+H_2^2 S_{zzs}+2(S_{zz}-S_{zzs})H_2^2]}$$

[3.94]

We now consider the second term of [3.92] we have

$$E_4(\beta_{z2,ml}|z) \cong \frac{\frac{S_{zz}}{S_{zzs}} E_4(S_{y2zs}|z)}{E_4(S_{y2y2s}|z) + \frac{[S_{zz}-S_{zzs}]}{S^2_{zzs}} E_4(S^2_{y2zs}|z)} . \qquad [3.95]$$

Subtituting [3.86] ,[3.88] and [3.89] in [3.95] we get

$$E_4(\beta_{z2,ml}|z) = \frac{H_2 S_{zz}}{K_2+H_2^2 S_{zzs} + H_2^2(S_{zz}-S_{zzs})} = \frac{H_2 S_{zz}}{S_{y2y2}} .$$

subtituting the value of K_2 from [1.12]
[3.96]

Substituting [3.96] and [3.94] in [3.92] we get

$$V_4[(\beta_{12.z} + \beta_{1z.2}\beta_{z2,ml})|z) =$$

$$\beta^2_{1z.2} \left[\frac{H_2^2 S_{zz}^2}{(K_2+H_2^2 S_{zzs})[K_2+H_2^2 S_{zzs}+2(S_{zz}-S_{zzs})H_2^2]} + \frac{H_2^2 S_{zz}^2}{S^2_{y2y2}}\right].$$

[3.97]

Substituting [3.97] and [3.91] in [3.84] we get the required result.

We see that the conditional variance expressions of the Fuller estimator and the Maximum likelihood estimator under model 4 given in corr 3.2 and theorem 3.6 are too complicated and cannot allow a simple comparison.We therefore make further assumptions of the underlying linear model.We assume that the independent variable y_2 is independent of the design variable z.i.e $H_2=0$. Making this substitution in Theorem 3.7 we get the conditional variance of the maximum likelihood estimator

$$V_4(\hat{\beta}_{12,ml}|s,z) \cong \frac{S_{y1y2}}{nS_{y2y2}} . \qquad [3.98]$$

and similarly substituting $H_2=0$ in corr 3.2 and for comparison purposes since the maximum likelihood estimator assumes constant weights we will also assume constant weights $w_i=1/n$ for the Fuller estimator.Therefore the conditional variance of the Fuller estimator is given by

$$V_4(\hat{B}_{12,f}|z,s) \cong \frac{S_{y1y2}}{nS_{y2y2}} [1+\bar{\underset{\sim}{t}}_s^T \underset{\sim}{S}_{tt}^{-1} \bar{\underset{\sim}{t}}_s], \qquad [3.99]$$

where $\bar{\underset{\sim}{t}}_s = n^{-1} \sum_s \underset{\sim}{t}_i$.

Comparing [3.98] and [3.99] we get

$$\frac{V_4(\hat{B}_{12,f}|z,s)}{V_4(\hat{B}_{12,ml}|z,s)} = 1+\bar{\underset{\sim}{t}}_s^T \underset{\sim}{S}_{tt}^{-1} \bar{\underset{\sim}{t}}_s . \qquad [3.100]$$

Since $\underset{\sim}{S}_{tt}$ is a positive definite matrix therefore from [3.100] we have

$$V_4(\hat{B}_{12,f}|z,s) \geq V_4(\hat{B}_{12,ml}|z,s). \qquad [3.101]$$

Equality sign holds only when the sample is balanced on the first and second moments in the case where y_2 is independent of z.From [3.101] we can therefore conclude that under our restrictions on model1 we find that the maximum likelihood estimator is more preferable in terms of minimum variance to the Fuller estimator with constant weights unless the sample is balanced on the mean,for example if the design used is an equal probability design the population mean and the sample mean are approximately equal.

3.4 MODIFIED FULLER ESTIMATORS

In the case of violation of the linearity assumption,we found in theorem 3.3 that the Fuller estimators are biased.The linear and nonlinear components of $\hat{\underset{\sim}{Q}}_f$ were found to be conditionally and unconditionally biased in general.A close look at these expectations reveal that there are cubic and quartic terms in these components.Since the proposed Fuller estimators accounts only for the linear and quadratic components ,we suspected that this might be the cause of the bias.

In order to adjust for these two terms we modified the adjustment vector $\underset{\sim}{t}_i$ defined for the Fuller estimators,so that it accounts for the linear,quadratic, cubic and quartic terms.Denote this modified vector by $\underset{\sim}{t}_{mi}$ defined as,

$$\underset{\sim}{t}_{mi}=\Big[\ (z_i-\bar{z}_U),\ (z_i-\bar{z}_U)^2-N^{-1}\sum_U(z_i-\bar{z}_U)^2, (z_i-\bar{z}_U)^3-N^{-1}\sum_U(z_i-\bar{z}_U)^3,$$

$$(z_i-\bar{z}_U)^4-N^{-1}\sum_U(z_i-\bar{z}_U)^4\Big].$$

We define the modified Fuller regression estimator,denoted by $\hat{\underset{\sim}{Q}}_{mf}$ as,

$$\hat{\underset{\sim}{Q}}_{mf}=\sum_s w_i\underset{\sim}{q}_i-\ \sum_s w_i\underset{\sim}{q}_i(\underset{\sim}{t}_{mi}-\bar{\underset{\sim}{t}}_m^{\ *})^T[\ \sum_s w_i(\underset{\sim}{t}_{mi}-\bar{\underset{\sim}{t}}_m^{\ *})(\underset{\sim}{t}_{mi}-\bar{\underset{\sim}{t}}_m^{\ *})^T]^{-1}\bar{\underset{\sim}{t}}_m^{\ *},$$

where

$$\underset{\sim}{t}_m^*=\sum_s w_i\underset{\sim}{t}_{mi},$$

$\underset{\sim}{t}_{mi}$ is as defined above and w_i denotes the weights defined in [3.6]

We did not attempt to look at the asymptotic properties of this estimator theoretically but we will study its asymptoptic properties empirically in chapter 4.

3.5 CONCLUSION

In this chapter we found that the Fuller estimators are asymptotically unbiased under the linear homoscedastic model but are asymptotically biased when the linearity assumption is violated.However if the weights are of Horvitz Thompson typewith ratio adjustment then the Fuller estimator with these weights is asymptotically unbiased under model 1 and retains its asymptotic unbiasedness properties when model 1assumptions are violated.When only the homoscedastic assumption is violated we proved that the Fuller estimators are robust to this violation.We can therefore conclude that when the population is linear and homoscedastic the Fuller estimators are preferable to the probability weighted estimators because they are approximately conditionally unbiased and are serious contenders to the maximum likelihood estimators.However when the population is linear and heteroscedastic then the Fuller estimators are the most preferable because they are approximately conditionally and unconditionally unbiased.Holmes(1987) found the maximum likelihood and the probability weighted estimators to be conditionally biased when the population is heteroscedastic. Though we proved that for unequal probability designs the maximum likelihood estimator has a smaller variance under model 1 as compared to the Fuller estimator,which of the two estimator to prefer will depend on the trade off between the variance and bias.To check the validity of the asymptotic assumptions made in this chapter and also to confirm the theoretical properties derived for the Fuller estimators we carried out a simulation study.The results of this study are reported in chapter 4.

CHAPTER 4

COMPARATIVE STUDY OF VARIOUS PARAMETRIC AND DESIGN BASED REGRESSION ESTIMATORS

4.1 INTRODUCTION

We will carry out two simulation studies,in the first study we compared the performance of the various estimators when the correlation structure and sample size are fixed and in the second study we will study the effects of varying the correlation structure ,sample size and the degree of nonlinearity for fixed sample design on various estimators. In the first study we will check the validity of the asymptotic results we proved in previous chapters.In order to do this, we carried out a simulation study with known properties.

In this study we will compare the ordinary least squares (ols) estimator which does not take into account the population structure,the probability weighted (pw)estimators which takes into account the population structure through the selection probabilities, the maximum likelihood (ml) estimator which takes into account the population structure through the sample design and the weighted version of the maximum likelihood (pwml) estimator with the estimators we proposed in chapters 2 and 3 under the;

(i) Linear homoscedastic model,

(ii) Quadratic homoscedastic model,

and

(iii) the Linear heteroscedastic model.

We will study extensively the empirical properties of the Fuller estimators i.e weighted Fuller (wf) and Equally weighted Fuller (ewf) ,and investigate whether the Fuller regression estimators have any significant gain in efficiency over the the probability weighted,maximum likelihood and probability weighted adjusted regression estimators,in the three cases.

In the second study we fix the sample design and the questions we address ourselves to are;

1.Do changes in population parameter values lead to predictable changes in performance of the six estimators?

2.If so, are the patterns similar over different estimators?

3 Do changes in the sample size lead to predictable changes in performance of the six estimators?

In order to be able to answer these quesions we will compare the performance of the six estimators for fixed sample design,when the,

(1) sample size is varied,

(2) correlation structure of the population under consideration is varied

and

(3) degree of nonlinearity between the variables is varied.

We now describe the first simulation study in the next section.

4.2 SIMULATION STUDY 1

The simulation study we carried out is similar to one carried out by Holmes(1987).We generated 10,000 finite population values of (y_{1i},y_{2i},z_i) i=1...10000 by first generating a value z_i from the uniform distribution U(0,10).Using the generated value of z_i the corresponding values of y_{2i} were obtained by

$$y_{2i}=m_2+H_2z_i+R_2z_i^2+\varepsilon_{2i},$$

where ε_{2i} is a random value selected from a normal distribution with mean 0 and variance w_{22}.
For each (y_{2i}, z_i) value generated the corresponding value of y_{1i} is generated by;

$$y_{1i}=m_1+H_1 z_i+R_1 z_i^2 +\varepsilon_{1i},$$

where ε_{1i} is a random value selected from a normal distribution with mean 0 and variance w_{11}.
Therefore given the value of z_i we have

$$\begin{bmatrix}\varepsilon_{1i}\\ \varepsilon_{2i}\end{bmatrix} \sim \text{MVN}\left[\quad 0 \quad , \begin{pmatrix} w_{11} & w_{12}\\ w_{21} & w_{22}\end{pmatrix}\right],$$

where ε_{1i} and ε_{2i} are generated by

$$\varepsilon_{2i}=w_{22}^{1/2}\eta_{2i},$$

and

$$\varepsilon_{1i}=w_{12}\ /w_{22}^{1/2}\eta_{2i}+(w_{11}-w_{12}^2/w_{22})^{1/2}\eta_{1i},$$

where η_{2i} and η_{1i} are values generated from independent N(0,1) variables.

Choosing various values of the parameters given in table 4.1 we can generate data under the;

(i) Linear homoscedastic model (model 1).($R_1=R_2=0$ and w_{ij} are constants).

(ii) Nonlinear Homoscedastic model (model 2).y_2 and y_1 are both nonlinear in z_i^2 ($R_1 \neq 0, R_2 \neq 0$ and w_{ij} are constants).

(iii)Linear heteroscedastic model (model 3).($R_1=R_2=0$ and w_{ij} are known functions of z).

Following Holmes(1987) we choose the following values for the parameters to generate the data for the three models.

TABLE 4.1 Table of parameter values for all the three models.

PARAMETERS	MODEL1	MODEL 2	MODEL3
m_2	10.0	10.0	10.0
H_2	1.0	0	0.5
R_2	0	-0.05	0
w_{22}	4.0	4.0	$8.0-1.2z+0.1z^2$
m_1	15.0	15.0	15.0
H_1	1.5	0	1.25
R_1	0	-0.075	0
w_{11}	5.0	5.0	$12.0-1.4z+0.135z^2$
w_{12}	2.0	2.0	$6.0-0.5z+0.06z^2$

These parameters values for generating data for the three models were chosen so that

(i) The regression function of y_1 and y_2 on z is a monotonically increasing function of z.

(ii) The regression of y_1 on y_2 defined in section 3.2 is approximately linear so that the regression coefficient β_{12} will be a meaningful parameter to estimate.

(iii) For model 3 the W_{ij} are quadratic functions of z.

This finite population was then stratified according to increasing value of z_i and then divided into five equal sized strata of 2000 units each such that the first stratum contained the first 2000 smallest values of z_i and the fifth strata contained the last 2000 largest values of z_i. From this stratified population a sample of size 300 was selected using the following sample designs;

TABLE 4.2 Table of sample designs and their symbols used in the conditional plots.

SAMPLE DESIGN			SAMPLE SIZES					SYMBOL
			n_1	n_2	n_3	n_4	n_5	
D1	proportionate	allocation	(60	60	60	60	60)	Δ
D2	increasing	allocation	(39	45	60	75	81)	∇
D3	increasing	allocation	(15	45	60	75	105)	+
D4	increasing	allocation	(15	15	30	90	150)	×
D5	increasing	allocation	(3	9	15	48	225)	□
D6	U-shaped	allocation	(90	45	30	45	90)	◇
D7	U-shaped	allocation	(135	12	6	12	135)	○
D8	U-shaped	allocation	(141	6	6	6	141)	*

For the various stratified sample designs we selected 1,000 independent samples of size 300 from the finite population. The sampling distribution of the various statistics under investigation were estimated from these 1,000 repeated samples. We carried out conditional and unconditional analyses.

To assess the conditional asymptotic properties of the estimators the 1,000 samples were divided into 20 groups of 50 samples each according to increasing values of $\Delta^F_{zz}=(s_{zzs}-S_{zz})/S_{zz}$ for the ewf,ols and ml estimators and $\Delta^{*F}_{zz}=(s^*_{zzs}-S_{zz})/S_{zz}$ for the pw , pwml and wf estimators respectively such that the first group contained the 50 samples with the smallest values of Δ^F_{zz}(or Δ^{*F}_{zz}) and so on upto the 20th group which contains the 50 samples with the largest values of Δ^F_{zz} (or Δ^{*F}_{zz}).We assume that the variation in Δ^F_{zz} (or Δ^{*F}_{zz}) within each group is small. The conditional distribution of the various estimators given Δ^F_{zz}(or Δ^{*F}_{zz}) could then be estimated.

LINEAR HOMOSCEDASTIC MODEL 1.
(a) UNCONDITIONAL ANALYSIS.

TABLE 4.3(linear homoscedastic model)

UNCONDITIONAL MEANS OF THE SIX REGRESSION ESTIMATORS

OVER 1,000 REPLICATIONS TRUE VALUE B_{12}=1.17

SAMPLE	MEANS					
DESIGN	$\hat{\beta}_{12,ols}$	$\hat{\beta}_{12,pw}$	$\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,pwml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,wf}$
D1	1.17	1.17	1.17	1.17	1.17	1.17
D2	1.15	1.17	1.17	1.17	1.17	1.17
D3	1.10	1.17	1.17	1.17	1.17	1.17
D4	1.06	1.17	1.17	1.17	1.17	1.17
D5	0.90	1.18	1.17	1.18	1.17	1.17
D6	1.22	1.17	1.17	1.17	1.16	1.17
D7	1.29	1.17	1.17	1.17	1.16	1.17
D8	1.29	1.17	1.17	1.17	1.16	1.17

TABLE 4.4 (linear homoscedastic model)

UNCONDITIONAL STANDARD DEVIATIONS OF THE SIX ESTIMATORS OVER 1,000 REPLICATIONS

SAMPLE DESIGN	STANDARD DEVIATIONS $\hat{\beta}_{12,ols}$	$\hat{\beta}_{12,pw}$	$\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,pwml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,wf}$
D1	0.0394	0.0394	0.0389	0.0389	0.0390	0.0390
D2	0.0386	0.0396	0.0383	0.0392	0.0394	0.0391
D3	0.0413	0.0520	0.0411	0.0514	0.0475	0.0514
D4	0.0433	0.0590	0.0440	0.0583	0.0598	0.0584
D5	0.0480	0.1070	0.0537	0.1060	0.0982	0.1050
D6	0.0358	0.0396	0.0358	0.0393	0.0388	0.0393
D7	0.0321	0.0670	0.0328	0.0662	0.0558	0.0663
D8	0.0307	0.0800	0.0313	0.0799	0.0585	0.0799

TABLE 4.5 (linear homoscedastic model)

UNCONDITIONAL MEAN SQUARE ERRORS OF THE SIX ESTIMATORS OVER 1,000 REPLICATIONS

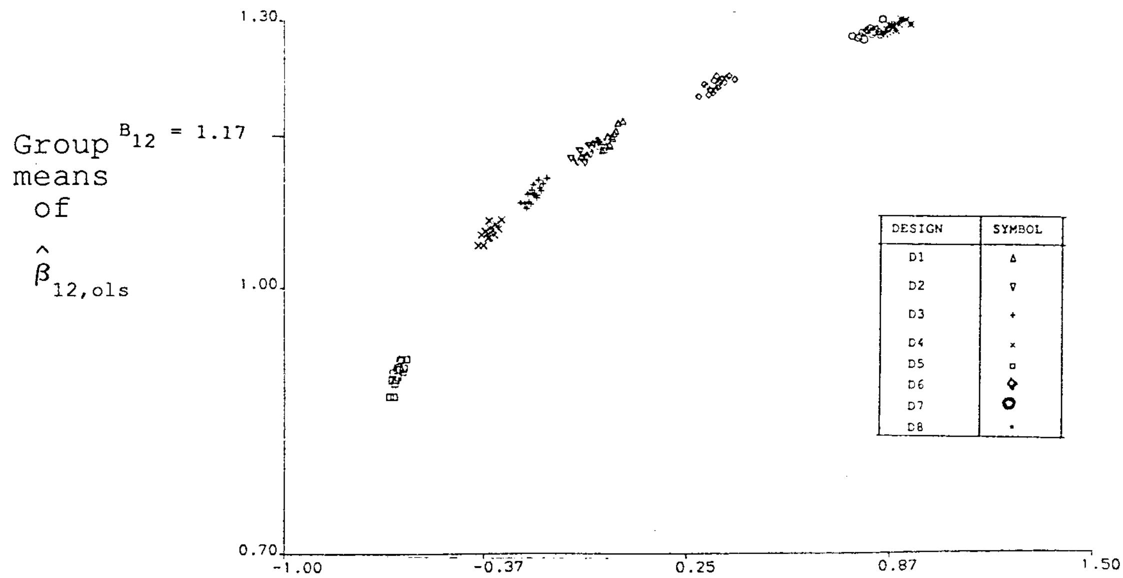
SAMPLE DESIGN	MEAN SQUARE ERRORS $\hat{\beta}_{12,ols}$	$\hat{\beta}_{12,pw}$	$\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,pwml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,wf}$
D1	0.0016	0.0016	0.0016	0.0016	0.0016	0.0016
D2	0.0021	0.0016	0.0015	0.0016	0.0016	0.0016
D3	0.0065	0.0028	0.0018	0.0027	0.0023	0.0027
D4	0.0145	0.0035	0.0020	0.0034	0.0036	0.0034
D5	0.0762	0.0115	0.0029	0.0111	0.0096	0.0111
D6	0.0042	0.0017	0.0013	0.0016	0.0016	0.0016
D7	0.0137	0.0045	0.0011	0.0044	0.0035	0.0044
D8	0.0146	0.0065	0.0010	0.0064	0.0041	0.0064

We see from table 4.3 that the design based estimators, maximum likelihood estimator and the equally weighted Fuller estimator in concordance with the theoretical results are approximately unconditionally unbiased under model 1.The ordinary least squares estimator which does not take into account the effect of the design is severely unconditional biased for unequal probability designs.From table 4.4 we note that the design based estimators have larger standard deviations than the model based and the equally weighted Fuller estimators.As seen from table 4.5 the maximum likelihood estimator is the most efficient across all the eight designs.We see that the equally weighted Fuller estimator is more efficient than the probability weighted estimator for the U-shaped designs.There does not seem to be any significant gain in efficiency using the probability weighted adjusted and the weighted Fuller estimators over the probability weighted estimator across all the designs.

(b) CONDITIONAL RESULTS.

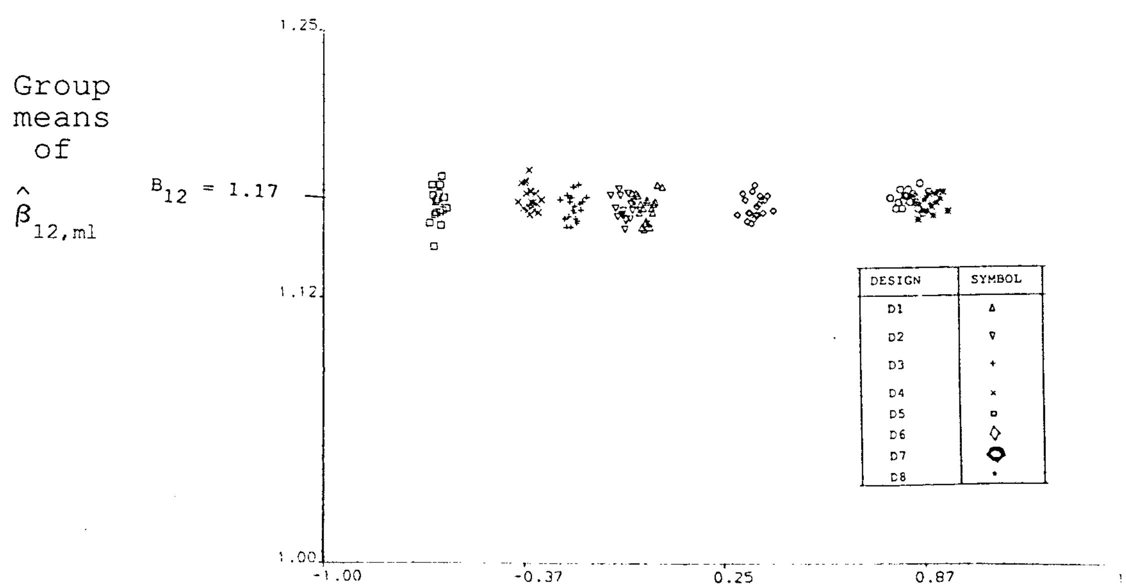
Figures 4.1-4.6 gives the plots of the group means of the ordinary least squares(ols),probability weighted(pw),maximum likelihood(ml),probability weighted adjusted (pwml),weighted Fuller(wf) and the equally weighted Fuller (ewf) estimators plotted against the group means of Δ_{zz}^{F} (for ols,ml,ewf) and Δ_{zz}^{*F} (for pw,pwml,wf). We see from the plots that the ols estimator and the probability weighted estimators are conditional biased for the unequal probability designs.The ml,ewf and wf estimators as expected from theoretical results are approximately conditionally unbiased.The probability weighted adjusted estimator seem to have removed the conditional bias in the probability weighted estimator and is approximately conditionally unbiased.

FIG 4.1 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,ols}$ VS GROUP MEANS OF Δ_{zz}^{F} (20 groups, 50 samples per group).



Group means of Δ_{zz}^{F} .

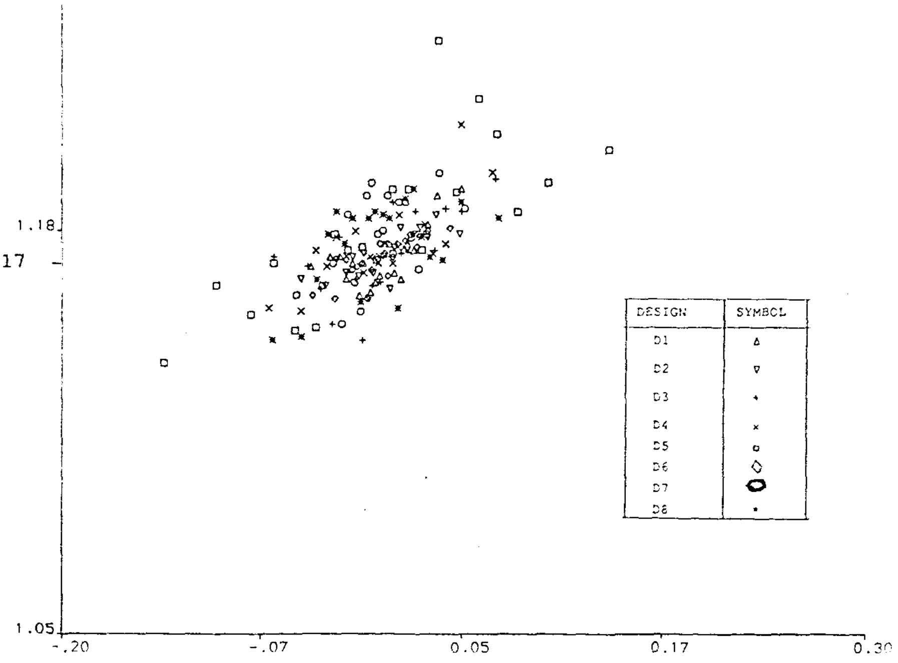
FIG 4.2 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,ml}$ VS GROUP MEANS OF Δ_{zz}^{F} (20 groups, 50 samples per group).



Group means of Δ_{zz}^{F} .

FIG 4.3 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,pw}$ VS GROUP MEANS OF Δ^{*F}_{ZZ} (20 groups, 50 samples per group).

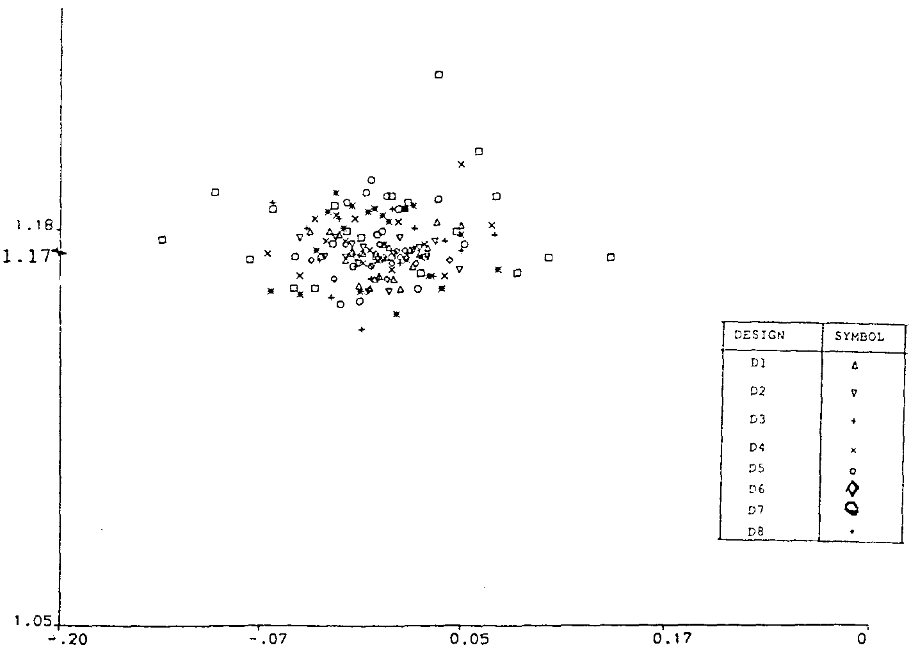
Group means of $\hat{\beta}_{12,pw}$



Group means of Δ^{*F}_{ZZ} .

FIG 4.4 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,pwml}$ VS GROUP MEANS OF Δ^{*F}_{ZZ} (20 groups, 50 samples per group).

Group means of $\hat{\beta}_{12,pwml}$



Group means of Δ^{*F}_{ZZ} .

FIG 4.5 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,wf}$ VS GROUP MEANS OF Δ^{*F}_{ZZ} (20 groups, 50 samples per group).

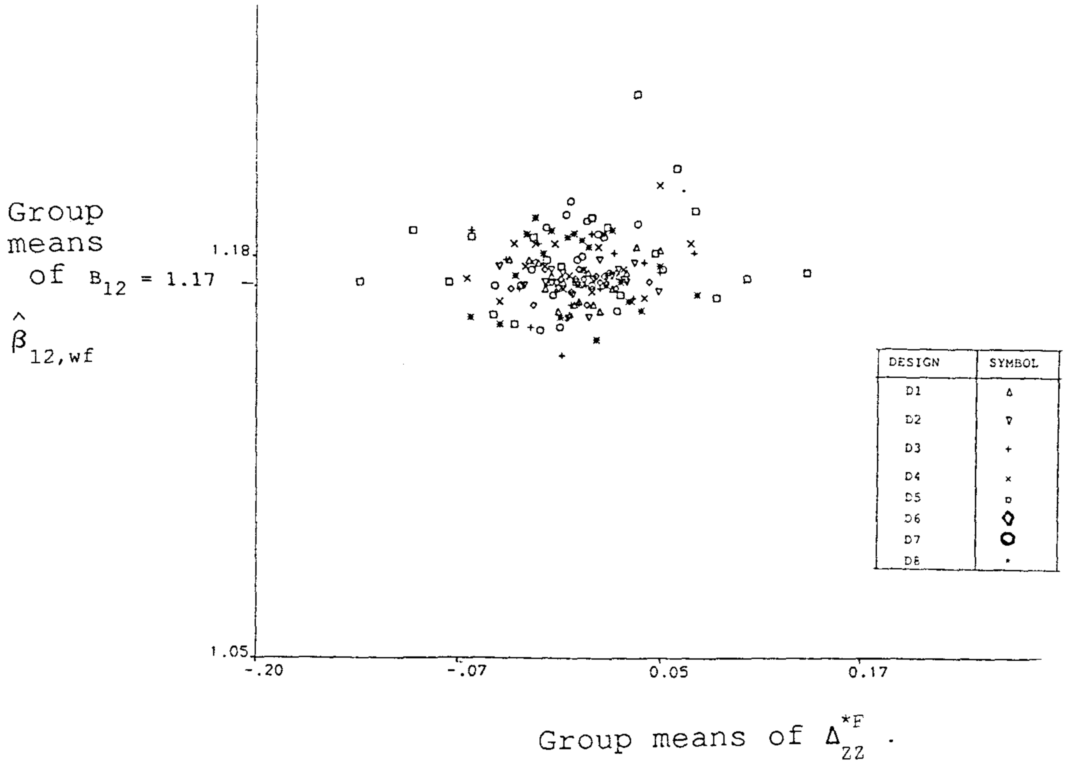
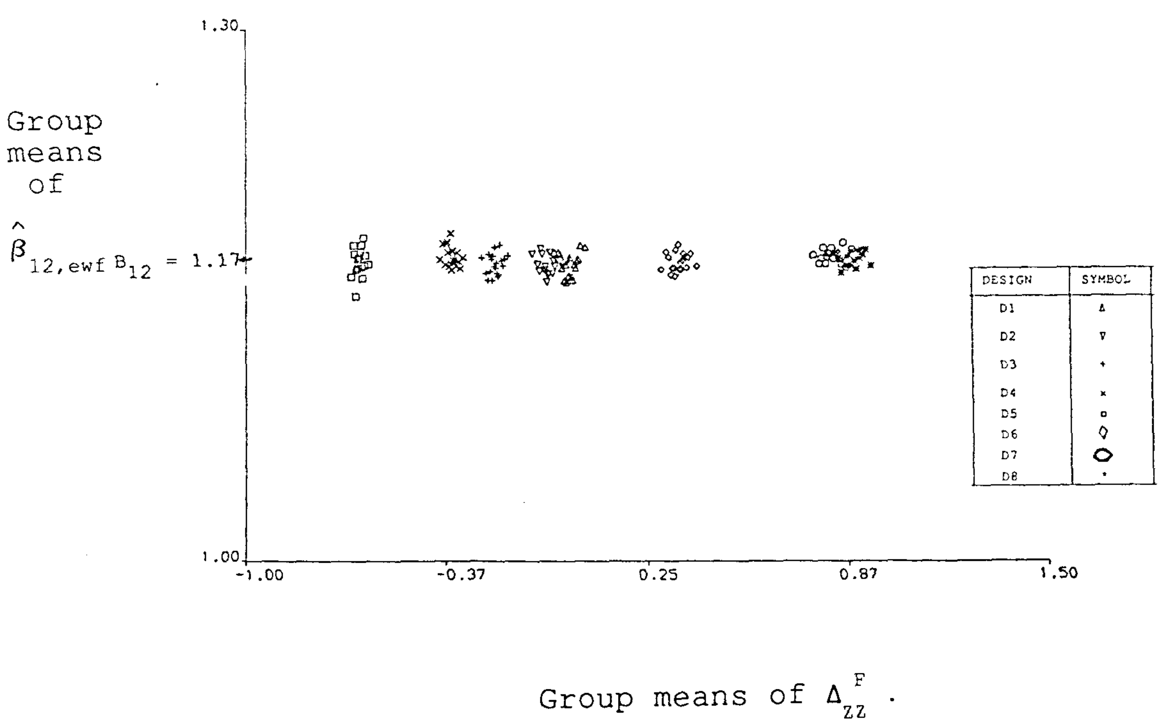


FIG 4.6 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,ewf}$ VS GROUP MEANS OF Δ^{F}_{ZZ} (20 groups, 50 samples per group).



We now investigate the empirical properties of the estimators studied above under the linear homoscedastic model and see whether ml,ewf,wf and pwml estimators retains their optimum properties when the linearity assumption is violated.

NONLINEAR HOMOSCEDASTIC MODEL 2.

(a) UNCONDITIONAL ANALYSIS.

TABLE 4.6 (quadratic homoscedastic model)

UNCONDITIONAL MEANS OF THE SIX REGRESSION ESTIMATORS OVER 1,000 REPLICATIONS

TRUE VALUE B_{12}=0.857

SAMPLE			MEANS			
DESIGN	$\hat{\beta}_{12,ols}$	$\hat{\beta}_{12,pw}$	$\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,pwml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,wf}$
D1	0.846	0.846	0.846	0.846	0.846	0.846
D2	0.804	0.845	0.819	0.845	0.849	0.845
D3	0.719	0.843	0.766	0.844	0.823	0.844
D4	0.685	0.847	0.761	0.847	0.870	0.847
D5	0.557	0.840	0.675	0.840	0.816	0.839
D6	0.915	0.845	0.849	0.845	0.843	0.845
D7	0.990	0.854	0.849	0.854	0.835	0.854
D8	0.994	0.848	0.844	0.848	0.821	0.848

TABLE 4.7 (quadratic homoscedastic model)

UNCONDITIONAL STANDARD DEVIATIONS OF THE SIX ESTIMATORS OVER 1,000 REPLICATIONS

SAMPLE	STARDARD DEVIATIONS					
DESIGN	$\hat{\beta}_{12,ols}$	$\hat{\beta}_{12,pw}$	$\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,pwml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,wf}$
D1	0.051	0.051	0.051	0.050	0.050	0.050
D2	0.052	0.054	0.052	0.052	0.054	0.053
D3	0.054	0.071	0.055	0.070	0.066	0.070
D4	0.056	0.081	0.058	0.080	0.081	0.080
D5	0.055	0.149	0.065	0.147	0.141	0.146
D6	0.049	0.054	0.055	0.049	0.053	0.054
D7	0.045	0.094	0.046	0.093	0.074	0.094
D8	0.043	0.114	0.044	0.114	0.079	0.114

TABLE 4.8 (quadratic homoscedastic model)

UNCONDITIONAL MEAN SQUARE ERRORS OF THE SIX ESTIMATORS OVER 1,000 REPLICATIONS

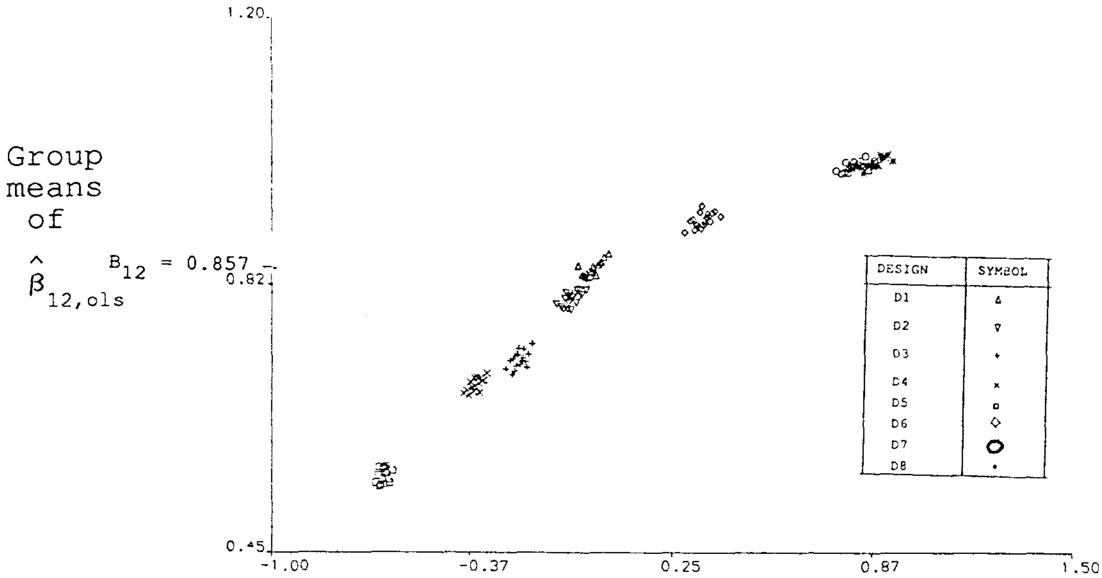
SAMPLE	MEAN SQUARE ERRORS					
DESIGN	$\hat{\beta}_{12,ols}$	$\hat{\beta}_{12,pw}$	$\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,pwml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,wf}$
D1	0.0027	0.0027	0.0026	0.0026	0.0026	0.0026
D2	0.0055	0.0030	0.0042	0.0030	0.0029	0.0030
D3	0.0221	0.0052	0.0114	0.0051	0.0056	0.0051
D4	0.0328	0.0066	0.0125	0.0065	0.0067	0.0064
D5	0.0931	0.0225	0.0372	0.0219	0.0217	0.0217
D6	0.0057	0.0031	0.0025	0.0031	0.0030	0.0030
D7	0.0197	0.0088	0.0022	0.0087	0.0059	0.0088
D8	0.0205	0.0130	0.0021	0.0130	0.0075	0.0131

Tables 4.6-4.8 gives the means,standard deviations and the mean square errors of the six estimators under study when the linearity assumption is violated.We see that the ols estimator is still severely biased for the unequal probability designs and the design based estimators as expected are approximately unconditional unbiased.The maximum likelihood estimator becomes severely downward biased for the increasing allocation design but remains approximately unconditionally unbiased for the U-shaped and the equal probability designs.The equally weighted Fuller estimator also becomes biased for the unequal probability designs.However we note that its bias for the increasing allocation designs is less than that of the maximum likelihood estimator.From table 4.7 we see that the design based estimators have higher standard deviations than the model based estimators and the equally weighted Fuller estimator.We see from table 4.8 that the design based estimators are the most efficient for the increasing allocation designs and the maximum likelihood estimator is the most efficient for the U-shaped designs.The equally weighted Fuller estimator is more efficient than the design based estimators for the U-shaped designs and also more efficient than the maximum likelihood estimator for the increasing allocation designs.It thus compromises the efficiency properties of the ml and the design based estimators across all the designs.

(b) CONDITIONAL RESULTS.

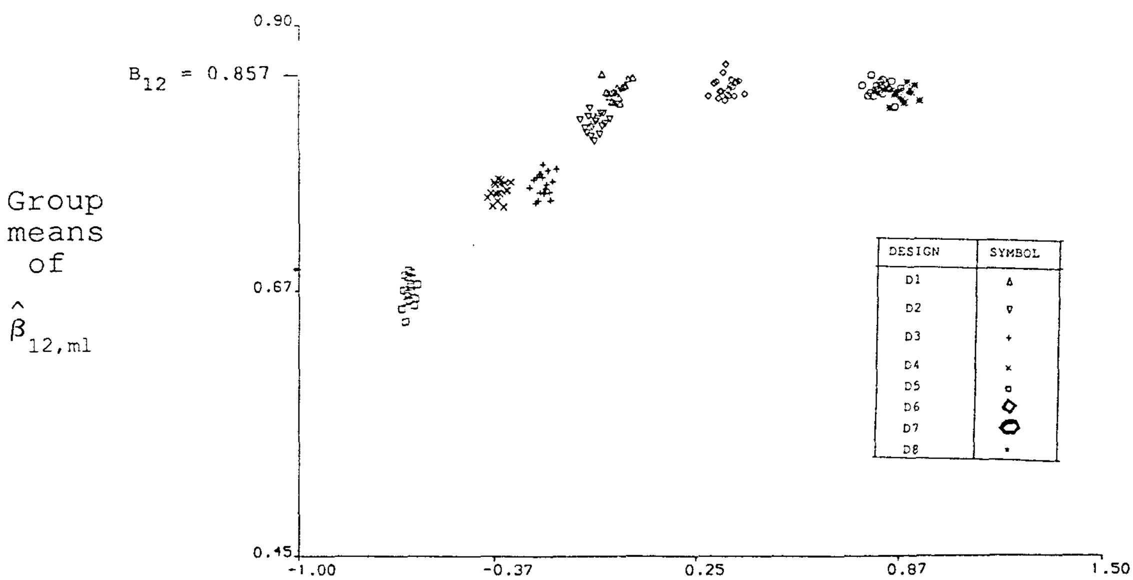
The conditional results for all the six estimators are given in figures 4.7-4.12.We see that the ols and pw estimators are conditionally biased across the unequal probability designs.We note that under model 2 the ml estimator is now conditionally biased across the increasing allocation designs but is approximately conditionally unbiased across the U-shaped and the equal probability designs.The probability weighted adjusted and the weighted Fuller

FIG 4.7 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,ols}$ VS GROUP MEANS OF Δ^{F}_{zz} (20 groups, 50 samples per group).



Group means of Δ^{F}_{zz} .

FIG 4.8 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,ml}$ VS GROUP MEANS OF Δ^{F}_{zz} (20 groups, 50 samples per group).



Group means of Δ^{F}_{zz} .

FIG 4.9 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,pw}$ VS GROUP MEANS OF Δ^{*F}_{ZZ} (20 groups, 50 samples per group).

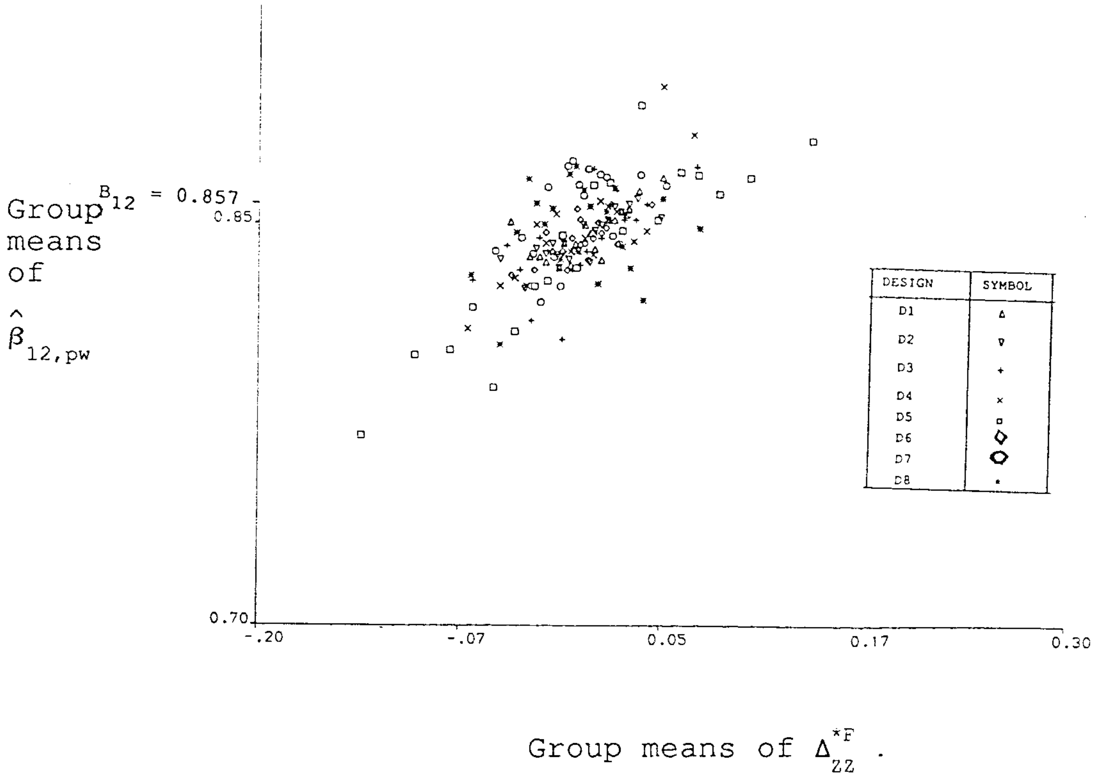


FIG 4.10 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,pwml}$ VS GROUP MEANS OF Δ^{*F}_{--} (20 groups, 50 samples per group).

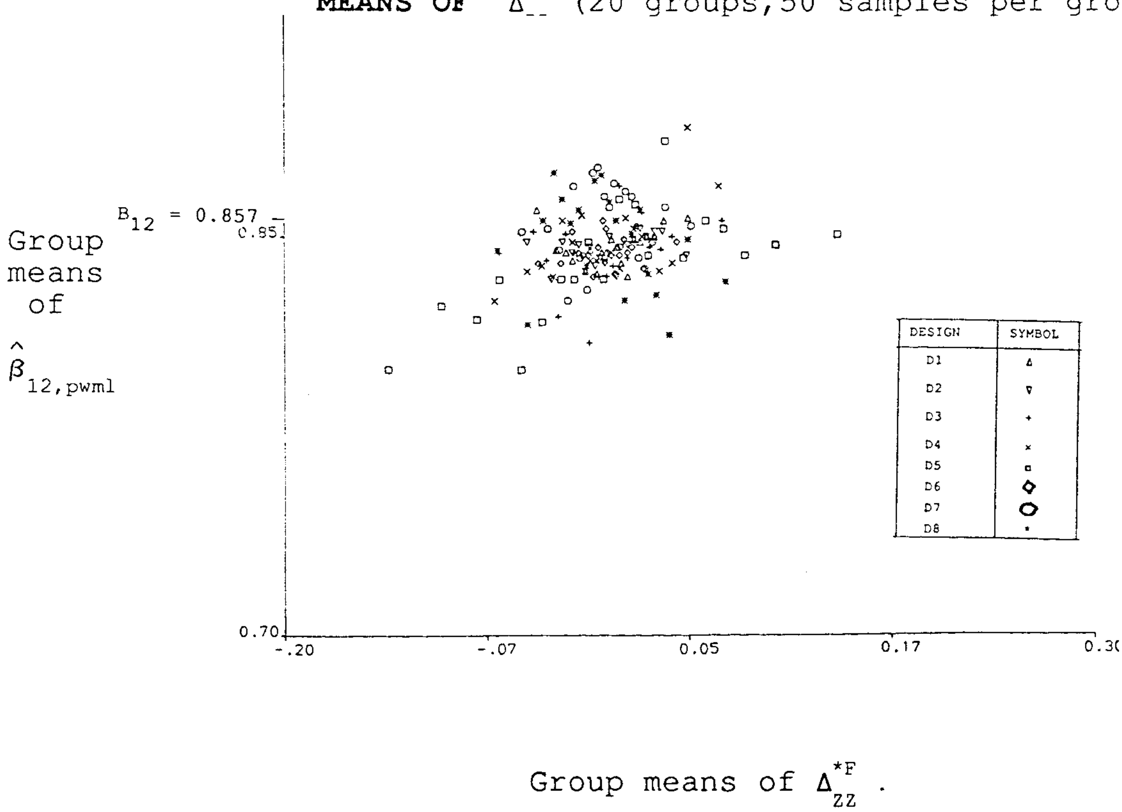


FIG 4.11 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,wf}$ VS GROUP MEANS OF Δ_{ZZ}^{*F} (20 groups,50 samples per group).

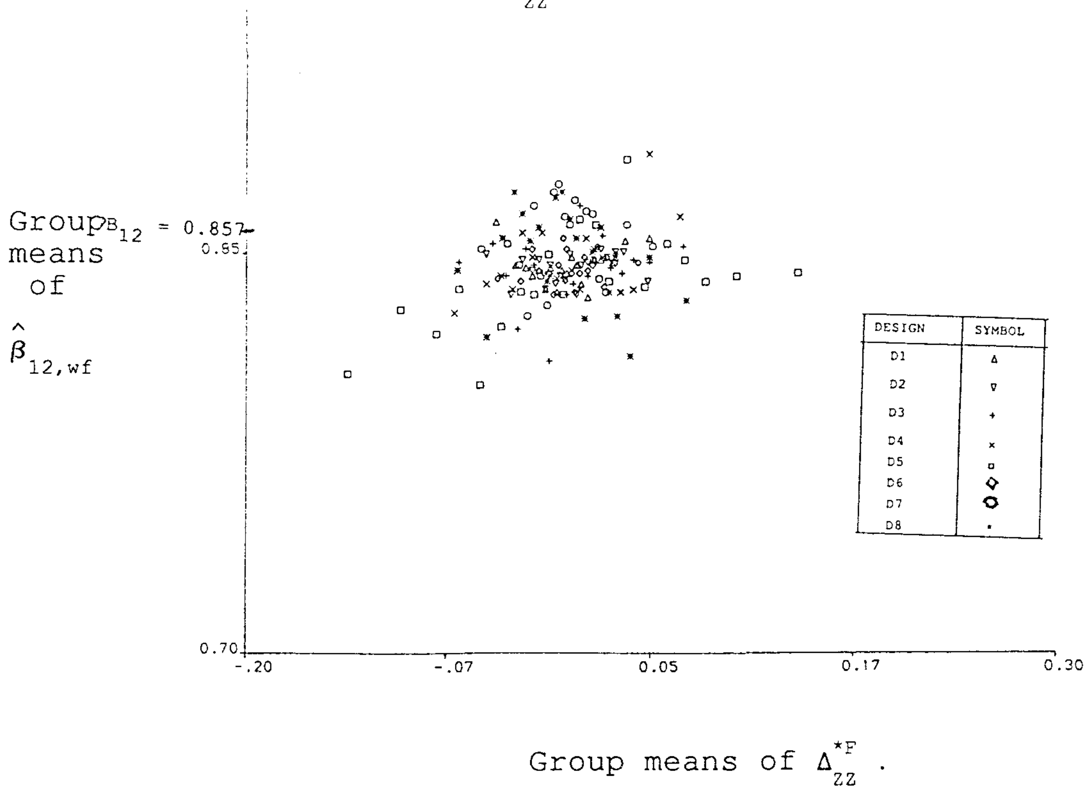
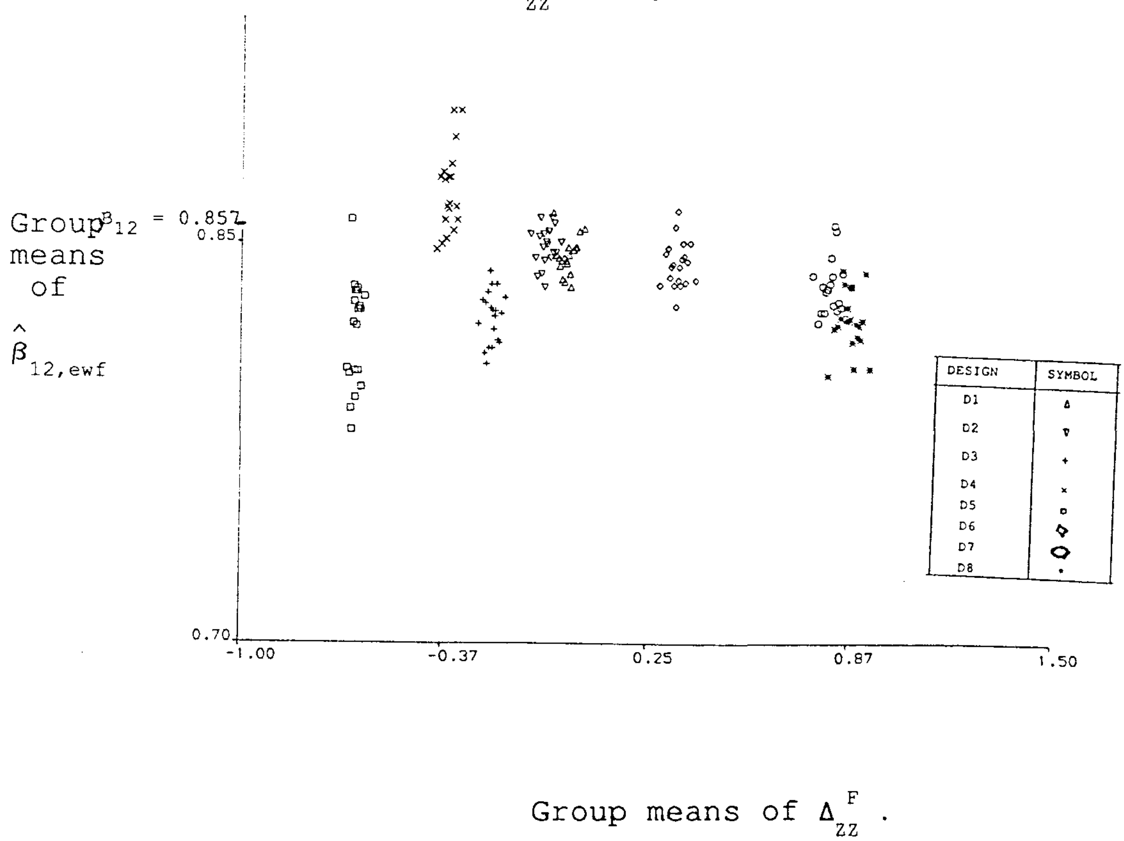


FIG 4.12 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,ewf}$ VS GROUP MEANS OF Δ_{ZZ}^{F} (20 groups,50 samples per group).



estimator also becomes conditionally biased across the unequal probability designs.For the equally weighted Fuller estimator we see that though it reveals some design specific bias its conditional bias for the increasing allocation designs is quite small compared to that of the ml estimator.

We will now investigate the empirical properties of the estimators when only the homoscedasticity assumption is violated.

LINEAR HETEROSCEDASTIC MODEL 3.

(a) UNCONDITIONAL ANALYSIS.

TABLE 4.9 (linear heteroscedastic model)

UNCONDITIONAL MEANS OF THE OF THE SIX REGRESSION ESTIMATORS OVER 1,000 REPLICATIONS TRUE VALUE B_{12}=1.44

SAMPLE			MEANS			
DESIGN	$\hat{\beta}_{12,ols}$	$\hat{\beta}_{12,pw}$	$\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,pwml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,wf}$
D1	1.44	1.44	1.44	1.44	1.44	1.44
D2	1.45	1.43	1.47	1.43	1.43	1.43
D3	1.43	1.43	1.50	1.43	1.43	1.43
D4	1.42	1.44	1.53	1.44	1.45	1.44
D5	1.32	1.44	1.53	1.44	1.43	1.44
D6	1.50	1.44	1.40	1.44	1.43	1.44
D7	1.57	1.44	1.38	1.44	1.44	1.44
D8	1.57	1.44	1.37	1.44	1.43	1.44

TABLE 4.10 (linear heteroscedastic model)

UNCONDITIONAL STANDARD DEVIATIONS OF THE SIX ESTIMATORS OVER 1,000 REPLICATIONS

SAMPLE	STARDARD DEVIATIONS					
DESIGN	$\hat{\beta}_{12,ols}$	$\hat{\beta}_{12,pw}$	$\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,pwml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,wf}$
D1	0.052	0.052	0.052	0.052	0.052	0.052
D2	0.051	0.057	0.051	0.057	0.057	0.057
D3	0.051	0.082	0.052	0.082	0.075	0.082
D4	0.048	0.091	0.052	0.090	0.094	0.090
D5	0.045	0.179	0.058	0.179	0.162	0.176
D6	0.053	0.054	0.052	0.053	0.052	0.053
D7	0.049	0.081	0.048	0.081	0.072	0.081
D8	0.048	0.105	0.048	0.106	0.080	0.105

TABLE 4.11 (linear heteroscedastic model)

UNCONDITIONAL MEAN SQUARE ERRORS OF THE SIX ESTIMATORS OVER 1,000 REPLICATIONS

SAMPLE	MEAN SQUARE ERRORS					
DESIGN	$\hat{\beta}_{12,ols}$	$\hat{\beta}_{12,pw}$	$\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,pwml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,wf}$
D1	0.0028	0.0028	0.0028	0.0028	0.0028	0.0028
D2	0.0026	0.0034	0.0033	0.0034	0.0034	0.0034
D3	0.0027	0.0068	0.0062	0.0069	0.0058	0.0068
D4	0.0031	0.0082	0.0101	0.0081	0.0089	0.0081
D5	0.0176	0.0321	0.0115	0.0320	0.0264	0.0311
D6	0.0059	0.0029	0.0039	0.0029	0.0028	0.0029
D7	0.0186	0.0066	0.0068	0.0066	0.0052	0.0066
D8	0.0193	0.0112	0.0079	0.0112	0.0067	0.0112

Tables 4.9-4.11 gives the means,standard deviations and mean square errors for the six estimators under model 3.We see that the ols and the ml estimators are severely biased for the unequal probability designs. The design based estimators as expected are approximately unconditionally unbiased.We note that the equally weighted Fuller estimator is approximately unconditional unbiased across all the designs.The model based estimators and equally weighted Fuller estimator have smaller standard deviations than the design based estimators .From table 4.9 we see that the equally weighted Fuller estimator is the most efficient estimator under model 3 for the U-shaped designs and with exception of the ols estimator which is more efficient for the increasing allocation design,it is the most efficient across all the sample designs,except for the extreme designs D5.

(b) CONDITIONAL RESULTS.

The conditional plots for the six estimators when the homoscedastic model assumption is violated are given in figures [4.13]-[4.18].We see from the plots that in line with the theoretical results proved by Holmes(1987) and those of the Fuller estimators we proved in chapter 3, the ordinary least squares,maximum likelihood,probability weighted adjusted and probability weighted estimators are conditional biased for the unequal probability design.The Fuller estimators are approximately conditional unbiased across all the eight designs.

FIG 4.13 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,ols}$ VS GROUP MEANS OF Δ^{F}_{ZZ} (20 groups, 50 samples per group).

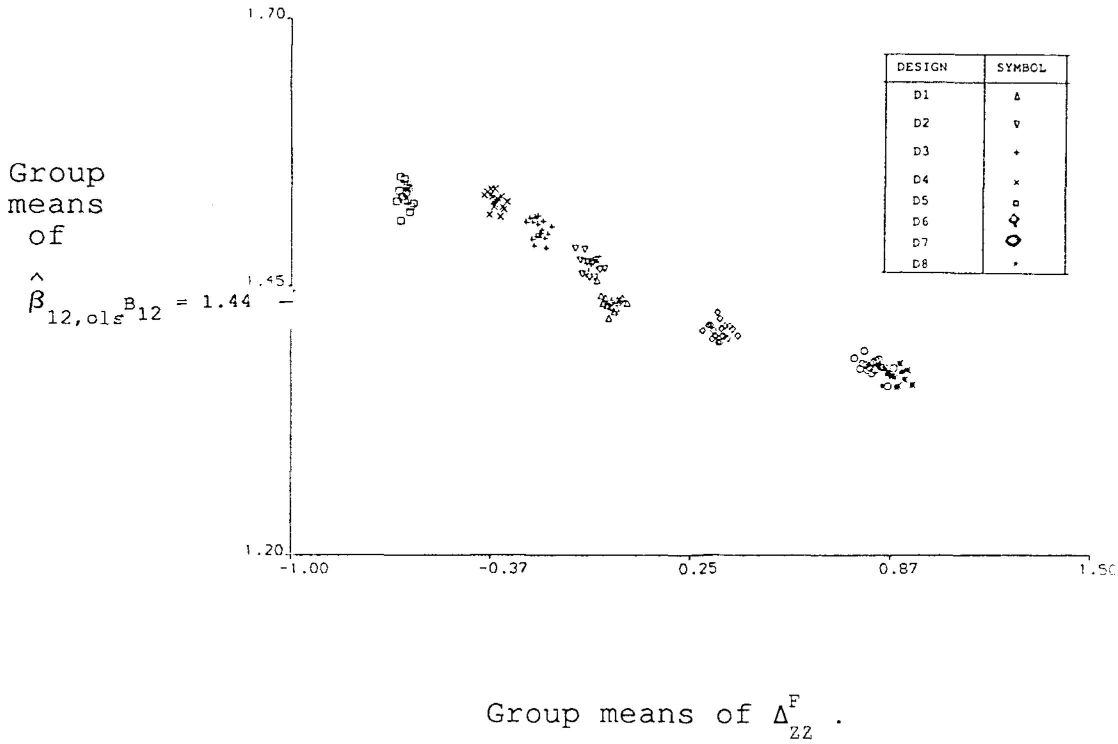


FIG 4.14 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,ml}$ VS GROUP MEANS OF Δ^{F}_{ZZ} (20 groups, 50 samples per group).

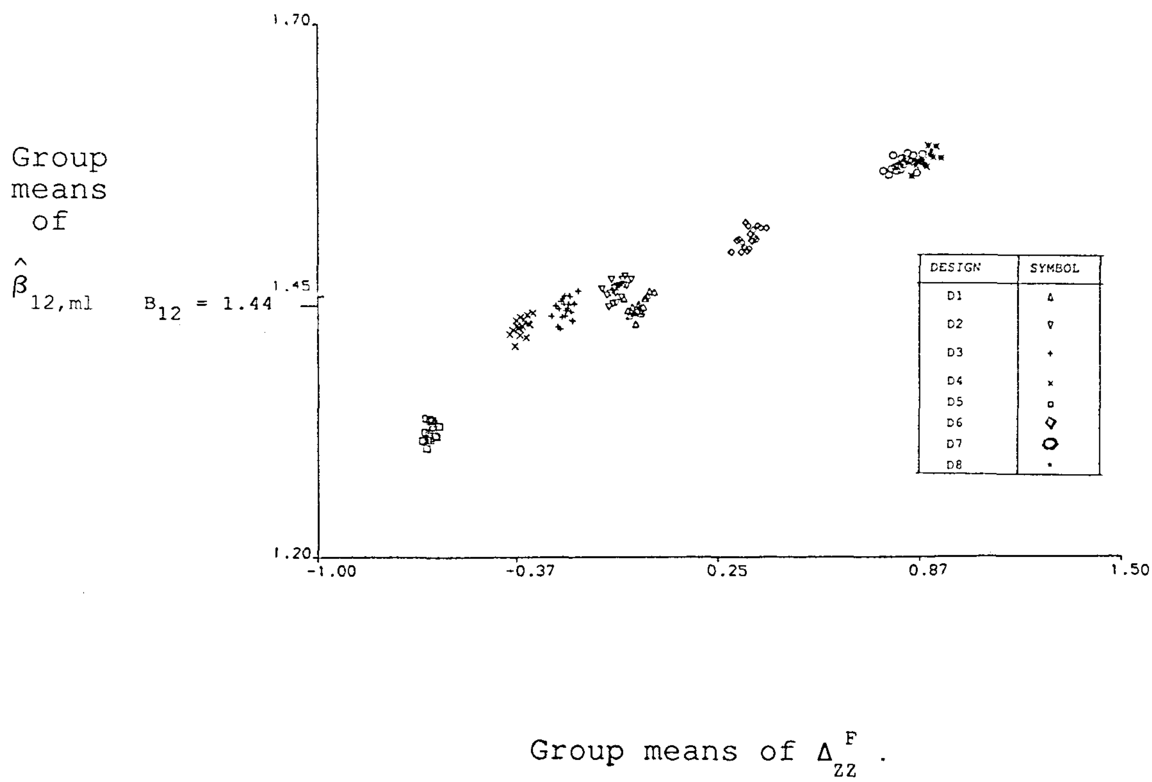


FIG 4.15 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,pw}$ VS GROUP MEANS OF Δ^{*F}_{zz} (20 groups, 50 samples per group).

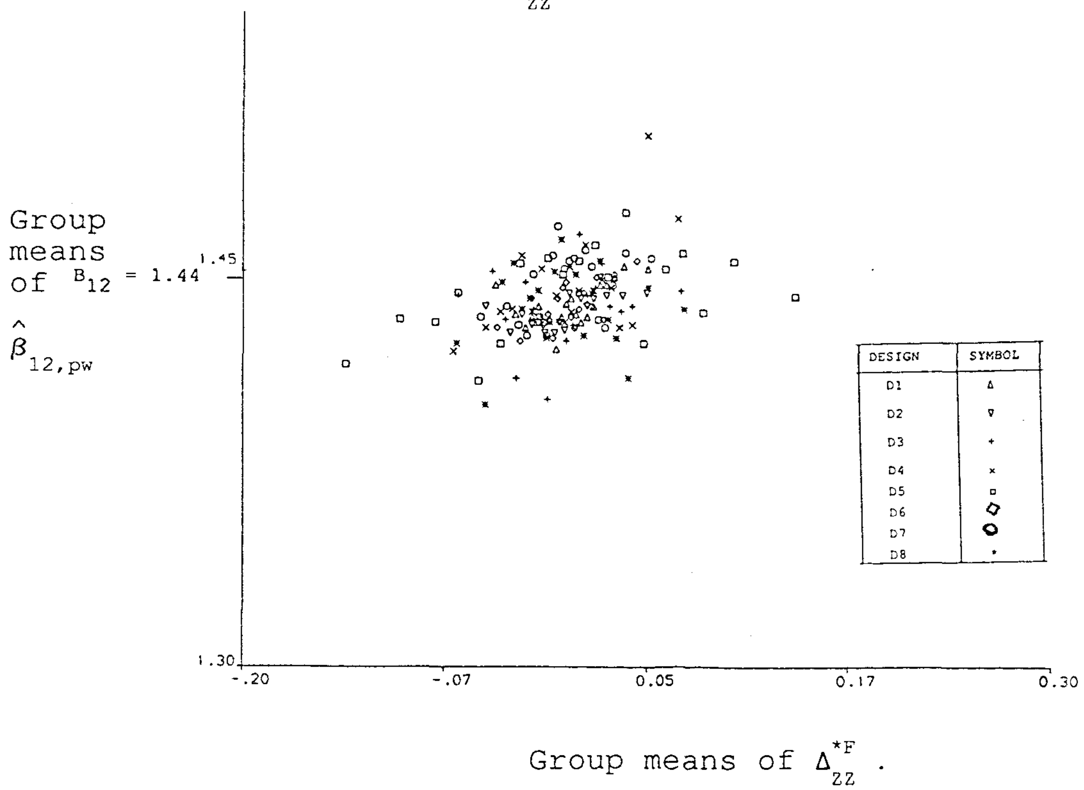


FIG 4.16 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,pwml}$ VS GROUP MEANS OF Δ^{*F}_{zz} (20 groups, 50 samples per group).

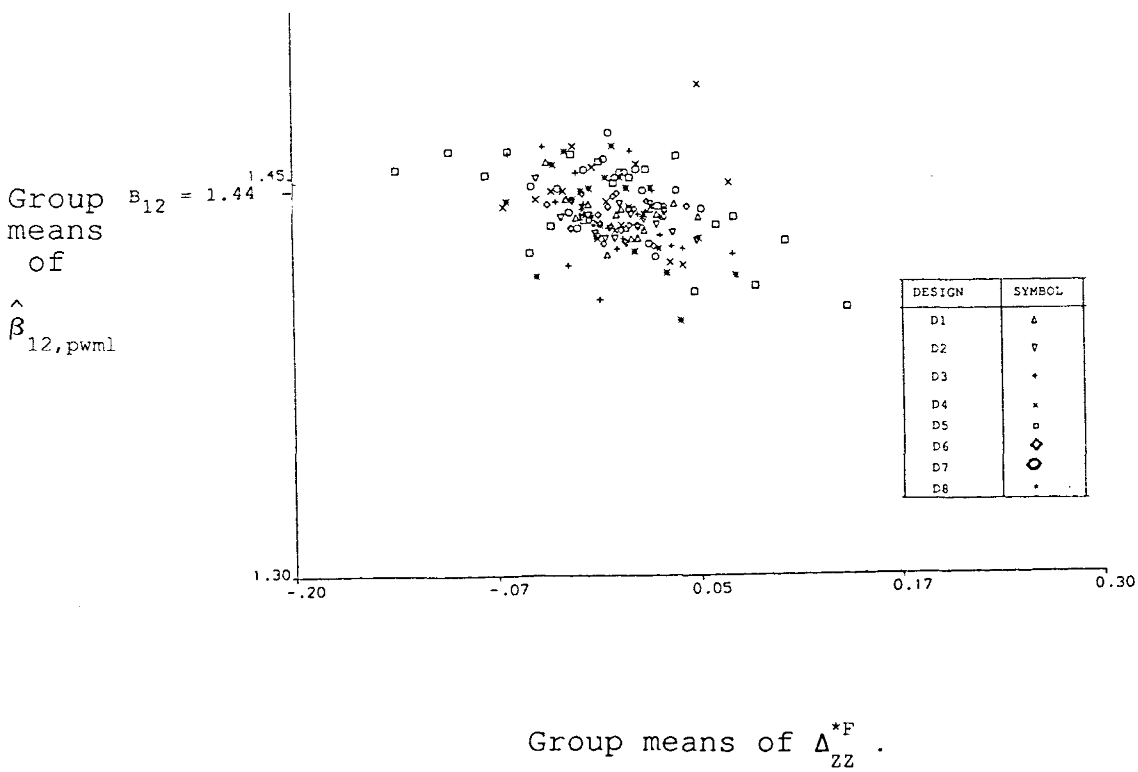


FIG 4.17 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,wf}$ **VS GROUP MEANS OF** Δ_{ZZ}^{*F} (20 groups, 50 samples per group).

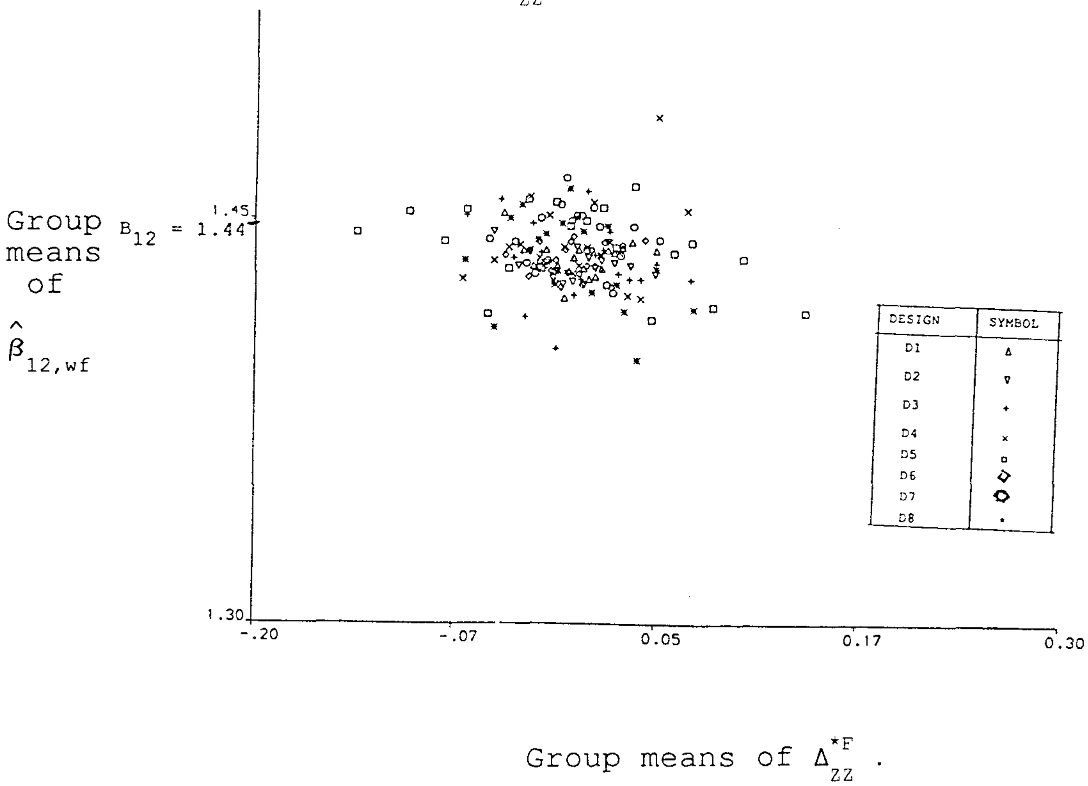
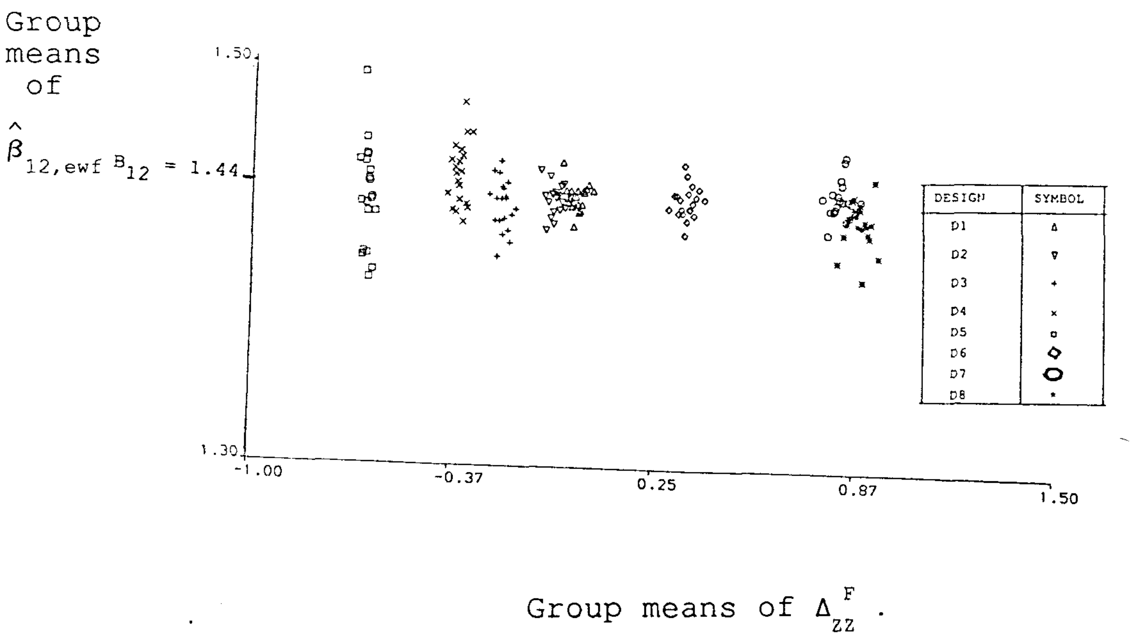


FIG 4.18 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,ewf}$ **VS GROUP MEANS OF** Δ_{ZZ}^{F} (20 groups, 50 samples per group).



To investigate further the performance of the six estimators we looked at the coverage properties of the estimators.For each procedure we calculate the frequency of which the 100(1-α) confidence intervals,

$$\hat{\beta}_{12}\pm z_{\alpha/2}[\hat{V}(\hat{\beta}_{12})]^{1/2},$$

cover the true value of the regression coefficient β_{12},where $z_{\alpha/2}$ denotes the standard normal deviate such that,

$$P(|Z|<z_{\alpha/2}]=\alpha.$$

$\hat{\beta}_{12}$ denotes the estimator of the regression coefficient,

$\hat{V}(\hat{B}_{12})$ denotes the estimator of its variance of the various estimators to be studied.These estimators of the variance of all the six estimators included in this study were derived by Holmes [1987].

Since all the eight designs are essentially only three sampling designs,the equal probability design,increasing allocation and U-shaped allocation design,we decided to study the properties of only three designs representing the three sampling schemes,these are D1,D5 and D8.

We present below the coverage probabilities for a 95% confidence interval (α=0.05).

TABLE 4.12 COVERAGE PROBABILITIES AVERAGED OVER 500 REPLICATIONS FOR EACH DESIGN.

ESTIMATOR	MODEL 1			MODEL 2			MODEL 3		
	SAMPLE DESIGNS								
	D1	D5	D8	D1	D5	D8	D1	D5	D8
$\hat{B}_{12,ols}$	97.4	0.4*	24.8*	98.4	0.8*	33.8*	98.4	53.0*	55.0*
$\hat{B}_{12,ml}$	93.8	96.6	95.6	96.8	37.2*	95.8	93.0	48.4*	64.6*
$\hat{B}_{12,pw}$	96.2	88.4*	95.4	96.6	91.4	94.6	97.6	91.2	93.8
$\hat{B}_{12,pwm}$	96.4	90.2	95.4	96.6	90.4	94.4	97.8	90.0	94.4
$\hat{B}_{12,ewf}$	96.2	71.8*	90.0	96.8	64.4*	89.8*	97.6	55.4*	93.2
$\hat{B}_{12,wf}$	96.3	89.1*	95.6	96.6	92.0	94.2	97.6	90.2	94.0

* indicates coverage is less than 90%.

Table 4.12 present the coverage properties for a nominal 90% confidence interval for all the six estimators.We see that the coverage properties for the equal probability design D_1 for all the estimators across the three sample designs and models considered is very good.For the ordinary least squares estimator,the coverage properties for the unequal probability designs D2 and D3 are very poor due to its severe bias in these designs across the three models.The coverage properties of the design based estimators as expected are quite good across all the three models and sample designs.This is because these estimators are approximately unconditionally unbiased.The coverage properties of the maximum likelihood estimator is quite good for all the designs in model 1.For model 2,the ml estimator does not have very good coverage properties for D2 because of its severe bias for this design,but since it is approximately unconditionally unbiased for the design D3 and D1 its coverage properties are good for these designs.For model 3,we see that the coverage properties for the ml estimators for the unequal propbaility design is very poor.The coverage properties for the equally weighted Fuller are good across all the models for designs D1 and D3,but are very poor for the increasing allocation design D2.

4.3 MODIFIED ESTIMATORS

In chapter 2, section 2.4 we proposed a modified maximum likelihood estimator which takes into account the quadratic nature of the population,denote this estimator by Qml. In chapter 3 section 3.4 we proposed a modified Fuller estimator.If the weights of the estimator are constant,then we denote the equally weighted modified Fuller estimator by Mewf.

We present below the unconditional results of the modified equally weighted Fuller estimator and the Quadratic maximum likelihood estimator.We will present results only for the designs D1,D5 and D8 when the population is linear and homoscedastic and when the population is nonlinear and homoscedastic.

TABLE 4.13(linear homoscedastic model).

UNCONDITIONAL MEANS OF THE FOUR ESTIMATORS OF REGRESSION COEFFICIENT OVER 1000 REPLICATIONS FOR EACH DESIGN.

N=10000,n=300 True value B_{12}=1.17

SAMPLE	MEANS			
DESIGN	$\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,Qml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,Mewf}$
D1	1.17	1.17	1.17	1.18
D5	1.17	1.14	1.17	1.19
D8	1.17	1.18	1.17	1.25

TABLE 4.14(linear homoscedastic model).

UNCONDITIONAL STANDARD DEVIATIONS OF THE FOUR ESTIMATORS OF REGRESSION COEFFICIENT OVER 1000 REPLICATIONS FOR EACH DESIGN.

N=10000,n=300

SAMPLE	STANDARD DEVIATIONS.			
DESIGN	$\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,Qml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,Mewf}$
D1	0.0389	0.0394	0.0390	0.0665
D5	0.0537	0.1010	0.0982	0.1440
D8	0.0313	0.0585	0.0708	0.2260

TABLE 4.15(linear homoscedastic model).

UNCONDITIONAL MEAN SQUARE ERRORS OF THE FOUR ESTIMATORS OF REGRESSION COEFFICIENT OVER 1000 REPLICATIONS FOR EACH DESIGN.

N=10000,n=300

SAMPLE	MEAN SQUARE ERRORS.			
DESIGN	$\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,Qml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,Mewf}$
D1	0.0016	0.0016	0.0016	0.0044
D5	0.0029	0.0115	0.0096	0.0212
D8	0.0010	0.0051	0.0041	0.0564

TABLE 4.16(nonlinear homoscedastic model).

UNCONDITIONAL MEANS OF THE FOUR ESTIMATORS OF REGRESSION COEFFICIENT OVER 1000 REPLICATIONS FOR EACH DESIGN.

N=10000,n=300 True value B_{12}=0.852

SAMPLE	MEANS			
DESIGN	$\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,Qml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,Mewf}$
D1	0.846	0.846	0.846	0.939
D5	0.675	0.958	0.816	0.907
D8	0.844	0.967	0.821	1.000

TABLE 4.17(nonlinear homoscedastic model).

UNCONDITIONAL STANDARD DEVIATIONS OF THE FOUR ESTIMATORS OF REGRESSION COEFFICIENT OVER 1000 REPLICATIONS FOR EACH DESIGN.

N=10000,n=300

SAMPLE	STANDARD DEVIATIONS.			
DESIGN	$\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,Qml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,Mewf}$
D1	0.0510	0.0510	0.0500	0.0741
D5	0.0650	0.1750	0.1460	0.1770
D8	0.0440	18.7	0.0790	0.1910

TABLE 4.18(nonlinear homoscedastic model).

UNCONDITIONAL MEAN SQUARE ERRORS OF THE FOUR ESTIMATORS OF REGRESSION COEFFICIENT OVER 1000 REPLICATIONS FOR EACH DESIGN.

N=10000,n=300

SAMPLE	MEAN SQUARE ERRORS.			
DESIGN	$\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,Qml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,Mewf}$
D1	0.0026	0.0027	0.0026	0.0121
D5	0.0372	0.0410	0.0217	0.0497
D8	0.0021	350.0	0.0075	0.0571

We see from tables [4.13]-[4.15] that when the data is linear and homoscedastic the modified equally weighted Fuller estimator and the quadratic maximum likelihood estimator are unconditionally biased for the unequal probability designs D5 and D9.We also note that as compared to the equally weighted Fuller and the maximum likelihood estimators the modified equally weighted Fuller (Mewf)and the quadratic maximum likelihood (Qml) estimators have higher standard deviations and are more inefficient.When the linearity assumption is violated we see from tables [4.16]-[4.18] that the Mewf and the Qml estimators are severely unconditionally biased and very inefficient.
We can therefore conclude from this results that adding more terms to the adjustment factor does not reduce the bias as expected but as expected the efficiency is decreased.

In the next section we describe the second simulation study we carried out,to study the effect of varying the nonlinear and correlation structures and the sample size on the performance of the estimators.

4.4 SIMULATION STUDY 2

Demets and Halperin(1977) carried out a simulation study in which they generated their survey and design variables from a standard trivariate normal distribution with zero means and unit variances.They compared the performance of the ordinary least squares estimators and the maximum likelihood estimators with varying correlations and sample sizes for a design in which the extremes on z are deliberately choosen i.e U-shaped design.Our simulation study is similar to theirs .We generated 10,000 finite population values of $(\tilde{y}_i, z_i)$ i=1....10,000 where $\tilde{y}_i=(y_{1i},y_{2i})$ and z_i are the survey variables and the design variables respectively from a joint multivariate normal distribution with unit variance and mean equal to 5.We took the mean equal to 5 arbitrarily to avoid generating negative values. Since our objective is to study the efficiency and robustness properties of the estimators we need to generate a population whereby the linear homoscedastic model assumptions are violated.Holmes (1987) as in study 1 generated such a population by generating the design variable from a uniform distribution and then the survey variables were obtained from relationships linking them with the design variable. In this study we will generate linear homoscedastic variables and then introduce nonlinearity and heteroscedasticity in such a way that we are able to control the degree of the violation of the two assumptions.This we do by transforming the three variables using a power series transformation.

Let

$$V_{y1i}=y_{1i}^{\lambda_1},$$

$$V_{y2i}=y_{2i}^{\lambda_2},$$

and

$$V_{zi}=z_i^{\lambda_3}.$$

be the transformed variables.

Since y_i and z are normal variables then under the normality model 1 we assume that,

$$E_1(\tilde{y}_i|z)=m+Hz,$$

and

$$V_1(\tilde{y}_i|z)= K,$$

where K,m and H are constants.

Define the transformed variables as

$$V_y=y^{\lambda},$$

and

$$V_z=z^{\lambda_3}.$$

We will now show that the transformed variables V_y are nonlinear in z and heteroscedastic if $\lambda\neq1$.

Let V_y and V_z denote the realized values of the random variables $\tilde{V}_y$ and $\tilde{V}_z$ respectively.

Now the conditional distribution of $\tilde{V}_y$ given V_z is given by

$$f(V_y|V_z)=\frac{f(V_y,V_z)}{f(V_z)}=\frac{f(y,z)\ |J_1|}{f(z)\ \ |J_2|} = f(y|z)\ |J_1|/|J_2| \quad [4.1]$$

where J_1 and J_2 are the Jacobians of the corresponding transformations.

Evaluating the Jacobians we get

$$|J_1| = 1/\ (\lambda\ \lambda_3\ y^{\lambda-1}\ z^{\lambda_3-1})\ ,$$

and

$$|J_2| = 1/\ (\ \lambda_3\ z^{\lambda_3-1}\ ).$$

Substituting these Jacobians in [4.1] we get

$$f(V_y|V_z)=f(y|z)/\lambda y^{\lambda-1}. \quad [4.2]$$

We can now use the conditional distribution in [4.2] to derive the conditional expectations and variances of $\tilde{V}_y$.

Now under model 1 the conditional expectation of $\tilde{V}_y$ given V_z is given by

$$E_1(\tilde{V}_y\ |V_z)=\int V_y\ f(V_y|V_z)dV_y$$

$$= E_1(\tilde{y}^{\lambda}|z),\text{using [4.2]}. \quad [4.3]$$

Example

We now derive the conditional expectations of $\tilde{V}_y$ given V_z and the conditional variance for a particular value of λ i.e let $\lambda=2$.

Using [4.3] we have

$$E_1(\tilde{V}_y|V_z)= E_1(\tilde{y}^2|z)$$

$$=V_1(\tilde{y}|z)+ [E_1(\tilde{y}|z)]^2$$
$$=K+[M+Hz]^2. \qquad [4.4]$$

and

$$V_1(\tilde{V}_y|V_z)= E_1(\tilde{V}_y^{\,2}|V_z) \;\tilde{-}[E_1(V_y\;|V_z)]^2$$

$$=E_1(\tilde{V}_y^{\,2}|V_z) \;-[K+[M+Hz]^2]^2$$
$$=E_1(\tilde{y}^4|z) \;-[K+[M+Hz]^2]^2. \qquad [4.5]$$

using (4.4)

We see from eq (4.4) and (4.5) that the transformed variables V_y are nonlinear in z and are heteroscedastic i.e their conditional variance depends on z.By varying the value of λ and λ_3 we can either increase or decrease the degree of nonlinearity.

Note that when $\lambda=1$ then

$$E_1(\tilde{V}_y\;|V_z) = E_1(\tilde{y}|z) \quad \text{and } E_1(\tilde{V}_y\;|V_z)=V_1(\tilde{y}|z)=K.$$

Thus for this value of λ V_y is linear and homoscedastic.We have therefore shown that using the power series transformation we can control the degree of violation of the linear homoscedastic model assumptions.

PARAMETER VALUES.

The correlation matrix of the survey variables $\underset{\sim}{\tilde{y}}$ and design variables $\underset{\sim}{\tilde{z}}$ is given by

$$\rho = \begin{bmatrix} 1 & & \\ \rho_{12} & 1 & \\ \rho_{1z} & \rho_{2z} & 1 \end{bmatrix}$$

Nathan and Holt(1980) have shown that the bias of the model based estimators when the Normality assumptions are violated is determined by the correlations ρ_{12},ρ_{2z} and $\rho_{1z.2}$.

Therefore to generate the observations we chose arbitrarily three values in the range 0.0 to 0.99 as our values taken by ρ_{12}, ρ_{2z} and $\rho_{1z.2}$. Since we need the correlation ρ_{1z} to generate the correlation matrix ρ and also the regression coefficent between $\underset{\sim}{y}_1$ and $\underset{\sim}{z}$ we use the following relationship to compute it.

$$\rho_{1z} = \rho_{1z.2} \sqrt{(1-\rho_{12}^2)(1-\rho_{2z}^2)} + \rho_{12}\rho_{2z} . \qquad [4.6]$$

We arbitrarily chose (0.1,0.5,0.9) labelled (C1,C2,C3) as the values conceived by the correlation coefficients and consider all the 3^3 possible combinations. Since our main objective is to study the performance of the six estimators and to study whether changes in the correlation structure lead to predictable changes in the estimators we need to set up an experiment such that it will be possible to study the effect of the changes of these parameter values and their interactions on the performance of the estimators independently of each other.This can be achieved by choosing the set of parameter values (levels) to be taken by each parameter(factor) to be studied and then compute such measures of performance like the mean square error for all the estimators.Since in our experiment we have considere d all the 3^3 possible combinations of thetreatments(correlation structure),which occurs at least once in each trial,our experiment is a *complete factorial experiment*.We will assume that the three factor interaction is zero and use its mean square error to test the main effects and the two factor interactions.We also choose arbitrary some values to be taken by the three powers of the transformation as (0.5,1,2) and considered all the 3^3 possible combinations.To study the effect of varying the nonlinear structure,since the scale of measurement across different nonlinear combinations is not constant,it is not possible to study the effect on the performance of the estimators across different nonlinear structures using the factorial design set up.We will therefore present the relative mean square errors of the six estimators relative to the probability weighted estimators for all the twenty seven nonlinear combinations.

FACTORIAL EXPERIMENT: SETUP

Let A_1, A_2 and A_3 be the three factors under study. For each of these three factors we decided to have three levels. So A_{1i}, A_{2j} and A_{3k} will denote the factors A_1, A_2 and A_3 at the i^{th}, j^{th} and k^{th} level respectively. Now the factorial experiment setup with the response R_{ijkl} in a trial with A_1 at the i^{th} level, A_2 at the j^{th} level and A_3 at the k^{th} level for the l^{th} trial of the treatment combination is given by

$$R_{ijkl}=\mu+A_{1i}+A_{2j}+A_{3k}+D_{ij}+E_{jk}+F_{ik}+G_{ijk}+\varepsilon_{ijkl},$$

where

μ denotes the true means of the treatment combinations,

A_{ti} denotes the true means of the treatment combinations in which $A_t (t=1,2,3)$ is at the i^{th} level,

D_{ij} is the interaction of A_{1i} and A_{2j} ,

E_{jk} is the interaction of A_{2j} and A_{3k},

F_{ik} is the interaction of A_{1i} and A_{3k} ,

G_{ijk} is the interaction of A_{1i} , A_{2j} and A_{3k}

and

ε_{ijkl} is the error term.

We studied the performance of the six estimators under a fixed correlation and nonlinear structure, for different sample sizes. i.e we considered the correlation structure (0.1,0,5,0.9) and the non linear structure (0.5,1,2) for the sample sizes 25,50,100,200,400, 800,1000, and 1200. We then varied the nonlinear structure (0.5,1,2) and considered all the 3^3 possible combinations for fixed sample size n=400 and fixed correlation structure (0.1,0.5,0.9). Lastly we varied the correlation structure (0.1,0.5,0.9) and considered all the 3^3 possible combinations for fixed sample size n=400 and fixed nonlinear structure(2,2,2).

The 27 different combinations of (ρ_{12} $\rho_{1z.2}$,ρ_{2z}) which we considered are:

	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}	M_{13}
ρ_{12}	0.9	0.1	0.5	0.9	0.9	0.1	0.1	0.9	0.1	0.9	0.9	0.5	0.5
$\rho_{1z.2}$	0.9	0.1	0.5	0.1	0.9	0.9	0.1	0.1	0.9	0.5	0.9	0.9	0.5
ρ_{2z}	0.9	0.1	0.5	0.9	0.1	0.9	0.9	0.1	0.1	0.9	0.5	0.9	0.9

	M_{14}	M_{15}	M_{16}	M_{17}	M_{18}	M_{19}	M_{20}	M_{21}	M_{22}	M_{23}	M_{24}	M_{25}
ρ_{12}	0.5	0.9	0.1	0.5	0.5	0.5	0.1	0.1	0.1	0.5	0.5	0.9
$\rho_{1z.2}$	0.9	0.5	0.5	0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.9	0.5
ρ_{2z}	0.5	0.5	0.5	0.5	0.1	0.1	0.1	0.5	0.9	0.9	0.1	0.1

	M_{26}	M_{27}
ρ_{12}	0.9	0.1
$\rho_{1z.2}$	0.1	0.9
ρ_{2z}	0.5	0.5

The 27 different combinations of (λ_1,λ_2,λ_3) which we considered are:

	K_1	K_2	K_3	K_4	K_5	K_6	K_7	K_8	K_9	K_{10}	K_{11}	K_{12}	K_{13}
λ_1	0.5	1	2	2	2	0.5	0.5	2	0.5	2	2	1	1
λ_2	0.5	1	2	0.5	2	2	0.5	0.5	2	1	2	2	1
λ_3	0.5	1	2	2	0.5	2	2	0.5	0.5	2	1	2	2

	K_{14}	K_{15}	K_{16}	K_{17}	K_{18}	K_{19}	K_{20}	K_{21}	K_{22}	K_{23}	K_{24}	K_{25}
λ_1	1	2	0.5	1	1	1	0.5	0.5	0.5	1	1	2
λ_2	2	1	1	0.5	1	0.5	1	0.5	1	0.5	2	1
λ_3	1	1	1	1	0.5	0.5	0.5	1	2	2	0.5	0.5

	K_{26}	K_{27}
λ_1	2	0.5
λ_2	0.5	2
λ_3	1	1

4.4.1 RESULTS

We present unconditional results for the U-shaped sample design.

TABLE 4.19 (fixed nonlinear and correlation structure)

UNCONDITIONAL MEAN SQUARE ERRORS OF THE SIX ESTIMATORS OVER 1,000 REPLICATIONS FOR U-SHAPED DESIGN.

SAMPLE	MSE OF THE ESTIMATORS					
SIZE	$\hat{\beta}_{12,ols}$	$\hat{\beta}_{12,pw}$	$\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,pwml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,wf}$
25	0.037	0.071	0.037	0.070	0.041	0.065
50	0.017	0.034	0.017	0.034	0.018	0.031
100	0.008	0.018	0.008	0.018	0.008	0.017
200	0.004	0.009	0.004	0.009	0.004	0.008
400	0.002	0.005	0.002	0.005	0.002	0.005
800	0.001	0.002	0.001	0.002	0.001	0.002
1000	0.001	0.002	0.001	0.002	0.001	0.002
1200	0.001	0.002	0.001	0.002	0.001	0.002

TABLE 4.20 (fixed nonlinear and sample size=400)

UNCONDITIONAL MEAN SQUARE ERRORS OF THE SIX ESTIMATORS OVER 1,000 REPLICATIONS FOR U-SHAPED DESIGN FOR 3^3 NONLINEAR COMBINATIONS.

Nonlinear Combination	RELATIVE MSE($\hat{\beta}_{12}$)=MSE($\hat{\beta}_{12}$)/ MSE($\hat{\beta}_{12,pw}$)				
	$\hat{\beta}_{12,ols}$	$\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,pwml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,wf}$
K1	2.4220	0.1938	0.9846	0.3608	0.9843
K2	2.4813	0.1636	0.9862	0.2933	0.9867
K3	2.1830	0.1691	0.9879	0.3483	0.9877
K4	1.8620	0.2299	0.9879	0.3847	0.9825
K5	2.1830	0.1773	0.9920	0.3563	0.9916
K6	1.8620	0.2400	0.9952	0.6061	0.9870
K7	2.4220	0.2202	0.9874	0.4044	0.9879
K8	1.7480	0.2324	0.9930	0.6768	0.9877
K9	2.3046	0.1925	0.9860	0.3745	0.9837
K10	2.1830	0.1724	0.9899	0.3171	0.9904
K11	1.8820	0.1949	0.9914	0.5321	0.9884
K12	2.4813	0.1715	0.9867	0.3732	0.9872
K13	1.8820	0.1961	0.9896	0.3455	0.9876
K14	2.3046	0.2019	0.9896	0.3807	0.9869
K15	2.4637	0.2113	0.9856	0.3694	0.9849
K16	2.2450	0.1691	0.9872	0.3116	0.9864
K17	2.4813	0.1647	0.9866	0.3278	0.9868
K18	2.2450	0.1717	0.9884	0.3767	0.9868
K19	2.4640	0.2041	0.9842	0.3680	0.9839
K20	2.4220	0.1976	0.9853	0.3481	0.9852
K21	2.4637	0.2415	0.9883	0.4636	0.9876
K22	2.2429	0.1794	0.9860	0.3456	0.9857
K23	2.2452	0.1694	0.9869	0.3476	0.9867
K24	1.8820	0.1954	0.9892	0.3323	0.9871
K25	2.3046	0.2124	0.9927	0.4750	0.9887
K26	1.8620	0.2364	0.9918	0.4571	0.9854
K27	1.7479	0.2283	0.9899	0.4243	0.9855

UNCONDITIONAL RESULTS OF THE MEAN SQUARE ERROR OF THE SIX ESTIMATORS WHEN THE CORRELATION STRUCTURE IS VARIED.

Fixed sample size n=400,fixed nonlinear structure (1,1,1) ond different combinations of the correlation structure (0.1,0.5,0.9)

TABLE 4.21 ESTIMATOR:Ordinary least squares.
CRITERIA:mean square error.

ANALYSIS OF VARIANCE TABLE.

UNITS	D.F	F-RATIO
C1	2	0.338
C2	2	1.141
C3	2	0.039
C1.C2	4	0.535
C1.C3	4	1.470
C2.C3	4	0.779
RESIDUAL.	8	
	26	

* means result is significant at 5% level of significance.

TABLE 4.22 ESTIMATOR:probability weighted
CRITERIA:mean square error.

ANALYSIS OF VARIANCE TABLE.

UNITS	D.F	F-RATIO
C1	2	0.443
C2	2	1.161
C3	2	0.090
C1.C2	4	0.434
C1.C3	4	1.355
C2.C3	4	0.883
RESIDUAL.	8	
Total	26	

* means result is significant at 5% level.

TABLE 4.23 ESTIMATOR:maximum likelihood
CRITERIA:mean square error.

ANALYSIS OF VARIANCE TABLE.

UNITS	D.F	F-RATIO
C1	2	0.618
C2	2	1.089
C3	2	0.112
C1.C2	4	0.587
C1.C3	4	2.436
C2.C3	4	0.975
RESIDUAL.	8	
TOTAL	26	

* means the result is significant at 5% level.

TABLE 4.24 ESTIMATOR:Probability weighted maximum likelihood.
CRITERIA:mean square error.

ANALYSIS OF VARIANCE TABLE.

UNITS	D.F	F-RATIO
C1	2	0.806
C2	2	1.111
C3	2	0.180
C1.C2	4	0.408
C1.C3	4	2.333
C2.C3	4	1.031
RESIDUAL.	8	
Total	26	

* means result is significant at 5% level

TABLE 4.25 ESTIMATOR:equally weighted Fuller.
CRITERIA:mean square error.

ANALYSIS OF VARIANCE TABLE.

UNITS	D.F	F-RATIO
C1	2	0.776
C2	2	1.139
C3	2	0.200
C1.C2	4	0.435
C1.C3	4	2.465
C2.C3	4	1.070
RESIDUAL.	8	
TOTAL	26	

* means result is significant at 5% level.

TABLE 4.26 ESTIMATOR:weighted Fuller.
CRITERIA:mean square error.

ANALYSIS OF VARIANCE TABLE.

UNITS	D.F	F-RATIO
C1	2	0.603
C2	2	1.078
C3	2	0.128
C1.C2	4	0.569
C1.C3	4	2.448
C2.C3	4	0.924
RESIDUAL.	8	
TOTAL	26	

* Means that result is significant at 5% level.

We see from table [4.19] that the weighted estimators i.e probability weighted, probability weighted adjusted and weighted Fuller regression estimators are very inefficient when the sample size taken is small.The equally weighted Fuller ,maximum likelihood and the ordinary least squares estimators are more efficient across all the sample sizes than the weighted estimators.All the six estimators becomes more efficient as the sample size is increased.
Tables [4.20] gives the relative mean square errors of all the estimators with respect to the probability weighted estimator.We see that the maximum lkelihood estimator and the equally weighted Fuller estimator are more efficient across all the nonlinear combinations than the design based estimators and the ordinary least squares estimator.The ols estimator is the least efficient across all the nonlinear combinations among the estimators considered.We also see that there does not seem to be any significant gain in efficiency using the Pwml or the Wf estimators over the probability weighted estimator.The performance across all the nonlinear combinations for the design based estimators varies very little,hence these estimators are robust to the variation of the nonlinear structure.We note that in the case of the maximum likelihood and equally weighted Fuller estimators ,their performance depends on the type of nonlinear combination.For the linear homoscedastic population i.e with the nonlinear combination (1,1,1) the performance of these estimators is optimal.Under any other nonlinear combination,the performance decreases.We can therefore conclude that the Ml and Ewf estimators are more sensitive to the variation of the nonlinear structure compared to the design based estimators.
Tables [4.21]-[4.26] gives the analysis of variance tables for the mean square errors of all the six estimators when the correlation structure of the population is varied.We see that across all the six estimators varying the correlation structure does not have a significant effect on the performance of the six estimators.
We can therefore conclude that the estimators are robust to

the variations of the correlation structure but are not robust to the variations of the degree of nonlinearity of the variables and the variations of the sample sizes.

4.5 CONCLUSION

In this chapter we have confirmed the asymptotic properties of the Fuller estimators empirically which we had proved theoretically in chapter 3.We found that if the population is linear and homoscedastic then the maximum likelihood estimator is the most efficient estimator across all the designs. We also found that there was no significant gain in efficiency using the weighted Fuller and the probability weighted adjusted estimator over the probability weighted estimator.When the linearity assumption is violated then across all the eight designs the probability weighted estimator was found to be the most efficient estimator for increasing allocation designs,Maximum likelihood estimator is the most efficient for the U-shaped designs and the equally weighted Fuller compromises the efficiency properties of the Pw estimator and the Ml estimator across the eight design,i.e the equally weighted Fuller estimator is more efficient than the probability weighted estimator for the U-shaped designs and also more efficient than the Maximum likelihood estimator for the increasing allocation designs when the linearity assumption is violated.This good performance of the probability weighted estimator for these type of designs was expected.This is because as Smith (1988) showed,the increasing allocation designs approximates size biased sampling and for this sampling scheme the probability weighted estimator is a method of moment model based estimator,hence is the appropriate estimator to use for these sampling designs. When only the Homoscedastic assumption is violated then we found that with exception of the ols estimator which was found to be more efficient than the ewf estimator, the equally weighted Fuller estimator is the most efficient estimator across all the designs and the design based estimators are more efficient than the model based estimators.Since the weighted Fuller estimator is approximately conditional unbiased when the population is linear and homoscedastic and also when the homoscedastic assumption is violated it has an advantage over the probability weighted and probability weighted adjusted estimators which were found to be conditional biased when the homoscedastic assumption is violated.For all the six estimators considered across the three models we found that they are approximately unconditional unbiased for the equal probability designs,i.e they are robust for this design.This result is expected because this sampling scheme approximates balanced sampling scheme,the use of balanced sampling permits the use of all the estimators because they are asymptotically unbiased within this class of designs.The disadvatage with balanced sampling as a criteria for robustness is that more efficient sample designs are excluded from this class.For example,increasing allocation design which is efficient for probability weighted estimators and the U-shaped designs which is efficient for the maximum likelihood estimator.see Royall and Herson (1973a). We also found that the estimators studied in this chapter are robust to the change in the correlation structure of the population ,the model based estimators and the equally weighted Fuller are very sensitive to the change in the nonlinear structure while the design based estimators are robust.From these empirical results we can conclude that weighting parametric estimators to introduce robustness (see Nathan and Holt (1980)) or increasing the number of terms in the adjustment terms like in Modified equally weighted Fuller and Quadratic estimators does not help us to get a robust efficient estimators which is satisfactory in all circumstances.Thus all the model based parametric procedures considered are not robust in either the conditional or the unconditional distribution framework,the design based procedures are robust in the unconditional distribution framework but are not robust in the conditional distribution framework and the weighted version of model based procedures are not robust in the conditional distribution framework.

In the remaining chapters we will deviate from the common ground where most authors have concentrated their efforts searching for robust efficient estimators that is,trying to compromise the good robustness properties of the weighted estimators and the good efficiency properties of the equally weighted estimators.We will derive estimators of the means and variance without making any distributional assumptions, thus bringing credibility to the estimators,since we have avoided the often questionable distributional assumptions.

CHAPTER 5

NONPARAMETRIC ESTIMATION OF REGRESSION COEFFICIENTS

In the preceding chapters,we found that model based procedures are not robust and the design based procedures were found to be robust unconditionally,inefficient in some cases and not robustconditionally.Attempts by various authors to introduce robustness in model based procedures by incorporating weights in them was not very successful, as we foundin our empirical studies that these weighted model based procedures,though robust unconditionally,they are not robust conditionally and were not efficient in some cases,see Nathan and Holt (1980).We therefore need procedures which are efficient and robust both conditionally and unconditionally.This chapter is devoted to the search of such procedures.

5.1 INTRODUCTION

Let the survey variables $\tilde{y}_i$ and design variables z_i be identically and independently distributed as $\tilde{y}$ and $\tilde{z}$ random variables respectively, where $\tilde{y}=(\tilde{y}_1 \ldots \tilde{y}_p)$ denote a px1 vector of survey variables and $\tilde{z}=(\tilde{z}_1 \ldots \tilde{z}_q)$ a qx1 vector of design variables.In chapter 1 equation (1.16) we defined the joint distribution of the survey variables $\tilde{y}$ and design variables $\tilde{z}$ indexed by the parameters $\theta=(\lambda,\phi)$ as;

$$f(y,z;\lambda,\phi)=f(y|z;\lambda)f(z;\phi).$$

After selecting a sample using a known sampling scheme

$f(s|\underset{\sim}{z})$ the survey variables observed in the sample are denoted by $\underset{\sim}{y}_s$ and the joint distribution of $\tilde{y}_s$,s and $\tilde{z}$ is given by;

$$f(\underset{\sim}{y}_s,\underset{\sim}{z},s;\underset{\sim}{\lambda},\underset{\sim}{\phi})=f(s|\underset{\sim}{z})f(\underset{\sim}{y}_s|\underset{\sim}{z};\underset{\sim}{\lambda})f(\underset{\sim}{z};\underset{\sim}{\phi}). \quad \text{see equation (1.17)}$$

From the conditional distribution $f(\underset{\sim}{y}_s|\underset{\sim}{z};\underset{\sim}{\lambda})$ we can estimate $\underset{\sim}{\lambda}$ and $\underset{\sim}{\phi}$ from the marginal distribution $f(\underset{\sim}{z};\underset{\sim}{\phi})$.

The parameter of interest is $\underset{\sim}{\theta}$ in the marginal distribution of $\tilde{y}$.This parameter is not directly observable from the data,but is a function of $\underset{\sim}{\lambda}$ and $\underset{\sim}{\phi}$ .The problem facing us is how to use the conditional distribution $f(\underset{\sim}{y}_s|\underset{\sim}{z};\underset{\sim}{\lambda})$ and the marginal distribution $f(\underset{\sim}{z};\underset{\sim}{\phi})$ to make inference about this parameter $\underset{\sim}{\theta}$.When the distributions are known,we can derive $\underset{\sim}{\theta}$ as a function of $\underset{\sim}{\lambda}$ and $\underset{\sim}{\phi}$ and then using some criterion,e.g maximum likelihood (Ml) estimation criterion,we can estimate $\underset{\sim}{\theta}$ given the Ml estimates of $\underset{\sim}{\lambda}$ and $\underset{\sim}{\phi}$ .When the distributions are of unknown form ,we can derive moments under certain conditions,for example under the assumption of linearity and homoscedasticity conditions,see chapter 1.But if the assumptions of linearity and homoscedasticity are not satisfied,then a model based procedure which is robust to the violation of these assumptions is not available.Of course we can use the randomization distribution which we know will be robust unconditionally but give poor results conditionally.

In this chapter,we propose a model based nonparametric procedure which overcomes many of the difficulties encountered with the other model based estimators.The results are established both theoretically in this chapter and empirically in chapter 6.

We will first consider the case where we have one design and survey variable,i.e p=1 and q=1.At a later stage we will generalize the results obtained to the case where p>1 and q=1.

In chapter 1 we formulated a parametric model in which the functional forms of the conditional expectations and variances were assumed known but with a finite set of unknown parameters. We generalize this parametric model by writing it in the more general form;

$$E(\tilde{y}|z)=\mu(z), \qquad [5.1]$$

and

$$V(\tilde{y}|z)=\sigma^2(z) \; . \qquad [5.2]$$

The conventional approach to this problem assumes that the functions $\mu(z)$ and $\sigma^2(z)$ are known functions of z up to some finite set of unknown parameters (see chapter1).In our previous chapters we followed this approach and found that the estimators derived in this approach by assuming that the survey and design variables have a joint multivariate normal distribution,are not robust to the mispecification of the functional forms $\mu(z)$ and $\sigma^2(z)$. This thesis is concerned with the search for robust estimators,so in order to protect estimators against model mispecification we now suppose that the functions $\sigma^2(z)$ and $\mu(z)$ are not known but are smooth.Then how do we carry out the estimation of the means and the variances of the survey variables?.

Throughout this chapter we will use the integral sign $\int$ to denote $\int_{-\infty}^{+\infty}$.

Now the mean of the marginal distribution of $\tilde{y}$ denoted by μ_y is given by

$$\mu_y=E(\tilde{y})=E_z(E(\tilde{y}|z))$$
$$=\int E(\tilde{y}|z)f(z)dz$$
$$=\int \mu(z)f(z)dz, \quad \text{using [5.1]}. \qquad [5.3]$$

A possible estimator of this mean is given by;

$$\hat{\mu}_y=\int\hat{\mu}(z)\hat{f}(z)dz, \qquad [5.4]$$

where $\hat{\mu}(z)$ is an estimator of $\mu(z)$ and $\hat{f}(z)$ is an estimator of $f(z)$.

The variance of $\tilde{y}$ denoted by σ_y^2 is given by;

$$\sigma_y^2=V(\tilde{y})=E_z(V(\tilde{y}|z))+V_z(E(\tilde{y}|z))$$

$$=\int V(\tilde{y}|z)f(z)dz+\int(E(\tilde{y}|z)-E(\tilde{y}))^2f(z)dz$$

$$=\int\sigma^2(z)f(z)dz+\int(\mu(z)-\mu_y)^2f(z)dz,$$

using [5.1],[5.2] and [5.3].

[5.5]

A possible estimator of this variance is then;

$$\hat{\sigma}_y^2=\int\hat{\sigma}^2(z)\hat{f}(z)dz+\int(\hat{\mu}(z)-\hat{\mu}_y)^2\hat{f}(z)dz, \qquad [5.6]$$

where $\hat{\sigma}^2(z)$ is an estimator of $\sigma^2(z)$.

To be able to evaluate [5.4] and [5.6] we need to estimate the functions $\mu(z)$,$\sigma^2(z)$ and $f(z)$.Parametrically this is not possible if we do not know the forms of these functions. We therefore resort to nonparametric methods to solve our problem. In the next section we propose nonparametric estimates of $\mu(z)$,$\sigma^2(z)$ and $f(z)$.

5.2 NONPARAMETRIC METHOD OF ESTIMATION

5.2.1 GENERAL NONPARAMETRIC ESTIMATES OF $\mu(z)$,$\sigma^2(z)$ AND $f(z)$

There are various nonparametric methods for estimating functions $\mu(z)$ and $\sigma^2(z)$.In this work we will construct estimators based on linear smoothing methods.Examples of some linear smoothing methods are kernel estimation (Gasser and Muller[1979];Nadaraya[1964]; Watson [1964] Priestley and Chao [1972]),local regression (Cleveland [1979];Cleveland and Devlin [1988];Friedman and Stuetzle [1982];Muller [1987]) smoothing splines (Craven and Wahba [1979];Hutchinson and DeHoog [1985]; Silverman [1985]). We present below general smooth estimates of $\mu(z)$, $\sigma^2(z)$ and $f(z)$.

(i) NONPARAMETRIC ESTIMATE OF f(z)

We assumed that the design variables $\tilde{z}_1, \tilde{z}_2, ., \tilde{z}_N$ are identically and independently distributed as random variable $\tilde{z}$ with probability density $f(z)$. A nonparametric estimate of $f(z)$ at a given point z is given by (see Parzen[1962]),

$$d\hat{F}(z)=\hat{f}(z)dz=\begin{cases} 1/N & \text{if } z=z_1,..,z_N, \\ 0 & \text{otherwise.} \end{cases} \qquad [5.7]$$

Since we assumed that the whole finite population of z is known we shall not attempt to construct a smooth estimate of $f(z)$ and [5.7] will be a good estimate to use in the estimators [5.4] and [5.6].
Substituting [5.7] in [5.4] and [5.6] we get the estimators of the mean and variance as;

$$\hat{\mu}_y=N^{-1}\sum_U \hat{\mu}(z_i),$$

and

$$\hat{\sigma}_y^2=N^{-1}[\sum_U \hat{\sigma}^2(z_i) + \sum_U(\hat{\mu}(z_i)-\hat{\mu}_y)^2].$$

We now need to construct smooth estimators of $\sigma^2(z)$ and $\mu(z)$.

(ii) SMOOTH LINEAR ESTIMATE OF $\mu(z)$

Most of the authors seem to have focussed their attention on the estimation of $\mu(z)$ (see Nadaraya [1964], Watson [1964] Priestley and Chao[1972]).
A smooth linear estimate of a function $\mu(z)$ denoted by $\hat{\mu}(z)$ can be written in general form as;

$$\hat{\mu}(z)=\sum_{j\in s} W_k(z,z_j)y_j. \qquad [5.8]$$

where $W_k(z,z_j)$ denotes a smoothing function with a bandwidth parameter k. This bandwidth parameter, also called the window span by some authors, determines the amount of smoothing to be done.For $W_k(z,z_j)$ to be a smoothing function it is assumed that it is Lipschitz continous(condition for a function to be smooth).If the sum of the weights $W_k(z,z_j)$ is equal to one then [5.8] is a weighted average of the y values.Eubank[1988] has reviewed various linear smooth estimates of $\mu(z)$ proposed in the literature. Using the integrated mean square error as a criterion for comparing the performance of different linear smooth estimates,Gasser and Engel [1990] found that none of the linear smooth estimates is uniformly best,but kernel estimates were found to have minimax optimal properties.In this work we will consider the estimates proposed by Nadaraya[1964] and Watson[1964] and by Priestley and Chao[1972] associated with the kernel function,which of late have attracted the attention of most authors.These two estimates have recently been studied extensively by Benedetti [1974],[1975] and [1977],Schuster and Yakowitz [1979],Gasser and Engel [1990]. Let us point out here that we are using the kernel estimates in a rather different context than that used by most authors.Most authors dealing with this problem of curve estimation,use observations recorded in experiments,and use these observations to estimate the curve $\mu(z)$.For example in the fitting of Raman spectra,which is a standard smoothing procedure in physical chemistry to determine location and height of spectral bands.(Hardle and Gasser [1984]).In our case we regard the observations as a sample drawn from a finite population using a known sampling scheme.Thus our study is an extension of nonparametric procedures to the estimation of regression coefficients in the analysis of complex surveys.

(i) PRIESTLEY AND CHAO (PC) ESTIMATE OF $\mu(z)$

Priestley and Chao[1972] proposed a smooth linear estimate of $\mu(z)$ denote it by $\hat{\mu}_{pc}(z)$;

$$\hat{\mu}_{pc}(z)=k^{-1}\sum_{S}(z_j-z_{j-1})w((z-z_j)/k)y_j, \qquad [5.9]$$

where they assumed that $0\le z_1\le z_2 \le z_3 .....\le z_n\le 1$. and $z_0=0$. k denotes the bandwidth parameter which determines the amount of smoothing to be done, w is called the kernel function with the following properties;

$$\left.\begin{array}{ll} \text{(i)} & w(t)\ge 0 \text{ for all } t. \\ \text{(ii)} & \int_{-\infty}^{\infty} w(t)dt=1. \\ \text{(iii)} & \int_{-\infty}^{\infty} w^2(t)dt<\infty. \end{array}\right\} \qquad [5.10]$$

This estimator is a generalization to regression of kernel methods introduced for density estimation by Rosenblatt[1956] and Parzen [1962].The factor z_j-z_{j-1} is considered more appropriate if the data are not equally spaced.If the data are equally spaced then this factor is equal to n^{-1} and for random unequally spaced data it is approximtely equal to n^{-1} (see Priestley-Chao [1972].We see from [5.9] that the sum of weights associated with the y values is not equal to one.This becomes an important issue at the endpoints of the range of the functions being estimated. Thus strictly speaking the PC estimator is not a weighted average of the y values.
However for large n and with random spacings of the design variables,

$$\sum_S W_k(z,z_j)=\sum_S(z_j-z_{j-1})w((z-z_j)/k)k^{-1} \sim \int_{-\infty}^{\infty} w(t)dt=1 \quad \text{as } n\Rightarrow\infty.$$

(ii) <u>NADARAYA-WATSON SMOOTH ESTIMATE OF $\mu(z)$</u>

Nadaraya[1964] and Watson[1964] independently proposed the following estimate of $\mu(z)$ denote it by $\hat{\mu}_{nw}(z)$,

$$\hat{\mu}_{nw}(z)=[\sum_S w((z-z_j)/k)y_j]/\sum_S w((z-z_j)/k). \qquad [5.11]$$

where k denotes the bandwidth parameter and the properties of kernel function w in this estimate are the same as those given for the kernel function in [5.10].For this estimator the sum of weights are equal to 1,hence [5.11] is a weighted average of the observations.

(iii) SMOOTH LINEAR ESTIMATE OF $\sigma^2(z)$

Not much has been done in the estimation of the variance function $\sigma^2(z)$.Most authors have assumed that this variance function is constant.Several authors have estimated this constant variance function to use in the construction of interval estimates of $\mu(z)$,see Rice [1984],Gasser,Sroka and Jennen-steinmetz [1986],Hall and Marron [1990] etc.In our case we assume that this variance function is not constant but a function of z. Following Hall and Marron [1990] we will propose a smooth estimate of this nonconstant $\sigma^2(z)$ to use in the construction of the estimator of the variance of $\tilde{y}$.

To derive a smooth linear estimate of $\sigma^2(z)$ we may reformulate the model given by [5.1] and [5.2] to get;

$$y=\mu(z)+\varepsilon,$$

where $E(\varepsilon|z)=0$,

and

$$V(\varepsilon|z)=\sigma^2(z).$$

For y_j and $\mu(z_j)$, $j\in s$ this model becomes

$$\left.\begin{aligned} &y_j=\mu(z_j)+\varepsilon_j, \\ &\text{where } E(\varepsilon_j|z_j)=0 \\ &\text{and} \\ &V(\varepsilon_j|z_j)=\sigma^2(z_j). \end{aligned}\right\} \qquad [5.12]$$

Also assume that $E(\varepsilon_j^4|z_j)=k_4(z_j)$.

Now the estimate of the residual term is given by

$$\hat{\varepsilon}_j = y_j - \hat{\mu}(z_j).$$

The square of the estimate of this residual term $\varepsilon_j, j \in s$ is given by;

$$\hat{\varepsilon}_j^2 = (y_j - \hat{\mu}(z_j))^2. \qquad [5.13]$$

To smooth [5.13] we choose a smoother function $W_h^*(z, z_j)$ with a bandwidth parameter h.Using [5.13] we get;

$$\hat{\sigma}^2(z) = \sum_{j \in s} W_h^*(z, z_j)(y_j - \hat{\mu}(z_j))^2. \qquad [5.14]$$

which is a smooth estimate of $V(\tilde{y}|z)$.
The reason why we choose a different smoother function with a different bandwidth parameter from the smoother function of the estimate of the conditional mean $\hat{\mu}(z)$ is because the rates of convergence of the estimates of the mean and variance to their true values may not be the same.Hall and Marron[1990] have shown that the optimal estimation of $\sigma^2(z)$ demands less smoothing than the optimal estimation of $\mu(z)$.We now propose the corresponding PC and NW kernel estimates of $\sigma^2(z)$.

(i) PROPOSED PRIESTLEY AND CHAO (PC) ESTIMATE OF $\sigma^2(z)$

Using [5.14] we propose a corresponding PC estimate of $\sigma^2(z)$ denoted by $\hat{\sigma}_{pc}^2(z)$;

$$\hat{\sigma}_{pc}^2(z) = \sum_{j \in s} h^{-1}(z_j - z_{j-1}) w^*((z - z_j)/h)(y_j - \hat{\mu}_{pc}(z_j))^2. \qquad [5.15]$$

where $w^*(.)$ is a kernel function with bandwidth parameter h whose properties are the same as those of w(.)in [5.10] and z_j's are ordered and lie in the interval (0,1).

(ii) PROPOSED NADARAYA-WATSON SMOOTH ESTIMATE OF $\sigma^2(z)$

Using [5.14] we propose a corresponding NW estimate of $\sigma^2(z)$ denoted by $\hat{\sigma}^2_{nw}(z)$ as;

$$\hat{\sigma}^2_{nw}(z)=[\sum_s w^*((z-z_j)/h)(y_j-\hat{\mu}_{nw}(z_j))^2]/\sum_s w^*((z-z_j)/h), \qquad [5.16]$$

where h denotes the bandwidth parameter and the properties of $w^*(.)$ are same as those in [5.10].

Note that we have used the same notation for the bandwidth parameters for the means and the variances of the two estimators.This does not mean that they are the same,it is only for notational convenience.

Subtituting [5.7],[5.8] and [5.14] in [5.4] and [5.6] we get the smooth linear estimators of the mean and variance of $\tilde{y}$ as;

$$\hat{\mu}_y = N^{-1}\sum_{i\in U}\sum_{j\in s} W_k(z_i,z_j)y_j, \qquad [5.17]$$

and

$$\hat{\sigma}^2_y=N^{-1}\sum_{i\in U}\sum_{j\in s} W^*_k(z_i,z_j)(y_j-\hat{\mu}(z_j))^2+N^{-1}\sum_{i\in U}(\hat{\mu}(z_i)-\hat{\mu}_y)^2. \qquad [5.18]$$

[5.17] and [5.18] are the nonparametric estimators of the mean and the variance of $\tilde{y}$.The advantage of this general representation of the nonparametric estimators is to enable the study of other linear smooth estimators which can be written in the forms [5.8] and [5.14].We get the corresponding PC and NW estimators of the means and variances by substituting the co r responding smoothing functions in the expressions [5.17] and [5.18].

5.3.1 CHOICE BETWEEN THE NADARAYA-WATSON AND PRIESTLEY-CHAO KERNEL ESTIMATORS

A lot of debate is currently going on as to which of

the two kernel estimators i.e NW or PC kernel estimators is preferable.The NW kernel estimator was intuitively motivated as an estimator of a conditional expectation which suggests a context where the design values are realizations of the random design variables (Nadaraya [1964],Watson [1964]).Euback [1988] using this concept derived the NW kernel estimator.The PC estimator is based on the assumption that the design values are fixed,i.e the design values are restricted to some interval.Priestley and Chao [1972] assumed that the design values are ordered and $z \in [0,1]$.The estimator they proposed depends on these assumptions.
Some of the questions which are being raised by various authors like Benedetti [1974],Gasser and Jennen-Steinmetz [1988],Gasser and Engel [1990] etc. are;

(i) Does the NW kernel estimator make sense in the fixed design framework?,and how does it perform relative to its performance in the random design framework?
(ii) Does the PC kernel estimator make sense in the random design framework?,and how does it perform relative to its performance in the fixed design framework?
(iii)How does the performance of these two estimators in both design frameworks compare?

Benedetti [1974] extended the NW kernel estimator to the case where the design values are fixed and from her empirical studies reported that the NW kernel estimator is superior to the PC estimator in terms of the mean integrated square error criterion.Her results were not very surprising.As is well known from density estimation,one of the major drawbacks of the kernel estimators is that they are highly biased at the endpoints of the data.Assuming that both the PC and NW estimators use the same kernel function ,then they have a common factor

$$\sum_{s} W((z-z_j)/k)y_j, \qquad [5.19]$$

see eqns [5.9] and (5.11].
The kernel function $W((z-z_j)/k)$ gives more weight to those

observations y_j associated with z_j closer to z,i.e in the middle of the distribution of z and very small weights to those observations associated with z_j at the end of the data.In the case of the PC estimator [5.19] is multiplied by a term $(z_j-z_{j-1})/k$ which does not depend on the position of z.As z_j nears the endpoints of the distribution the weights given to the observations are very small,hence making the PC estimator asymptotically biased at the endpoints.In the case of the NW estimator,the common factor [5.19] is multiplied by $[\sum_s W((z-z_j)/k)]^{-1}$ which depend on the position of z.In the middle of the distribution of z this quantity is very small and increases as z_j tends towards the ends of the distribution.Thus at the endpoints this factor helps the common factor [5.19] to compensate for the bias of estimating the y_j's associated with the z_j at the enpoints.Benedetti [1974] asserted without proof that the NW estimator may be asymptotically unbiased at the endpoints.

Data collected in complex surveys usually does not impose any restriction on the values of the design values.However we can introduce the restictions on the design values as required in the PC estimator by transforming the data appropriately.If the design values lies in the intervel (0,1) then we can analyse the data using both the PC and the NW kernel estimators.

5.3.2 NONPARAMETRIC SMOOTH ESTIMATOR OF THE REGRESSION COEFFICIENT

Since regression analysis is the most commonly used statistical technique in the analysis of data from complex surveys this thesis focuses attention on the search for a robust efficient estimator of the regression coefficients. Let us partition the px1 dimensional random vector $\tilde{y}$ into two random vectors $\tilde{y}_1$ and $\tilde{y}_2$,such that $\tilde{y}_1$ is a p_1x1 vector of dependent variables and $\tilde{y}_2$ is a p_2x1 vector of

independent variables where $p=p_1+p_2$. In this work we will consider the case where y_1, y_2 and the design variable z are scalar.Extension to the multivariate case where p>2 but q=1 can easily be done.
The partitioned superpopulation mean and variance covariance matrix are given by;

$$\underset{\sim}{\mu}_y = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \text{and} \quad \underset{\sim}{\Sigma}_{yy} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Our interest is in the point estimation of the regression coefficient of the slope, i.e

$$\beta_{12} = \Sigma_{12}\Sigma_{22}^{-1}.$$

As in previous chapters the design variable does not occur explicitly in the regression equation.We assume that the regression of y1 on y2 is approximately linear so that β_{12} is a reasonable parameter to estimate.
Using [5.18] we define the nonparametric estimator of the regression coefficient β_{12} denoted by $\hat{\beta}_{12}$ as;

$$\hat{\beta}_{12} = \hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1} .$$

In the next section we derive the asymptotic properties of the Nadaraya-Watson and the Priestley -Chao type kernel estimators of the mean ,variance and regression coefficient.

5.4 ASYMPTOTIC PROPERTIES OF THE NADARAYA-WATSON AND THE PRIESTLEY-CHAO TYPE ESTIMATORS OF THE POPULATION MEAN, VARIANCE AND REGRESSION COEFFICIENT

Before we investigate the asymptotic properties of the nonparametric estimators of the finite population mean and finite population variance we note that if we consider the mean square error criterion as a measure of performance of the kernel estimates i.e

$$MSE(\hat{\mu}(z))=Var(\hat{\mu}(z))+[E(\hat{\mu}(z))-\mu(z)]^2 ,$$

and

$$MSE(\hat{\sigma}^2(z))=Var(\hat{\sigma}^2(z))+[E(\hat{\sigma}^2(z))-\sigma^2(z)]^2.$$

then since the bias term depend on the unknown functions $\mu(z)$ and $\sigma^2(z)$ it is impossible to calculate this bias unless some assumptions are made about these functions and the kernel functions W and W^*.We therefore assume that it is known apriori that the functions $W, W^*, \mu(z)$ and $\sigma^2(z)$ are "smooth". A function is said to be smooth if it satisfies **Lipschitz conditions** (see Priestley and Chao[1972]).

As a prelude to the derivation of the theorems on the consistency of the kernel estimators we first give the following Lemmas.

From [5.11] and [5.16] we see that we can rewrite the NW kernel estimator of the mean and variance as;

$$\hat{\mu}_{nw}(z)=[\sum_{j\in s} w((z-z_j)/k)y_j]/\sum_{s} w((z-z_j)/k)$$

$$=(nkf(z))^{-1}[\sum_{j\in s} w((z-z_j)/k)y_j][f(z)/(nk)^{-1}\sum_{s} w((z-z_j)/k)]$$

$$=\mu^*_{nw}(z)[f(z)/\hat{f}1(z)]. \qquad [5.20]$$

where $\mu^*_{nw}(z)=(nkf(z))^{-1}\sum_{j\in s} w((z-z_j)/k)y_j$,

$\hat{f}1(z)=(nk)^{-1}\sum_{s} w((z-z_j)/k)$,

f(z) is the marginal density function of z.

and

$$\hat{\sigma}^2_{nw}(z)=[\sum_{j\in s} w^*((z-z_j)/h)(y_j-\hat{\mu}_{nw}(z_j))^2]/\sum_{s} w^*((z-z_j)/h)$$

$$=[(nhf(z))^{-1}[\sum_{j\in s} w^*((z-z_j)/h)(y_j-\hat{\mu}_{nw}(z_j))^2]f(z)/[(nh)^{-1}\sum_{s} w^*((z-z_j)/h)]$$

$$=\sigma^{*2}_{nw}(z)[f(z)/\hat{f}2(z)], \qquad [5.21]$$

where $\sigma^{*2}_{nw}(z)=(nhf(z))^{-1}\sum_{j\in s} w^{*}((z-z_j)/h)(y_j-\hat{\mu}_{nw}(z_j))^2$,

$\hat{f}2(z)=(nh)^{-1}\sum_{s} w^{*}((z-z_j)/h)$.

$f(z)$ is the marginal density function of z.

THEOREM 5.1

Let $\mu(z)$ and $w((z-z_j)/k)$ be smooth functions then the smooth estimate $\hat{\mu}_{nw}(z)$ is a consistent estimate of $\mu(z)$ i.e

$$\lim_{\substack{n\Rightarrow\infty\\ k\Rightarrow 0}} E(\hat{\mu}_{nw}(z))=\mu(z),$$

and

$$\lim_{\substack{n\Rightarrow\infty\\ k\Rightarrow 0\\ nk\Rightarrow\infty}} Var(\hat{\mu}_{nw}(z))=0.$$

PROOF.(see Watson(1964)).

To prove the consistency properties of the NW estimate of the variance,following Watson [1964] we will prove that the last factor in [5.21] tends to unity in probability i.e

$$\hat{f}2(z)/f(z)\xrightarrow{p}1$$

and

$$E(\sigma^{*2}_{nw}(z))\Rightarrow \sigma^2(z) \text{ as } h\Rightarrow 0, n\Rightarrow\infty$$

$$V(\sigma^{*2}_{nw}(z))\Rightarrow 0 \text{ as } nh\Rightarrow\infty, n\Rightarrow\infty\ h\Rightarrow 0.$$

Then combining these two results we deduce the consistency of $\hat{\sigma}^2_{nw}(z)$.

LEMMA 5.1

Let $f(z)$ and $w((z-z_j)/h)$ be smooth functions then the smooth estimate $\hat{f}2(z)$ is a consistent estimate of $f(z)$ i.e

$$\lim_{\substack{n\to\infty \\ k\to 0}} E(\hat{f}2(z))=f(z),$$

and

$$\lim_{\substack{n\to\infty \\ k\to 0 \\ nk\to\infty}} Var(\hat{f}2(z))=0.$$

PROOF.(see Tapia and Thompson [1978] pp 56).

Using lemma 5.1 we can deduce that $\hat{f}2(z)\xrightarrow{p}f(z)$, hence $\hat{f}2(z)/f(z)\xrightarrow{p}1$ by lemma [3.1].

LEMMA 5.2

Let $\sigma^2(z)$, $f(z)$ and $w^*((z-z_j)/h)$ be smooth functions then

$$\lim_{\substack{n\to\infty \\ h\to 0}} E(\sigma^{*2}_{nw}(z))=\sigma^2(z),$$

and

$$\lim_{\substack{n\to\infty \\ h\to 0 \\ nh\to\infty}} Var(\sigma^{*2}_{nw}(z))=0.$$

PROOF.

From [5.21] we have

$$\sigma^{*2}_{nw}(z)=(nhf(z))^{-1}\sum_{j\in s}[w^*((z-z_j)/h)(y_j-\hat{\mu}_{nw}(z_j))^2$$

Taking conditional expectation given s and z we get

$$E(\sigma^{*2}_{nw}(z)|s,z)=(nhf(z))^{-1}\sum_{j\in s}[w^*((z-z_j)/h)E[(y_j-\hat{\mu}_{nw}(z_j))^2|z)$$

Adding and subtracting $\mu(z_j)$ inside the expectation and taking limits as $n\to\infty$, and $h\to 0$ we get

$$\lim_{\substack{n\to\infty \\ h\to 0}} E(\sigma^{*2}_{nw}(z)|s,z)=\lim_{\substack{n\to\infty \\ h\to 0}}(nhf(z))^{-1}[\sum_{j\in s}[w^*((z-z_j)/h)$$
$$[E[(y_j-\mu(z_j))^2]|z)+E[(\mu(z_j)-\hat{\mu}_{nw}(z_j))^2|z)]]$$

since cross product term is equal to zero,

$$=\lim_{\substack{n\Rightarrow\infty \\ h\Rightarrow 0}}(nhf(z))^{-1}[\sum_{j\in s}[w^{*}((z-z_j)/h)[E[(y_j-\mu(z_j))^2]|z)$$

$$+\lim_{\substack{n\Rightarrow\infty \\ h\Rightarrow 0}}(nhf(z))^{-1}[\sum_{j\in s}[w^{*}((z-z_j)/h)E[(\mu(z_j)-\hat{\mu}_{nw}(z_j))^2|z)]] \qquad [5.22]$$

Considering the second term of [5.22] we get

$$\lim_{\substack{n\Rightarrow\infty \\ h\Rightarrow 0}}(nhf(z))^{-1}[\sum_{j\in s}[w^{*}((z-z_j)/h)E[(\mu(z_j)-\hat{\mu}_{nw}(z_j))^2|z)]]$$

$$=\lim_{h\Rightarrow 0}(hf(z))^{-1}[\int[w^{*}((z-t)/h)f(t)Var[\hat{\mu}_{nw}(t)|z)]]dt$$

=0.using theorem [5.1] [5.23]

Substituting [5.23] in [5.22] we get

$$\lim_{\substack{n\Rightarrow\infty \\ h\Rightarrow 0}} E(\sigma^{*2}_{nw}(z)|s,z)$$

$$=\lim_{\substack{n\Rightarrow\infty \\ h\Rightarrow 0}}(nhf(z))^{-1}\sum_{j\in s}[w^{*}((z-z_j)/h)[E[(y_j-\mu(z_j))^2]|z]$$

$$=\lim_{\substack{n\Rightarrow\infty \\ h\Rightarrow 0}}(nhf(z))^{-1}\sum_{j\in s}w^{*}((z-z_j)/h)\sigma^2(z_j) \text{ using [5.12].}$$

$$= \lim_{h\Rightarrow 0}(hf(z))^{-1}\int w^{*}((z-t)/h)]\sigma^2(t)f(t)dt$$

$$= \lim_{h\Rightarrow 0}(f(z))^{-1}\int w^{*}(v)\sigma^2(z-hv)\, f(z-hv)dv$$

substituting $(z-t)/h=v$.

Using taylor series expansion and taking limits as $h\Rightarrow 0$ we get

$$\lim_{\substack{n\Rightarrow\infty \\ h\Rightarrow 0}}E(\sigma^{*2}_{nw}(z)|s,z)=\sigma^2(z)\int w^{*}(v)dv$$

$$=\sigma^2(z), \qquad [5.24]$$

since $\int w^{*}(v)dv=1$, from [5.10].

Thus $\lim_{\substack{n\Rightarrow\infty \\ h\Rightarrow 0}} E(\sigma^{*2}_{nw}(z)|s,z)=\sigma^{2}(z)$.

Averaging [5.24] over all the possible samples we get

$$\lim_{\substack{n\Rightarrow\infty \\ h\Rightarrow 0}} E_p E(\sigma^{*2}_{nw}(z)|s,z)=\sigma^{2}(z).$$

$$\Rightarrow \lim_{\substack{n\Rightarrow\infty \\ h\Rightarrow 0}} E(\sigma^{*2}_{nw}(z))=\sigma^{2}(z). \qquad [5.25]$$

Using eqn [5.12] the variance of a variance is of $O(n^{-1})$ and so

$$\lim_{\substack{n\Rightarrow\infty \\ h\Rightarrow 0 \\ nh\Rightarrow\infty}} V(\sigma^{*2}_{nw}(z)|s,z)=0. \qquad [5.26]$$

Now

$$\lim_{\substack{n\Rightarrow\infty \\ h\Rightarrow 0 \\ nh\Rightarrow\infty}} Var(\sigma^{*2}_{nw}(z))=E_p \lim_{\substack{n\Rightarrow\infty \\ h\Rightarrow 0 \\ nh\Rightarrow\infty}}[Var(\sigma^{*2}_{nw}(z)|s,z)]$$

$$+V_p[\lim_{\substack{n\Rightarrow\infty \\ h\Rightarrow 0 \\ nh\Rightarrow\infty}}[E(\sigma^{*2}_{nw}(z)|s,z)]$$

$$=0. \text{ using [5.25] and [5.26]} \qquad [5.27]$$

Using eqns [5.25] and [5.27] we deduce the result.

THEOREM 5.2

Let $\sigma^{2}(z)$ and $w^{*}((z-z_j)/h)$ be smooth functions then the smooth estimate $\hat{\sigma}^{2}_{nw}(z)$ is a consistent estimate of $\sigma^{2}(z)$ i.e

$$\lim_{\substack{n\Rightarrow\infty \\ h\Rightarrow 0}} E(\hat{\sigma}^{2}_{nw}(z))=\sigma^{2}(z),$$

and

$$\lim_{\substack{n\Rightarrow\infty \\ h\Rightarrow 0 \\ nh\Rightarrow\infty}} Var(\hat{\sigma}^2_{nw}(z))=0.$$

PROOF

From [5.22] we see that the estimate $\hat{\sigma}^2_{nw}(z)$ is a product of $\sigma^{*2}_{nw}(z)$and $[f(z)/\hat{f}2(z)]$.Following Watson [1964] we use lemma [5.1] and lemma [5.2] to deduce the result.

Similar results as those proved in theorem [5.1] and [5.2] holds for a general case $p>1$ and $q=1$.

THEOREM 5.3

Let $\mu(z),\sigma^2(z),w^*((z-z_j)/h)$ and $w((z-z_j)/k)$ be smooth fuctions then the kernel NW estimator of the population mean and covariance matrix derived in [5.17] and [5.18] are consistent estimators i.e;

$$\lim_{\substack{n\Rightarrow\infty \\ k\Rightarrow 0 \\ N\Rightarrow\infty, n/N\Rightarrow c(constant)}} E(\hat{\underset{\sim}{\mu}}_{y,nw})=\underset{\sim}{\mu}_y,$$

and

$$\lim_{\substack{n\Rightarrow\infty \\ h\Rightarrow 0 \\ nh\Rightarrow\infty \\ N\Rightarrow\infty, n/N\Rightarrow c(constant)}} Var(\hat{\underset{\sim}{\mu}}_{y,nw})=0,$$

and

$$\lim_{\substack{n\Rightarrow\infty \\ k,h\Rightarrow 0 \\ N\Rightarrow\infty, n/N\Rightarrow c(constant) \\ nk,nh\Rightarrow\infty}} E(\hat{\underset{\sim}{\Sigma}}_{yy,nw}=\underset{\sim}{\Sigma}_{yy},$$

and

$$\lim_{\substack{n\Rightarrow\infty \\ k,h\Rightarrow 0 \\ nk,nh\Rightarrow\infty \\ N\Rightarrow\infty, n/N\Rightarrow c(constant)}} Var(\hat{\underset{\sim}{\Sigma}}_{yy,nw})=0 .$$

PROOF

Applying lemma [3.1] to theorem [5.1] and [5.2] we get the result.

THEOREM 5.4

Let $\mu(z), \sigma^2(z), w^*((z-z_j)/h)$ and $w((z-z_j)/k)$ be smooth fuctions then the NW kernel estimator of the regression coefficient is a consistent estimator of β_{12} i.e,

$$\lim_{\substack{n\Rightarrow\infty \\ h,k\Rightarrow 0 \\ N\Rightarrow\infty, n/N\Rightarrow \text{ constant}) \\ nk,nh\Rightarrow\infty}} E(\hat{B}_{12,nw}) = \beta_{12},$$

and

$$\lim_{\substack{n\Rightarrow\infty \\ k,h\Rightarrow 0 \\ nk,nh\Rightarrow\infty \quad n/N\Rightarrow \text{ constant}) \\ N\Rightarrow\infty}} Var(\hat{B}_{12,nw}) = 0.$$

PROOF

Since the estimator of the regression coefficient β_{12} is a function of the estimators of the covariance $\hat{\Sigma}_{12,nw}$ and $\hat{\Sigma}_{22,nw}$ which we have proved are consistent estimators in theorem [5.3],then applying lemma [3.1] to theorem [5.3] we get the required result.

To derive the asymptotic properties of the PC estimators,we assume that the following conditions are satisfied.

(i) $0 \le z_1 \le \ldots\ldots \le z_n \le 1$

(ii) z_j's are randomly spaced.

THEOREM 5.5

Let $\mu(z)$ and $w((z-z_j)/k)$ be smooth functions then the smooth estimate $\hat{\mu}_{pc}(z)$ is a consistent estimate of $\mu(z)$ i.e

$$\lim_{\substack{n\Rightarrow\infty \\ k\Rightarrow 0}} E(\hat{\mu}_{pc}(z))=\mu(z),$$

and

$$\lim_{\substack{n\Rightarrow\infty \\ k\Rightarrow 0 \\ nk\Rightarrow\infty}} Var(\hat{\mu}_{pc}(z))=0.$$

PROOF.(see Priestley and Chao[1972]).

We now look at the asymptotic properties of the proposed estimator of the variance in [5.15].

THEOREM 5.6.

Let $\sigma^2(z)$ and $w^*((z-z_j)/h)$ be smooth functions then $\hat{\sigma}^2_{pc}(z)$ is a consistent estimate of $\sigma^2(z)$,i.e

$$\lim_{\substack{n\Rightarrow\infty \\ h\Rightarrow 0}} E(\hat{\sigma}^2_{pc}(z))=\sigma^2(z),$$

and

$$\lim_{\substack{n\Rightarrow\infty \\ h\Rightarrow 0 \\ nh\Rightarrow\infty}} Var(\hat{\sigma}^2_{pc}(z))=0.$$

PROOF.

From [5.15] we have

$$\hat{\sigma}^2_{pc}(z)=h^{-1}\sum_{j\in s}(z_j-z_{j-1})w^*((z-z_j)/h)(y_j-\hat{\mu}_{pc}(z_j))^2$$

Taking conditional expectation given s and z we get

$$E(\hat{\sigma}^2_{pc}(z)|s,z)=h^{-1}\sum_{j\in s}(z_j-z_{j-1})w^*((z-z_j)/h)E(y_j-\hat{\mu}_{pc}(z_j))^2|z)$$

Adding and subtracting $\mu(z_j)$ inside the expectation and then taking limits as $n\Rightarrow\infty$,and $h\Rightarrow 0$ we get

$$\lim_{\substack{n\Rightarrow\infty \\ h\Rightarrow 0}} E(\hat{\sigma}^2_{pc}(z)|s,z)$$

$$=\lim_{\substack{n\Rightarrow\infty \\ h\Rightarrow 0}} h^{-1}[\sum_{j\in s}[w^*((z-z_j)/h)[E[(y_j-\mu(z_j))^2]|z)$$

$$+E[(\mu(z_j)-\hat{\mu}_{pc}(z_j))^2|z)]](z_j-z_{j-1})$$

$$=\lim_{\substack{n\Rightarrow\infty\\h\Rightarrow 0}} h^{-1}[\sum_{j\in s}[w^{*}((z-z_j)/h)[E[(y_j-\mu(z_j))^2]|z](z_j-z_{j-1})$$
$$+\lim_{\substack{n\Rightarrow\infty\\h\Rightarrow 0}} h^{-1}[\sum_{j\in s}[w^{*}((z-z_j)/h)E[(\mu(z_j)-\hat{\mu}_{pc}(z_j))^2|z)]](z_j-z_{j-1})$$

[5.28]

Considering the second term of [5.28] we get

$$\lim_{\substack{n\Rightarrow\infty\\h\Rightarrow 0}} h^{-1}[\sum_{j\in s}[w^{*}((z-z_j)/h)E[(\mu(z_j)-\hat{\mu}_{pc}(z_j))^2|z)]](z_j-z_{j-1})$$

$$\cong\lim_{h\Rightarrow 0} h^{-1}[\int_0^1[w^{*}((z-t)/h)f(t)E[(\mu(t)-\hat{\mu}_{pc}(t))^2|z)]]dt$$

since with random spacing $z_j-z_{j-1}\cong n^{-1}$.

$$=\lim_{h\Rightarrow 0}\ [\int_{-(1-z)/h}^{z/h}\ [w^{*}(v)E[(\mu(z-hv)-\hat{\mu}_{pc}(z-hv))^2|z)]]f(z-hv)dv$$

substituting $(z-t)/h=v$.

Expanding using taylors series ,taking limits as $h\Rightarrow 0$ and using theorem 5.5 we get

$$\lim_{\substack{n\Rightarrow\infty\\h\Rightarrow 0}} h^{-1}[\sum_{j\in s}(z_j-z_{j-1})[w^{*}((z-z_j)/h)E[(\mu(z_j)-\hat{\mu}_{nw}(z_j))^2|z)]]=0.$$

[5.29]

Substituting [5.29] in [5.28] we get

$$\lim_{\substack{n\Rightarrow\infty\\h\Rightarrow 0}} E(\hat{\sigma}_{pc}^{2}(z)|s,z)$$

$$=\lim_{\substack{n\Rightarrow\infty\\h\Rightarrow 0}} h^{-1}\sum_{j\in s}(z_j-z_{j-1})[w^{*}((z-z_j)/h)[E[(y_j-\mu(z_j))^2]|z]$$

$$=\lim_{\substack{n\Rightarrow\infty\\h\Rightarrow 0}} h^{-1}\sum_{j\in s}(z_j-z_{j-1})[w^{*}((z-z_j)/h)\sigma^2(z_j)$$

using [5.12].

since with random spacing $z_j-z_{j-1}\cong n^{-1}$.

$$=\lim_{h\Rightarrow 0} h^{-1}\int_0^1 w^*((z-t)/h)]\sigma^2(t)f(t)dt$$

$$=\lim_{h\Rightarrow 0}\ [\int_{-(1-z)/h}^{z/h} w^*(v)\sigma^2(z-hv)\ f(z-hv)dv$$

substituting $(z-t)/h=v$.

Using taylor series expansion and taking limits as $h\Rightarrow 0$ we get

$$\lim_{\substack{n\Rightarrow\infty\\ h\Rightarrow 0}} E(\hat{\sigma}^2_{pc}(z)|s,z)=f(z)\ \sigma^2(z)\int w^*(v)dv$$

$$= \sigma^2(z). \qquad [5.30]$$

since $\int w^*(v)dv=1$. from [5.10] and

$$f(z)=\begin{cases} 1 & \text{if } z\in[0,1] \\ 0 & \text{otherwise}\end{cases}$$

Averaging [5.30] over all the possible samples we get

$$\lim_{\substack{n\Rightarrow\infty\\ h\Rightarrow 0}} E_p E(\hat{\sigma}^2_{pc}(z)|s,z)=\sigma^2(z).$$

$$\Rightarrow \lim_{\substack{n\Rightarrow\infty\\ h\Rightarrow 0}} E(\hat{\sigma}^2_{pc}(z))=\sigma^2(z). \qquad [5.31]$$

Using eqn [5.12] the variance of a variance is of $O(n^{-1})$ and so

$$\lim_{\substack{n\Rightarrow\infty\\ h\Rightarrow 0\\ nh\Rightarrow\infty}} V(\hat{\sigma}^2_{pc}(z)|s,z)=0. \qquad [5.32]$$

Now

$$\lim_{\substack{n\Rightarrow\infty\\ h\Rightarrow 0\\ nh\Rightarrow\infty}} Var(\hat{\sigma}^2_{pc}(z))=E_p \lim_{\substack{n\Rightarrow\infty\\ h\Rightarrow 0\\ nh\Rightarrow\infty}}[Var(\hat{\sigma}^2_{pc}(z)|s,z)]$$

$$+V_p[\lim_{\substack{n\Rightarrow\infty\\ h\Rightarrow 0}}[E(\hat{\sigma}^2_{pc}(z)|s,z)]$$

$nh \Rightarrow \infty$

=0. using [5.30] and [5.32] [5.33]

Using [5.31] and [5.33] we deduce the result.
On similar lines as theorems [5.3] and [5.4] we use theorems [5.5],[5.6] and lemma [3.1] to prove that the Priestley-Chao estimator of the mean,covariance matrix and the regression coefficient are consistent estimators.

5.5 CONCLUSION

In this chapter we have proposed estimators which do not depend on any parametric model assumptions hence should be robust to the violations of any parametric model assumptions.These nonparametric estimators answers the main question of this thesis 'the search for a robust estimator'.Since previous work on nonparametric estimators have concentrated on the curve fitting for a given set of data,we have extended this concept to that of estimation of complex statistics in sample surveys.We have shown that the proposed estimators are consistent for a given set of consistency conditions.For this set of consistency conditions we proved that the kernel estimators are both approximately conditionally and unconditionally unbiased.Since these nonparametric estimators do not depend on any model assumptions,they should be robust,hence we can conclude that these model based nonparametric procedures are robust in both the conditional and the unconditional distribution frameworks.In chapter 4,we deduced that the weighted estimators were robust in the unconditional distribution framework but were found to be very inefficient in some circumstances and severely conditionally biased.The model based parametric procedures were found not to be robust in both the conditional and the unconditional distribution frameworks when the underlying parametric model assumptions are violated.The questions lingering our minds is whether these model based nonparametric robust estimators have any significant gain in efficiency over the model based parametric estimators when the underlying parametric model assumptions are violated and over the robust design based estimators.To answer these ques t ions we carried out an empirical study reported in chapter 6.

CHAPTER 6

COMPARATIVE STUDY OF THE NONPARAMETRIC REGRESSION ESTIMATORS

6.1 INTRODUCTION

To compare the performance of the nonparametric estimators of the regression coefficient studied theoretically in chapter 5 and the parametric estimators studied in the preceding chapters we carried out an empirical study. In the case of the Priestley-Chao and Nadaraya-Watson type estimators we must first choose a kernel function.

6.2.1 CHOICE OF KERNEL FUNCTIONS

Most authors assert that the performance of the kernel estimators does not depend critically on the choice of the kernel function but rather on the bandwidth parameter. Since the Gaussian kernel is the most widely used function we chose to use it to study the nonparametric estimators of the mean and covariance matrix.

For the Kernel estimate we considered the following smoother function

$W(z_i,z_j)=c_iW_k(z_i,z_j)$, where $i\in U$ and $j\in s$.

W_k is the kernel function and k is the bandwidth parameter.

The Gaussian kernel smoother function is given by

$$W_k(z_i,z_j)=c_i\exp(-(z_i-z_j)^2/2k^2) \text{ ,where } i\in U \text{ and } j\in s.$$

For the NW estimator,

$$c_i=1/\sum_{j\in s}\exp(-(z_i-z_j)^2/2k^2)$$

and for the PC estimator $c_i=(z_j-z_{j-1})/k$.
Since with c_i defined this way we argued theoretically in section 5.3.1 that the PC estimator will be biased at the endpoints,hence the global bias of the estimator might be very severe.We thus decided to modify the PC estimator so that the sum of weights is equal to one i.e

$$\sum_j W_k(z_i,z_j)=1.$$

We achieved this by defining

$$c_i=(z_j-z_{j-1})/\sum_j(z_j-z_{j-1})\exp(-(z_i-z_j)^2/2k^2).$$

We call the resulting modified PC estimator with Nadaraya-Watson type of weights, the MPC estimator. This modified PC estimator is a weighted average of the observations.
In this study we will use the same kernel function i.e Gaussian kernel function for the estimators of the mean and variance though there is no reason why different kernel functions cannot be used.

6.2.2 CHOICE OF THE BANDWIDTH PARAMETER

Since the choice of the bandwidth parameter determines the performance of the nonparametric estimators,a lot of work is currently going on to determine its optimal value in terms of minimizing the mean integrated square error see Rice(1984),Marron and Park(1989) etc. Since in our research we are sampling from a finite population,we will determine the optimal bandwidth parameter(k)in terms of minimizing the mean square error over repeated samples.We will calculate the mean square error of the kernel estimators for different values of the bandwidth parameter and whichever value of k gives the least mean square error to the kernel estimator over repeated samples,is taken as the optimal value.To get a rough idea of the neighbourhood within which the optimal value of the bandwidth parameter might lie we use the following formulae given by Silverman[1986];

$$0.25\sigma n^{-1/5} \leq k \leq 1.5\sigma n^{-1/5} . \qquad [6.1]$$

where σ is the standard deviation of the design variables and n is the sample size taken.

In our empirical studies we will evaluate the value of k at the boundaries of the interval and using these values calculate the mean square error of our estimators.We take as our starting point the value of k,k_{opt} say in the two limits which gives the least value of the mean square error.By considering other values of k in the neighbourhood of k_{opt} we try to get an estimator with the least mean square error.

In the next section we outline the simulation study we carried out to study the performance of the parametric and the nonparametric estimators.Since in chapter 4 we found that the adjusted estimators and the equally weighted Fuller estimator to be preferable for different sample designs to the other estimators,we will only compare these adjusted estimators,i.e Maximum likelihood,probability weighted adjusted estimators and the equally weighted Fuller estimator with the proposed nonparametric estimators.

6.3 SIMULATION STUDY

In this simulation study we studied two types of simulations.The first will be repeated sampling from a multivariate normal population to compare the performance of the nonparametric estimators with the Maximum likelihood,probability weighted adjusted estimators and the equally weighted Fuller estimators whose performance is optimal when the data is multivariate normal.In the second we sample repeatedly from a 'Real'population to check the performance of the estimators when the multivariate normal model assumptions are violated.

In this chapter we carried out both the conditional and unconditional analysis.The former will allow us to assess whether a particular estimator is good in some samples and poor for others whereas the latter averages over all possible samples for a particular design.

6.3.1 REPEATED SAMPLING FROM A MULTIVARIATE NORMAL POPULATION

We considered cases where (q= 1 i.e one design variable) and generated 6,962 finite population values $\underset{\sim}{x}=(\underset{\sim}{y}_i,\underset{\sim}{z}_i)$ i=1...6962 from a multivariate normal distribution where $\underset{\sim}{y}_i=(y_{1i},y_{2i})$ and the mean vector $\underset{\sim}{\mu}_x$ and the covariance matrix $\underset{\sim}{\Sigma}_{xx}$ were chosen to be those estimated for a set of variables from the 1975 U.K Family expenditure survey, the correlation matrix and the standard deviation of the two y values and the z-variable obtained from the actual FES are given below.

Table of parameter values from the real population

VARIABLE		S.D	CORRELATION MATRIX		
y_1	Expenditure on all items	0.668	1		
y_2	Total income	0.849	0.75	1	
z	Expenditure on food	0.658	0.41	0.28	1

The design variable is the expenditure on food,independent variable is the total income and the dependent variable is the total expenditure. This finite population was stratified into five strata according to increasing values of the design variable,such that the first strata contains 1393 units with lowest values of z,second,third,fourth contain 1392 units each and the fifth contains the last 1393 units with the highest z values.

The sample designs used were based on those used by Holt Smith and Winter [1980]. Denote a stratified random sampling design by $(n_1 \ldots n_h)$ with n_h units selected from the h^{th} stratum h =1,..,5.Unlike in chapter 4,where we considered eight sample designs,we will consider only three designs representing all the eight designs,since essentially the eight designs considered represented only three sampling schemes,the equal probability,increasing allocation and U-shaped designs at different degrees of extremity.

sample designs.			n_1	n_2	n_3	n_4	n_5	symbol
D1	proportional	allocation	20	20	20	20	20	Δ
D2	increasing	allocation	5	9	16	30	40	∇
D3	U-shaped	allocation	40	8	4	8	40	+

Before we can study the performance of the nonparametric regression estimators for all the three sample designs,our first task is to select a bandwidth parameter k which gives the nonparametric estimator the least mean square error for each sample design. From [6.1] we see that the bandwidth parameter does not depend on the type of the kernel function used but on the sample size taken and the standard deviation of the design variable.We will therefore choose an appropriate Bandwidth for only one of the nonparametric estimators and apply the same for all the other nonparametric estimators.Arbitrarily we choose the Nadaraya-Watson estimator for choosing the bandwidth for different designs.

We selected a number of values of k using [6.1] as our quideline.The selected values for each sample design are given in the table below and in figures 6.1-6.3 we present plots of the mean square errors,standard deviations and biases of the Nw estimator against different values of k chosen for the three sample designs.

Table of values selected for choosing the optimal bandwidth Parameter, standard deviations, biases and mean square errors of the NW regression estimator for each design. over 100 replications.

Design D1

Bandwidth Parameter	Standard deviation	Bias	Mean square Error
0.4	0.0534	0.0188	0.0032 *
0.5	0.0542	0.0226	0.0035
0.6	0.0542	0.0247	0.0036
0.7	0.0542	0.0256	0.0036

* means optimal bandwidth

Design D2

Bandwidth Parameter	Standard deviation	Bias	Mean square Error
0.4	0.0556	0.0237	0.0037
0.6	0.0541	0.0259	0.0036
0.7	0.0531	0.0269	0.0035*
0.8	0.0524	0.0271	0.0035

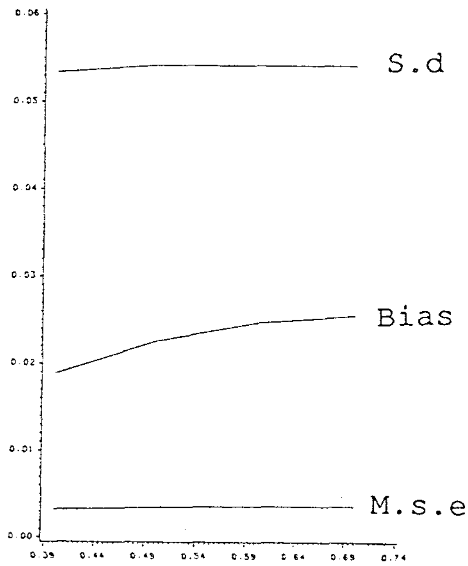
* means optimal bandwidth

Design D3

Bandwidth Parameter	Standard deviation	Bias	Mean square Error
0.9	0.0692	0.0007	0.0048
1.0	0.0679	0.0016	0.0046
2.0	0.0603	0.0285	0.0045*
3.0	0.0583	0.0422	0.0052

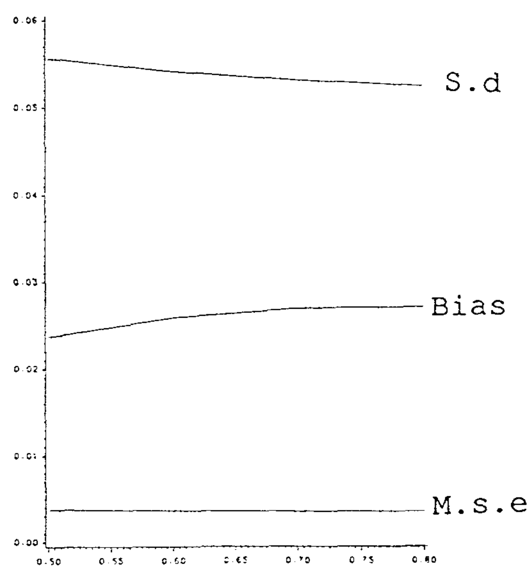
* means optimal bandwidth

FIG 6.1 PLOT OF THE MEAN SQUARE ERRORS, STANDARD DEVIATIONS AND ABSOLUTE BIASES FOR DESIGN D1 VS BANDWIDTH PARAMETER VALUES.



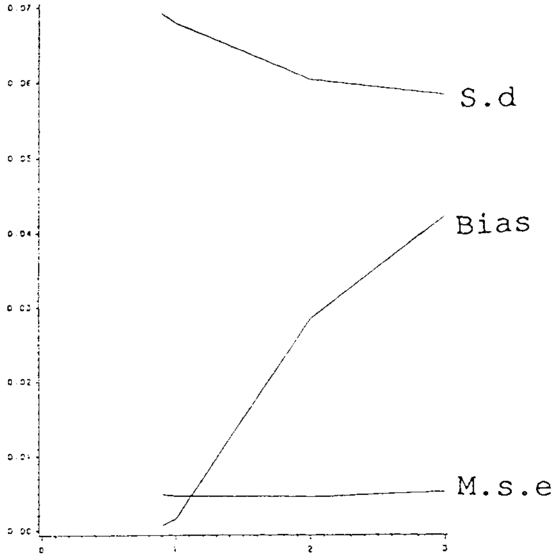
Bandwidth parameter values (K)

FIG 6.2 PLOT OF THE MEAN SQUARE ERRORS, STANDARD DEVIATIONS AND ABSOLUTE BIASES FOR DESIGN D2 VS BANDWIDTH PARAMETER VALUES.



Bandwidth parameter values (K)

FIG 6.3 **PLOT OF THE MEAN SQUARE ERRORS, STANDARD DEVIATIONS AND ABSOLUTE BIASES FOR DESIGN D3 VS BANDWIDTH PARAMETER VALUES.**



Bandwidth parameter values (K)

We see from the plots that when the bandwidth parameter is equal to 0.4,0.7 and 1.0,the Nw estimator has the least mean square error for the design D1,D2 and D3 respectively.We consider this value of k as the optimal bandwidth and we will regard the Nw estimator with this bandwidth as optimal.As expected we observe that as the bandwidth parameter increases the bias of the NW kernel estimator increases.We observed that as k increases the rate of decrease of the variance is very small.

Once the value of k is chosen we then compared the performance of the nonparametric estimators with the parametric estimators proposed by other authors to solve the problem stated in section 5.1.Empirical studies by Holmes[1987], Skinner,Holt and Smith[1989] and also in chapter 4 concluded that the adjusted estimators perform better than the unadjusted ones when the data is multivariate normal.Since the maximum likelihood and equally weighted Fuller estimator have optimal properties when the population is multivariate normal ,it will therefore be interesting to compare the properties of the nonparametric estimators with the maximum likelihood (Ml),probability weighted adjusted (Pwml) and equally weighted Fuller (ewf) estimators when the data is multivariate normal. We will study the performance of the nonparametric estimators, maximum likelihood, probability weighted adjusted and equally weighted Fuller estimators both unconditionally and conditionally.

For the various stratified sample designs we selected 1,000 independent samples of size 100 from the finite population. The sampling distribution of the various statistics under investigation were estimated from these 1,000 repeated samples. We obtain the unconditional results by averaging the statistics under investigation over all the 1000 samples.

(ii) UNCONDITIONAL ANALYSIS

We first present the unconditional absolute biases of the Priestley-Chao amd the Modified Priestley-Chao estimators.

TABLE 6.1 UNCONDITIONAL ABSOLUTE BIASES OF THE PC AND MPC ESTIMATORS OF REGRESSION COEFFICIENT OVER 1000 REPLICATIONS FOR EACH DESIGN.

N=6962,n=100 TRUE VALUE β_{12}=0.595

SAMPLE DESIGN	ABSOLUTE BIASES OF $\hat{\beta}_{12,mpc}$	$\hat{\beta}_{12,pc}$
D1	0.0046	0.3817
D2	0.0109	0.3825
D3	0.0071	0.3827

We see from table 6.1 that as expected from theoretical results the Priestley-Chao(PC) estimator is severely biased as compared to its modified version.We therefore dropped the PC estimator in favour of its modified version for further analysis.

We present the unconditional results of the five estimators of the regression coefficient below,

TABLE 6.2 UNCONDITIONAL ABSOLUTE BIASES OF THE FIVE ESTIMATORS OF REGRESSION COEFFICIENT OVER 1000 REPLICATIONS FOR EACH DESIGN.

N=6962,n=100 TRUE VALUE β_{12}=0.595

SAMPLE DESIGN	ABSOLUTE BIASES OF $\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,pwml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,nw}$	$\hat{\beta}_{12,mpc}$
D1	0.0003	0.0003	0.0001	0.0185	0.0046
D2	0.0007	0.0019	0.0014	0.0269	0.0109
D3	0.0026	0.0018	0.0008	0.0159	0.0071

TABLE 6.3 UNCONDITIONAL STANDARD DEVIATIONS OF THE FIVE ESTIMATORS OF REGRESSION COEFFICIENT OVER 1000 REPLICATIONS FOR EACH DESIGN.

SAMPLE	STANDARD DEVIATIONS				
DESIGN	$\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,pwml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,nw}$	$\hat{\beta}_{12,mpc}$
D1	0.0500	0.0500	0.0501	0.0507	0.0689
D2	0.0522	0.0693	0.0510	0.0531	0.0907
D3	0.0486	0.0710	0.0494	0.0503	0.0819

TABLE 6.4 UNCONDITIONAL MEAN SQUARE ERRORS OF THE FIVE ESTIMATORS OF REGRESSION COEFFICIENT OVER 1000 REPLICATIONS FOR EACH DESIGN.

SAMPLE	MEAN SQUARE ERRORS				
DESIGN	$\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,pwml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,nw}$	$\hat{\beta}_{12,mpc}$
D1	0.0025	0.0025	0.0025	0.0029	0.0048
D2	0.0027	0.0048	0.0026	0.0035	0.0084
D3	0.0024	0.0050	0.0024	0.0028	0.0068

Tables 6.2-6.4 gives the unconditional results of the absolute bias,standard deviations and the mean square errors of the five estimators.We see that as expected

from theoretical resul t s derived in chapter 3 and by Holmes(1987) the maximum likelihood ,probability weighted adjusted and the equally weighted Fuller estimators are approximately unconditionally unbiased when the data is multivariate normal.Empirical results in table 6.2 seems to support Gasser and Engel(1990) qualitative argument that the sample expectation of the Nadaraya-Watson (NW) estimator is severely biased when the data is linear and that of the Priestley-Chao (MPC) estimator has a very small bias compared to the NW estimator.This pattern is similar across all the three sample designs considered.

The Ml and ewf estimators have smaller variances than all the other estimators.We note that the standard deviations of the MPC estimator is higher than those of the NW estimator by a factor 1.4 for equal probability design and 1.6 for the unequal probability designs.This result fits well with the theoretical results given by Gasser and Engel(1990) who proved that for random designs the variance of the MPC estimator is higher than that of the NW estimator by a factor 1.5 for a simple random design.The mean square error results given in table 6.4 indicate that the ML estimator is the most efficient across all the sample designs.The ewf estimator is more efficient than the Pwml estimator for the U-shaped design.The low standard deviations of the NW estimator compensates for the severe bias bringing down the mean square error so that it is more efficient than the Pwml and MPC estimators for the unequal probability designs.The high variances of the MPC estimators makes the estimator very inefficient as compared to all the other estimators.

(ii) CONDITIONAL ANALYSIS

To assess the conditional asymptotic properties of the estimators the 1,000 samples were divided into 20 groups of 50 samples each according to increasing values of $\Delta^{F}_{zz}=(S_{zzs}- S_{zz})/S_{zz}$ for NW,Mpc,ml and ewf estimators and $\Delta^{*F}_{zz}=(S^{*}_{zzs}- S_{zz})/S_{zz}$ for the pwml estimators respectively such that the first group contained the 50 samples with the smallest values of Δ^{F}_{zz} (or Δ^{*F}_{zz}) and so on upto the 20th group which contains the 50 samples with the largest values of Δ^{F}_{zz} (or Δ^{*F}_{zz}). We assume that the variation in Δ^{F}_{zz} (or Δ^{*F}_{zz}) within each group is small.The conditional distribution of the various estimators given Δ^{F}_{zz} (or Δ^{*F}_{zz}) could then be estimated.

Figure [6.4]-[6.8] gives the plots of the group means of the pwml,ml,ewf and the nonparametric estimators conditional on s and z,plotted against the group means of Δ^{*F}_{zz} and Δ^{F}_{zz} for design based and model based, equally weighted Fuller and the nonparametric estimators respectively.

FIG 6.4 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,ml}$ VS GROUP MEANS OF Δ^{F}_{zz} (20 groups,50 samples per group).

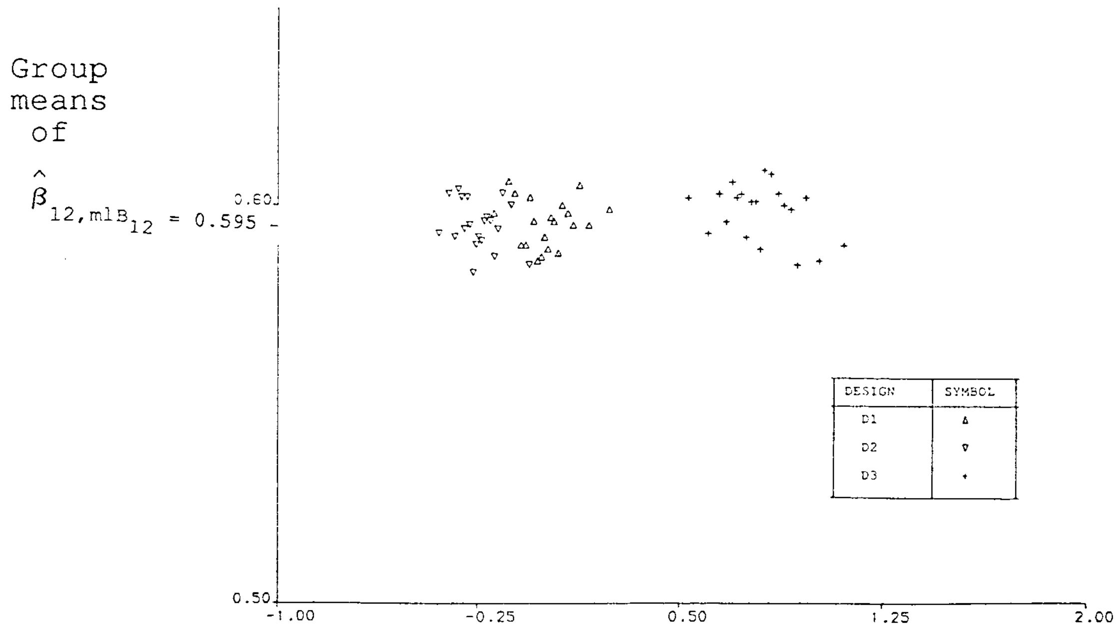


FIG 6.5 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,pwml}$ VS GROUP MEANS OF Δ^{*F}_{ZZ} (20 groups, 50 samples per group).

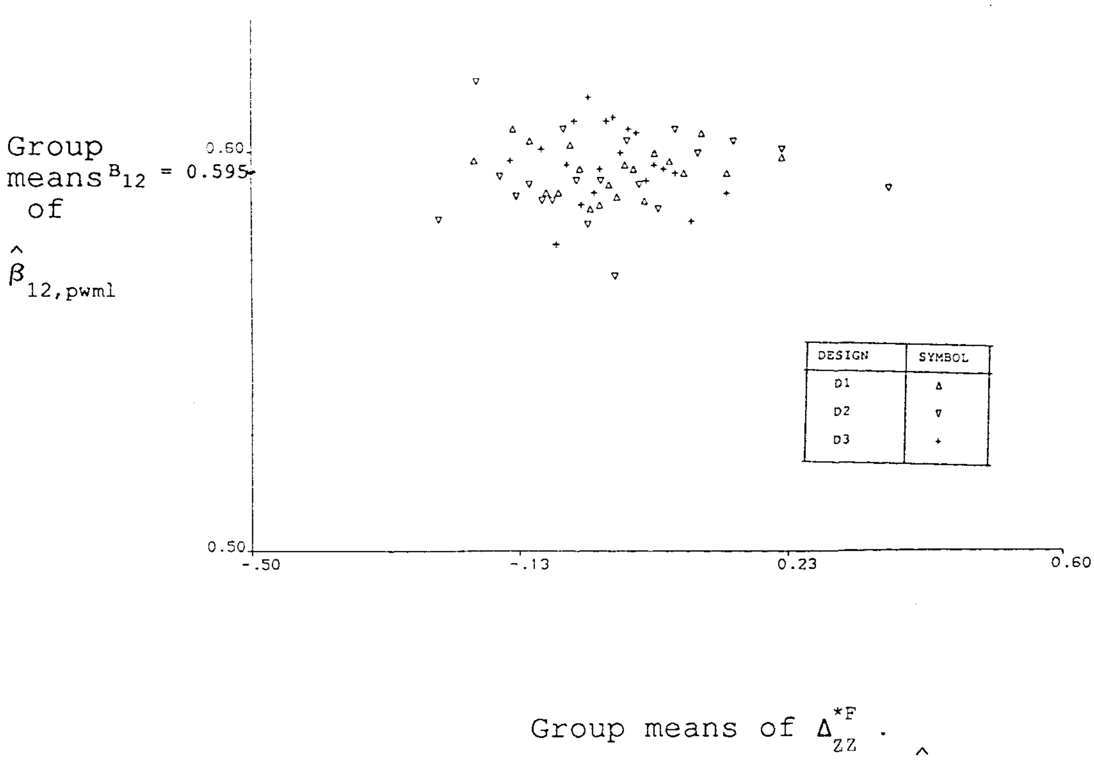


FIG 6.6 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,ewf}$ VS GROUP MEANS OF Δ^{F}_{ZZ} (20 groups, 50 samples per group).

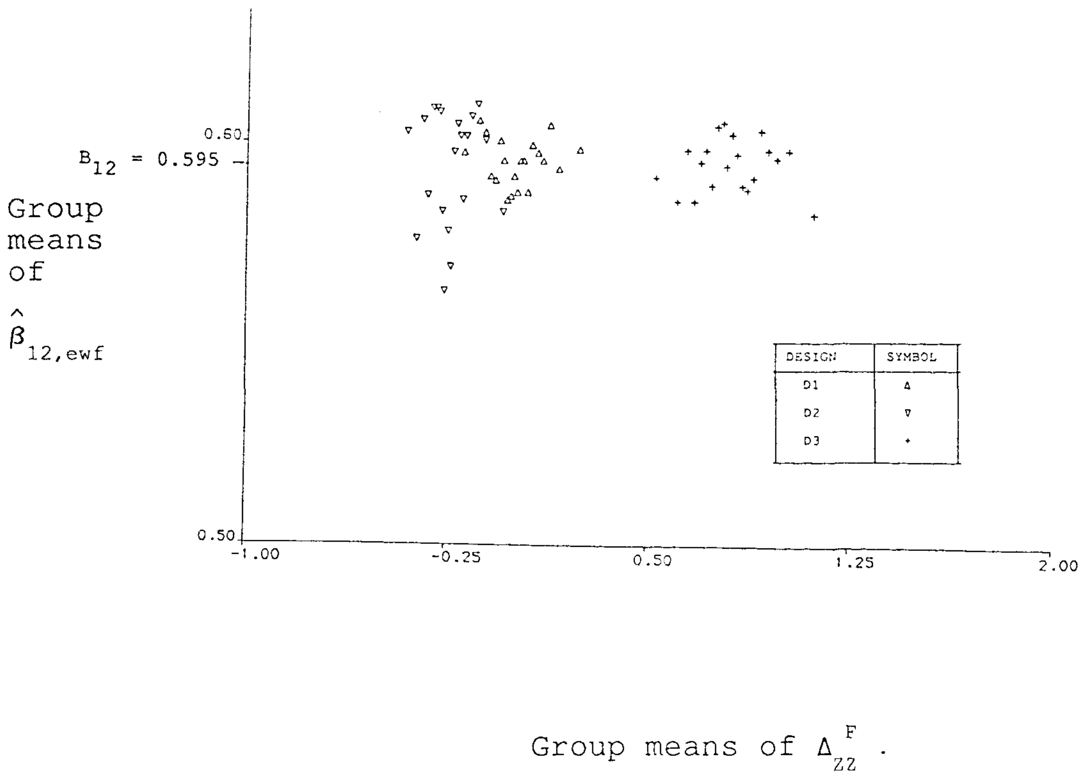
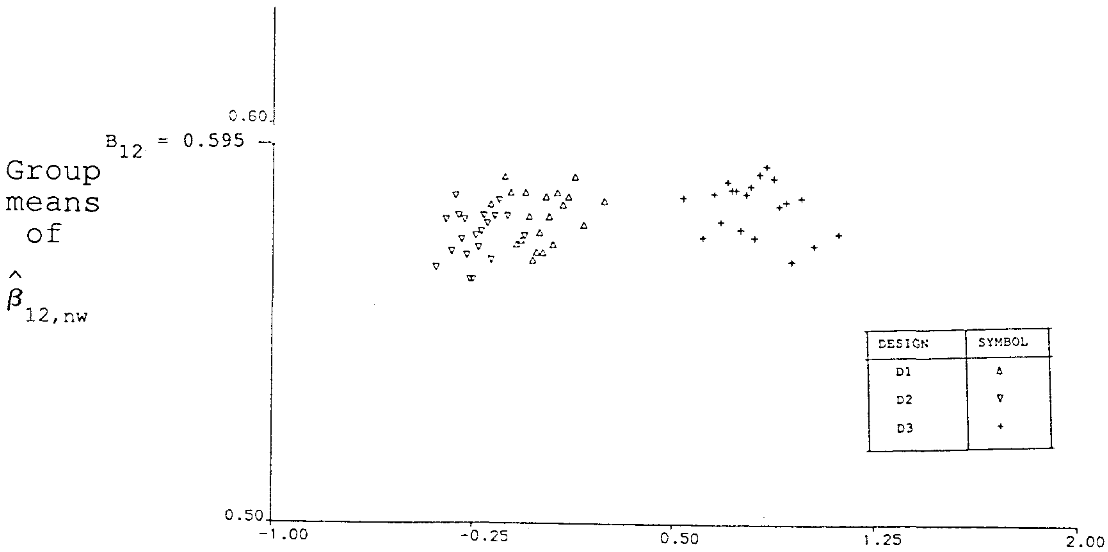
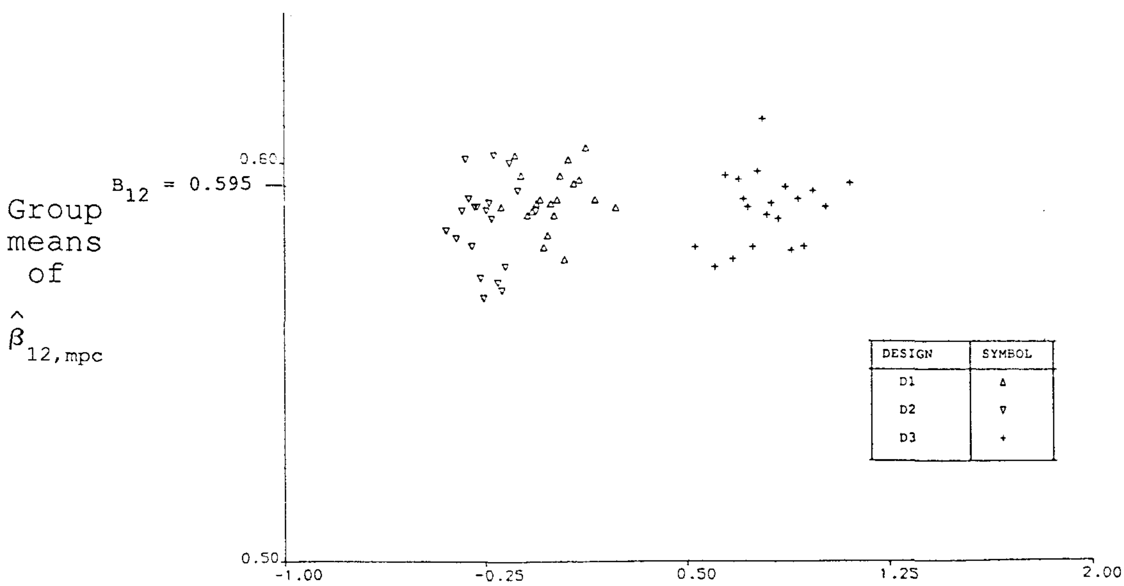


FIG 6.7 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,nw}$ VS GROUP MEANS OF Δ_{zz}^{F} (20 groups, 50 samples per group).



Group means of Δ_{zz}^{F} .

FIG 6.8 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,mpc}$ VS GROUP MEANS OF Δ_{zz}^{F} (20 groups, 50 samples per group).



Group means of Δ_{zz}^{F} .

We see from figure [6.4],[6.5] and [6.6] that as expected for the Ml,Pwml and ewf estimators, they are approximately conditionally unbiased.We also note from fig [6.7] and [6.8] that the NW and MPC estimators are also approximately conditionally unbiased for all the three designs.

We now investigate the performance of the five regression estimators when the population is 'Real'.

6.3.2 REPEATED SAMPLING FROM A MULTIVARIATE " REAL" POPULATION

In this simulation study we consider 6962 actual data points which make up the Family expenditure survey defining the finite population.We consider the same variables as in section 6.3.1 and sample repeatedly from this population to investigate the robustness properties of the five regression estimators.We expect the real population to violate at least one of the normality assumptions.

To investigate the structure of the real population we present some plots in figures [6.9]-[6.11]. We see from figures [6.9]-[6.11] that there is no clear evidence to suggest that the expenditure on all items and the total income are linear in the design variable,the expenditure on food.However figure [6.10] does seem to suggest a linear relationship between the expenditure on all items and the total income.We can therefore deduce that the parameter β_{12} is a meaningful parameter to estimate.

As in section 6.3.1 the first task facing us is to choose an appropriate bandwidth parameter for the NW estimator. In figures [6.12]-[6.14] we present plots of the mean square error,standard deviations and absolute biases of the Nw estimator against different values of k chosen for the three sample designs.

Table of values selected for choosing the optimal bandwidth Parameter, standard deviations, biases and mean square errors of the NW regression estimator for each design over 100 replications.

Design D1

Bandwidth Parameter	Standard deviation	Bias	Mean square Error
0.6	0.0910	0.0049	0.0083 *
0.7	0.0910	0.0050	0.0083
0.8	0.0910	0.0056	0.0083
0.9	0.0910	0.0064	0.0084

* means optimal bandwidth

Design D2

Bandwidth Parameter	Standard deviation	Bias	Mean square Error
0.9	0.0604	0.0236	0.0042
1.0	0.0602	0.0238	0.0042*
1.5	0.0603	0.0270	0.0044
2.0	0.0609	0.0304	0.0046

* means optimal bandwidth

Design D3

Bandwidth Parameter	Standard deviation	Bias	Mean square Error
0.5	0.0894	0.0001	0.0080
0.6	0.0880	0.0052	0.0078*
0.7	0.0878	0.0089	0.0078
0.8	0.0878	0.0109	0.0079

* means optimal bandwidth

FIG 6.9 PLOT OF THE EXPENDITURE ON ALL ITEMS (y_1) VS EXPENDITURE ON FOOD (z) FOR THE WHOLE POPULATION.

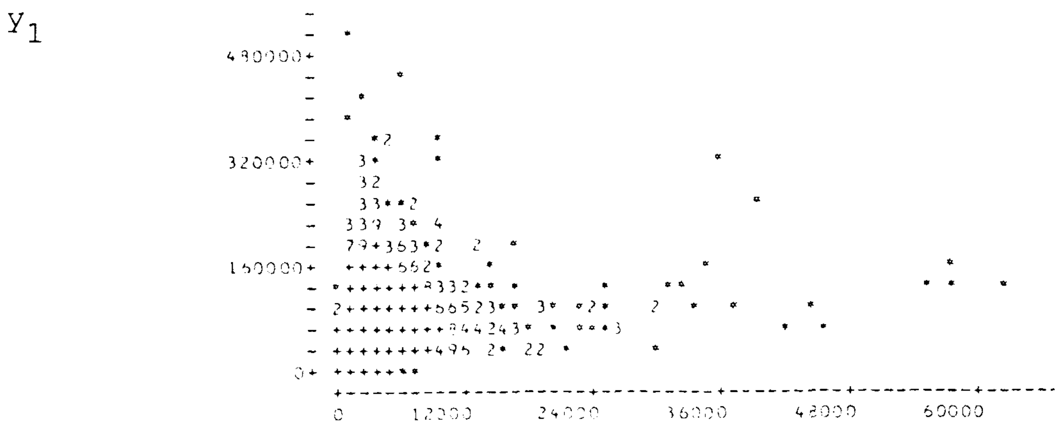


FIG 6.10 PLOT OF THE EXPENDITURE ON ALL ITEMS (y_1) VS TOTAL INCOME (y_2) FOR THE WHOLE POPULATION.

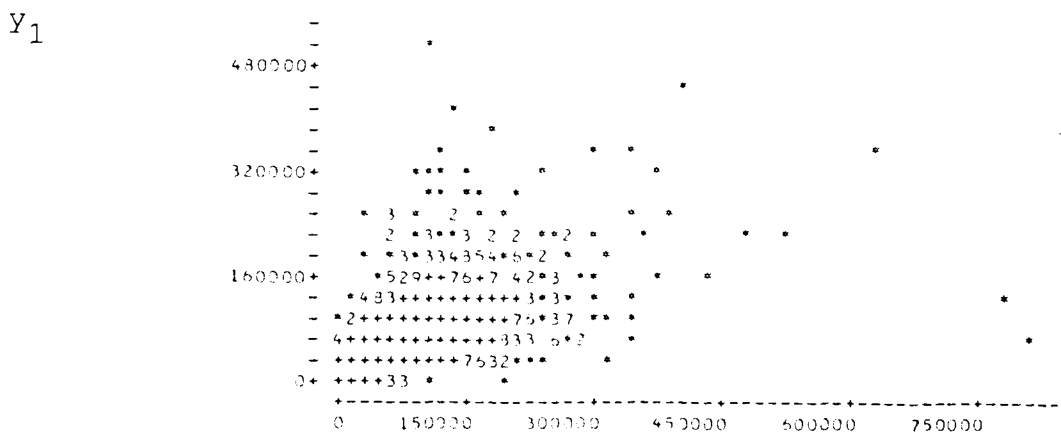
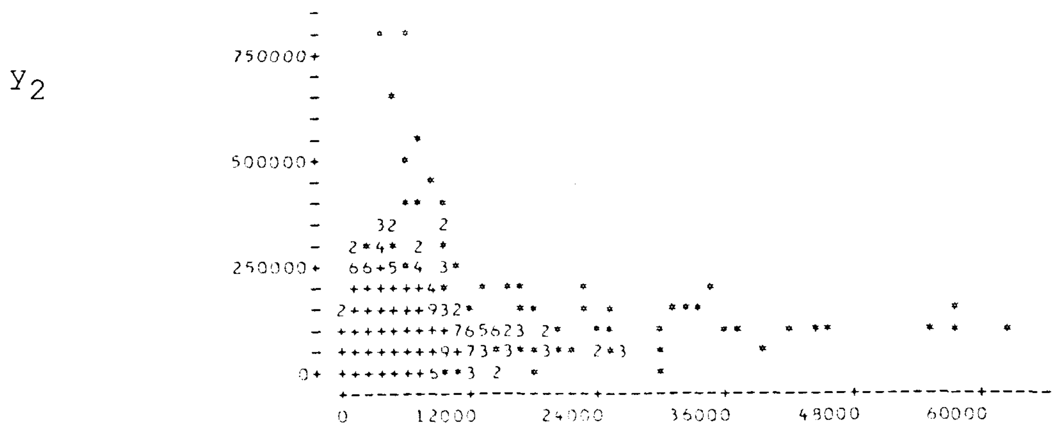
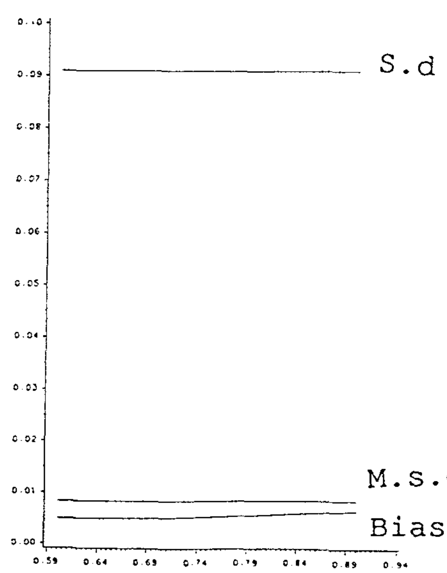


FIG 6.11 PLOT OF THE TOTAL INCOME (y_2) VS EXPENDITURE ON FOOD (z) FOR THE WHOLE POPULATION.



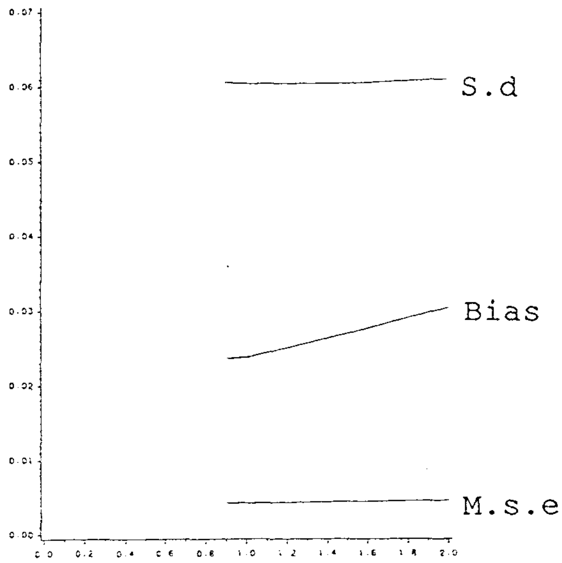
Expenditure on food (z)

FIG 6.12 PLOT OF THE MEAN SQUARE ERRORS, STANDARD DEVIATIONS AND ABSOLUTE BIASES FOR DESIGN D1 VS BANDWIDTH PARAMETER VALUES.



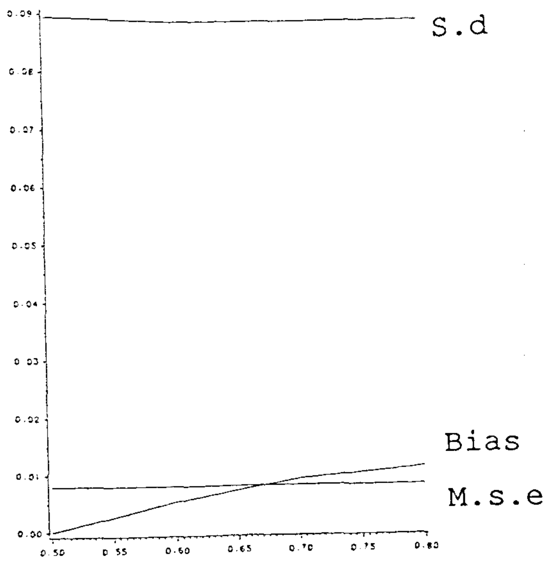
Bandwidth parameter values (K)

FIG 6.13 PLOT OF THE MEAN SQUARE ERRORS, STANDARD DEVIATIONS AND ABSOLUTE BIASES FOR DESIGN D2 VS BANDWIDTH PARAMETER VALUES.



Bandwidth parameter values (K)

FIG 6.14 PLOT OF THE MEAN SQUARE ERRORS, STANDARD DEVIATIONS AND ABSOLUTE BIASES FOR DESIGN D3 VS BANDWIDTH PARAMETER VALUES.



Bandwidth parameter values (K)

We see from the plots that when the bandwidth parameter is equal to 0.6 the Nw estimator has the least mean square error for the probability designs D1 and D3 and for the increasing allocation design D2 this optimal bandwidth is equal to 1.0.Increasing or decreasing this optimal value of k makes the estimator more inefficient.

Holmes(1987),Skinner,Holt and Smith(1989) found that the Maximum likelihood estimator was not robust to the violation of the normality model assumptions.The probability weighted adjusted estimator proposed by Nathan and Holt (1980) which was expected to have the robustness properties of the probability weighted regression estimator and the high efficiency properties of the Ml estimator was found to be robust unconditionally but it did not have any significant gain in efficiency over the probability weighted estimator.In chapter 4 we found that the equally weighted Fuller estimator compromises the poor efficiency properties of the Ml estimator for increasing allocation designs and the poor efficiency properties of the Pwml estimator for the U-shaped designs when the normality assumptions are violated but was found not to be robust to the violation of the linearity assumption.In section 6.3.1 we found that the NW estimator is more efficient than the Pwml when the data is multivariate normal.Does it still retain this desired property when the multivariate normal model assumptions are violated?.Since nonparametric estimators do not make any model assumptions they are robust to the violations of the multivariate normal model assumptions,but do they have any significant gain in efficiency over the Ml estimator when the multivariate normal model assumptions are violated?.We present the unconditional results averaged over 1000 samples for the three sample designs D2 with optimal k=1.0 and D1,D3 with optimal k=0.6.

(i) UNCONDITIONAL ANALYSIS

TABLE 6.5 UNCONDITIONAL ABSOLUTE BIASES OF THE FIVE ESTIMATORS OF REGRESSION COEFFICIENT OVER 1000 REPLICATIONS FOR EACH DESIGN.

N=6962, n=100 TRUE VALUE β_{12}=0.576

SAMPLE	ABSOLUTE BIASES OF				
DESIGN	$\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,pwml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,nw}$	$\hat{\beta}_{12,mpc}$
D1	0.0245	0.0245	0.0249	0.0056	0.0337
D2	0.0260	0.0408	0.0401	0.0060	0.0260
D3	0.0128	0.0355	0.0326	0.0072	0.0388

TABLE 6.6 UNCONDITIONAL STANDARD DEVIATIONS OF THE FIVE ESTIMATORS OF REGRESSION COEFFICIENT OVER 1000 REPLICATIONS FOR EACH DESIGN.

SAMPLE	STANDARD DEVIATIONS				
DESIGN	$\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,pwml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,nw}$	$\hat{\beta}_{12,mpc}$
D1	0.111	0.111	0.111	0.111	0.134
D2	0.106	0.132	0.105	0.108	0.147
D3	0.111	0.122	0.112	0.111	0.138

TABLE 6.7 UNCONDITIONAL MEAN SQUARE ERRORS OF THE FIVE ESTIMATORS OF REGRESSION COEFFICIENT OVER 1000 REPLICATIONS FOR EACH DESIGN.

SAMPLE	MEAN SQUARE ERRORS				
DESIGN	$\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,pwml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,nw}$	$\hat{\beta}_{12,mpc}$
D1	0.0130	0.0130	0.0132	0.0121	0.0190
D2	0.0120	0.0192	0.0126	0.0117	0.0229
D3	0.0125	0.0161	0.0136	0.0123	0.0205

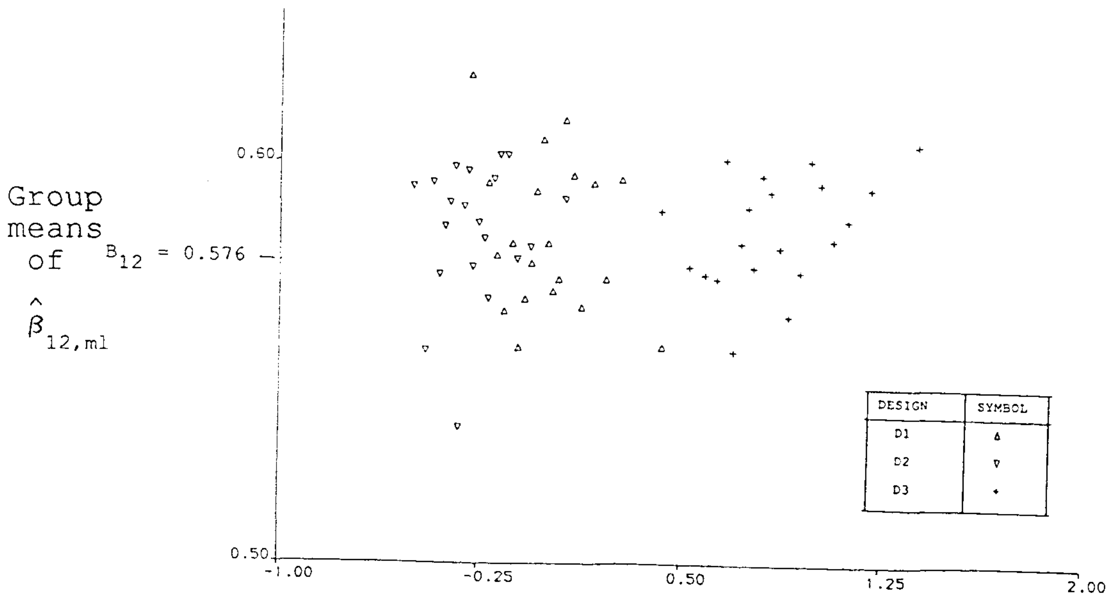
Tables [6.5]-[6.7] gives the unconditional absolute biases,standard deviations and mean square errors of the five estimators when the population is real.We see that the maximum likelihood estimator is biased for the increasing allocation design but is approximately unbiased for the equal probability and the U-shaped designs.The probability weighted adjusted estimator is also severely biased for the extreme designs D2 and D3.The equally weighted Fuller estimator though biased for the unequal probability designs D2 and D3 its bias for these designs is smaller than that of the Pwml estimator. For the increasing allocation design the bias of the MPC estimator is comparable to that of the Ml estimator.The NW estimator has the least bias for all the designs among all the five estimators.Gasser and Engel [1990] illustrated that the expectation of the NW estimator is a sigmoid function of z and the expectation of the MPC estimator is approximately linear,so when the data is not multivariate normal the NW estimator should have a very small bias and the MPC estimator should be severely biased.Our empirical studies confirm this assertion.From table[6.6] we see that the Ml,ewf and NW estimators have the least standard deviations.The Pwml estimator has higher standard deviations but comparatively less than those of the MPC estimator.As expected for random design the standard deviations for the MPC estimator are higher than those of the NW estimator.(see Gasser and Engel [1990]).Table [6.7] gives the mean square errors of all the five estimators for the three designs.We see that the NW estimator is the most efficient estimator across all the three probability designs considered. We note that the equally weighted Fuller estimator is more efficient than the Pwml estimator for the U-shaped design though less efficient across all the designs than the Ml and NW estimators.

We now look at the conditional properties of the five estimators.

(ii) **CONDITIONAL ANALYSIS**

The conditional plots of the group means of the Pwml,Ml,ewf and the nonparametric regression estimators plotted against Δ_{zz}^{*F} and Δ_{zz}^{F} are given in figures [6.15]-[6.19].

FIG 6.15 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,ml}$ VS GROUP MEANS OF Δ_{zz}^{F} (20 groups,50 samples per group).



Group means of Δ_{zz}^{F} .

FIG 6.16 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,pwml}$ VS GROUP MEANS OF Δ^{*F}_{zz} (20 groups, 50 samples per group).

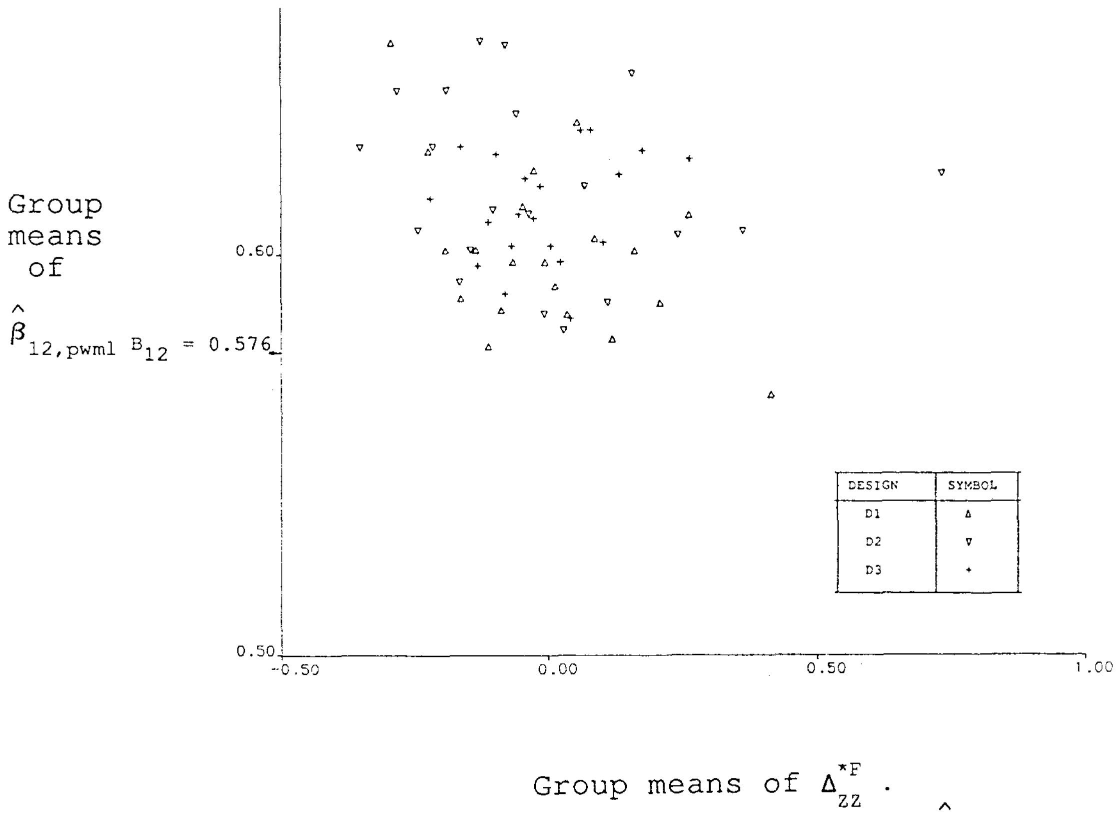


FIG 6.17 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,ewf}$ VS GROUP MEANS OF Δ^{F}_{zz} (20 groups, 50 samples per group).

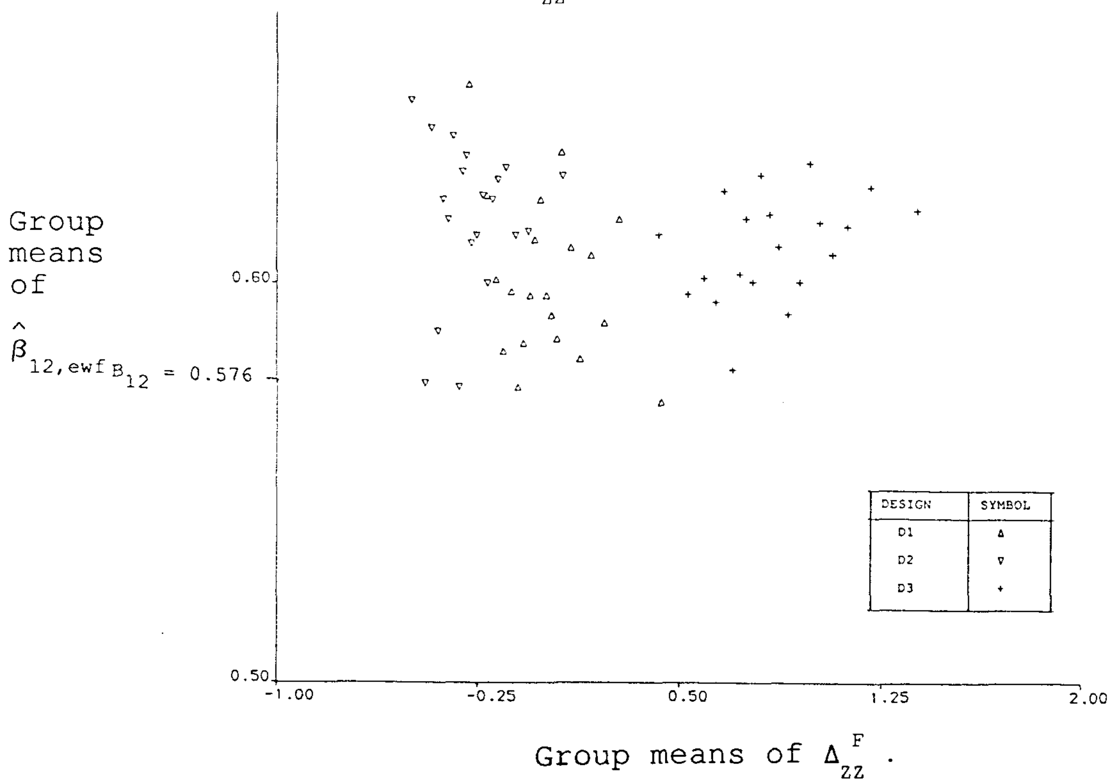


FIG 6.18 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,nw}$ VS GROUP MEANS OF Δ_{ZZ}^{F} (20 groups, 50 samples per group).

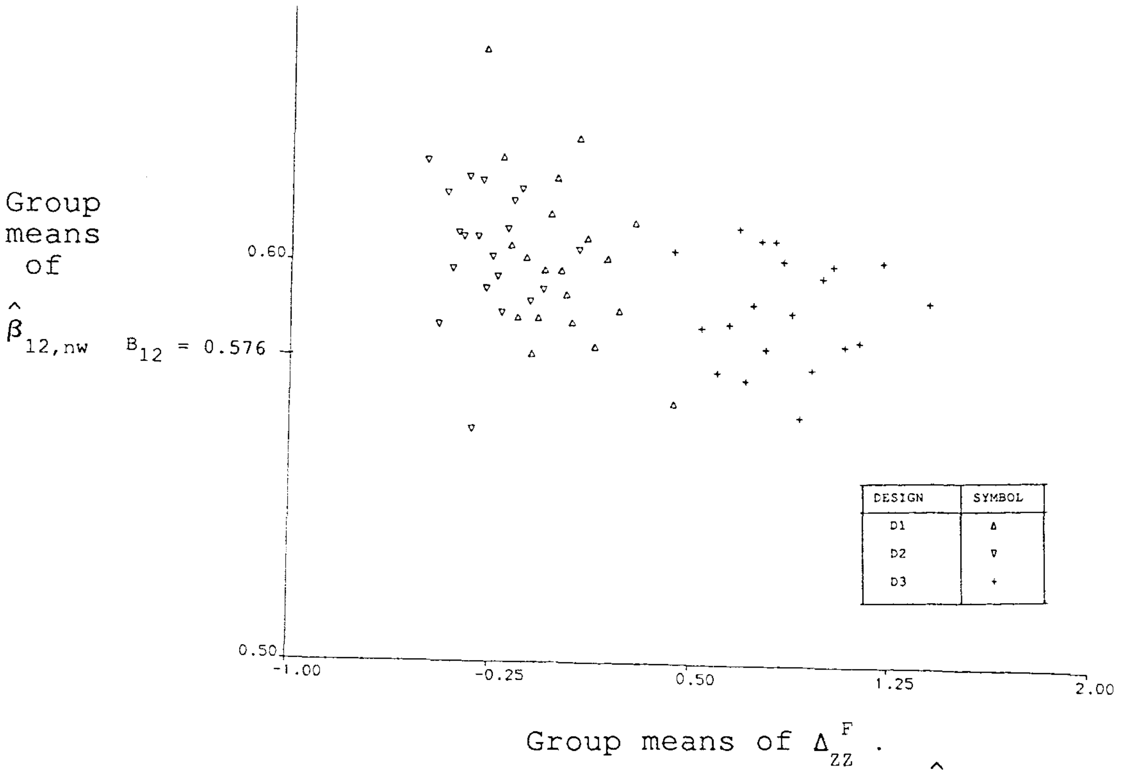
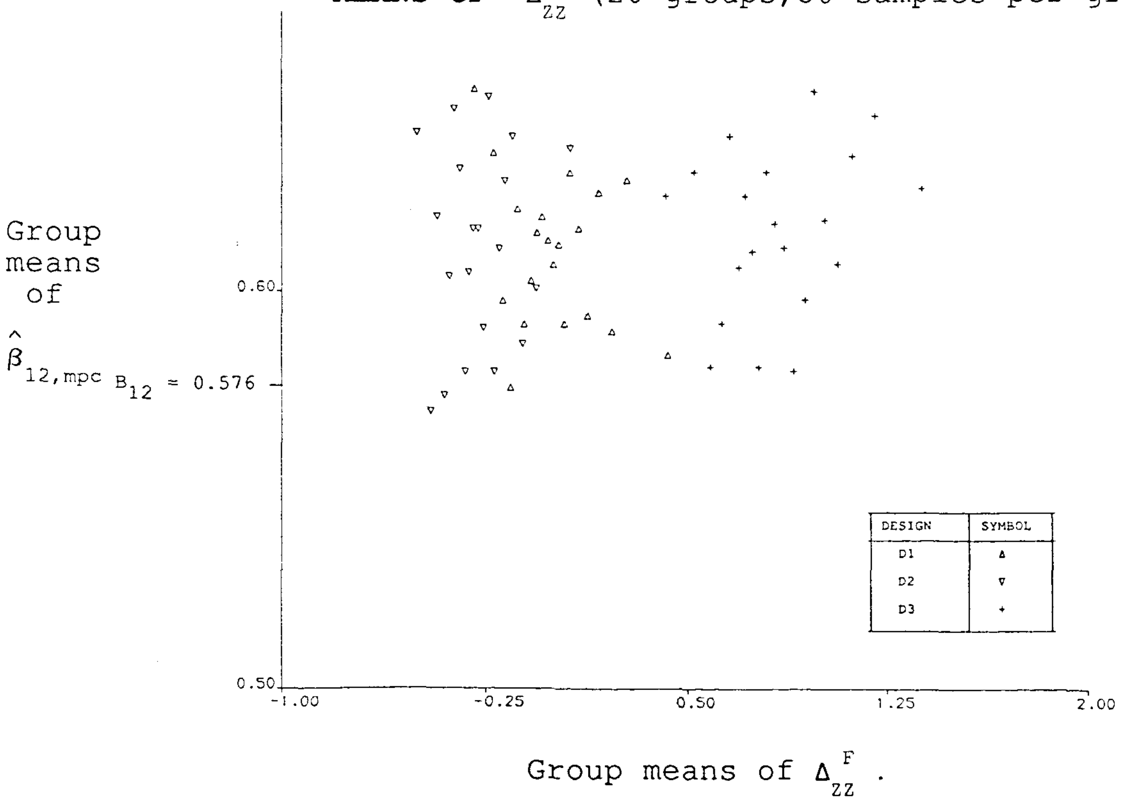


FIG 6.19 SCATTERGRAM OF GROUP MEANS OF $\hat{\beta}_{12,mpc}$ VS GROUP MEANS OF Δ_{ZZ}^{F} (20 groups, 50 samples per group).



We see from figures [6.15]-[6.19] that the Ml estimator looks conditionally unbiased for all the designs,the NW estimator is upward biased for the proportional and increasing allocation designs,but is approximately conditionally unbiased for the U-shaped design,and the MPC looks upward biased for all the designs.The Pwml and ewf are conditionally biased for all the designs.

Which of the five estimators to choose across all the sample designs will depend on the unconditional mean square errors and also may depend on the sample size taken.We investigated whether the observations we made above for a fixed sample size also holds for different sample sizes across the three designs.

TABLE 6.8 UNCONDITIONAL ABSOLUTE BIASES OF THE FIVE ESTIMATORS OF REGRESSION COEFFICIENT OVER 1000 REPLICATIONS FOR EACH SAMPLE SIZE AND DESIGN.

N=6962,n=100 TRUE VALUE β_{12}=0.576

DESIGN	SAMPLE SIZE	ABSOLUTE BIASES OF $\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,pwml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,nw}$	$\hat{\beta}_{12,mpc}$
	50	0.0477	0.0477	0.0494	0.0281	0.0445
D1	100	0.0245	0.0245	0.0249	0.0056	0.0190
	150	0.0165	0.0165	0.0163	0.0037	0.0181
	50	0.0399	0.0555	0.0581	0.0230	0.0454
D2	100	0.0260	0.0408	0.0401	0.0060	0.0350
	150	0.0221	0.0338	0.0326	0.0033	0.0326
	50	0.0352	0.0520	0.0735	0.0286	0.0645
D3	100	0.0128	0.0355	0.0326	0.0072	0.0388
	150	0.0072	0.0307	0.0268	0.0024	0.0315

TABLE 6.9 UNCONDITIONAL STANDARD DEVIATION OF THE FIVE ESTIMATORS OF REGRESSION COEFFICIENT OVER 1000 REPLICATIONS FOR EACH SAMPLE SIZE AND DESIGN.

DESIGN	SAMPLE SIZE	STANDARD DEVIATIONS $\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,pwml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,nw}$	$\hat{\beta}_{12,mpc}$
D1	50	0.144	0.144	0.144	0.141	0.167
	100	0.111	0.111	0.111	0.111	0.133
	150	0.097	0.097	0.097	0.097	0.118
D2	50	0.139	0.164	0.135	0.137	0.184
	100	0.106	0.132	0.105	0.108	0.147
	150	0.087	0.118	0.088	0.092	0.134
D3	50	0.141	0.151	0.142	0.140	0.162
	100	0.110	0.122	0.113	0.111	0.138
	150	0.093	0.110	0.094	0.094	0.119

TABLE 6.10 UNCONDITIONAL MEAN SQUARE ERRORS OF THE FIVE ESTIMATORS OF REGRESSION COEFFICIENT OVER 1000 REPLICATIONS FOR EACH SAMPLE SIZE AND DESIGN.

SAMPLE DESIGN	SIZE	MEAN SQUARE ERRORS $\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,pwml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,nw}$	$\hat{\beta}_{12,mpc}$
	50	0.0232	0.0232	0.0232	0.0206	0.0299
D1	100	0.0130	0.0130	0.0129	0.0117	0.0190
	150	0.0098	0.0098	0.0097	0.0094	0.0143
	50	0.0209	0.0299	0.0216	0.0194	0.0359
D2	100	0.0120	0.0192	0.0126	0.0121	0.0229
	150	0.0083	0.0150	0.0087	0.0084	0.0150
	50	0.0211	0.0256	0.0256	0.0205	0.0303
D3	100	0.0125	0.0161	0.0138	0.0123	0.0205
	150	0.0087	0.0131	0.0088	0.0089	0.0152

We note from tables [6.8]-[6.10] that as expected the absolute biases,mean square errors and the standard deviations decreases as the sample size is increased .This pattern is similar across all the sample designs considered. From table [6.10] we see that the Ml and NW estimators are more efficient than the other estimators across all the sample sizes and also across the three probability designs.So whatever the sample size taken the choice should be between the NW and Ml estimators.The Nw estimator seem to be more efficient for small sample sizes and as the sample size is increased the Ml estimator becomes slightly more efficient across all the designs.

To study the performance of the NW estimator further we considered other unequal probability designs,with the same real population but smaller in size.

SIMULATION STUDY 3.

We selected a simple random sample of size 2000 from the real population and considered this sample as our population of size 2000.As in section 6.3.1 we stratified this population into five equal sized strata according to the increasing values of z such that the first stratum contains the first 400 smallest z values and so on upto the fifth stratum which contains the last 400 units with the highest z values.A sample of size 100 was selected using the following sample designs;

	sample designs.	n_1	n_2	n_3	n_4	n_5
D4	increasing allocation	13	15	20	25	27
D5	increasing allocation	5	5	10	30	50
D6	U-shaped allocation	30	15	10	15	30
D7	U-shaped allocation	47	2	2	2	47

We will present only the unconditional mean square errors of the four estimators for the real population of size 2000.

Since this population is a random sample from the real population,we used the same optimal bandwidth parameters for the two designs which we obtained in section 6.3.2.

TABLE 6.12 UNCONDITIONAL MEAN SQUARE ERRORS OF THE FOUR ESTIMATORS OF REGRESSION COEFFICIENT OVER 100 REPLICATIONS FOR EACH DESIGN.

N=2000,n=100

SAMPLE DESIGN	MEAN SQUARE ERRORS $\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,pwml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,nw}$
D4	0.0122	0.0139	0.0142	0.0120
D5	0.0078	0.0154	0.0169	0.0069
D6	0.0094	0.0094	0.0102	0.0090
D7	0.0255	0.0082	0.0202	0.0077

We see from table 6.12 that the Nadaraya-Watson estimator of the regression coefficient is the most efficient across the four sample designs considered here.We can therefore conclude that the NW estimator is the most efficient robust estimator across all the sample designs considered,when the multivariate normal model assumptions are violated.

In chapter 4 we found that when only the linearity assumption is violated then the design based estimators are the most efficient for the increasing allocation designs,the model based estimators were found to be the most efficient for the U-shaped sample designs and the equally weighted Fuller estimator was found to compromise the efficiency properties of the two types of estimators.We also found that when only the Homoscedastic assumption is violated then the equally weighted Fuller estimator is the most efficient adjusted estimator across all the sample designs considered.The question we ask ourselves is;

Does the NW estimator which was found to be the most efficient among the estimators considered when the structure of the population is unknown still the best when,

(i) only the linearity assumption is violated?

(ii) only the homoscedasticity assumption is violated?

To investigate this we carried out a simulation study 1 in chapter 4 and generated a nonlinear homoscedastic and a linear heteroscedastic populations of size 6962.We then carried out simulation study 3 using the sample designs used in section 6.3.1. As was the case in previous studies the first task is to choose an appropriate bandwidth parameter for each design in the two populations.On similar lines as in previous sections we found that the optimal bandwidth parameters in the case of the nonlinear homoscedastic population were 0.6,1.0 and 2.0 for the sample designs D1,D2 and D3 respectively.For the linear heteroscedastic population the optimal bandwidth parameters were 0.6,2.0 and 1.0 for the sample designs D1,D2 and D3 respectively.We present below the unconditional mean square errors of the Ml,Pwml,ewf and Nw regression estimators averaged over 100 repeated samples.

TABLE 6.13 (nonlinear homoscedastic model).

UNCONDITIONAL MEAN SQUARE ERRORS OF THE FOUR ESTIMATORS OF REGRESSION COEFFICIENT OVER 100 REPLICATIONS FOR EACH DESIGN.

N=2000,n=100

SAMPLE DESIGN	MEAN SQUARE ERRORS $\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,pwml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,nw}$
D1	0.0073	0.0073	0.0074	0.0074
D2	0.0162	0.0143	0.0150	0.0118
D3	0.0047	0.0077	0.0048	0.0046

TABLE 6.14 (linear heteroscedastic model)

UNCONDITIONAL MEAN SQUARE ERRORS OF THE FOUR ESTIMATORS OF REGRESSION COEFFICIENT OVER 100 REPLICATIONS FOR EACH DESIGN.

N=2000,n=100

SAMPLE DESIGN	MEAN SQUARE ERRORS $\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,pwml}$	$\hat{\beta}_{12,ewf}$	$\hat{\beta}_{12,nw}$
D1	0.0080	0.0080	0.0082	0.0083
D2	0.0130	0.0177	0.0179	0.0113
D3	0.0110	0.0082	0.0065	0.0063

We see from tables [6.13] and [6.14] that for both populations the nonparametric Nw kernel estimator is the most efficient estimator across all the sample designs.

6.4 CONCLUSION

In this chapter we compared various nonparametric estimators with the design based estimators and parametric estimators and found that the Nadaraya-Watson nonparametric kernel regression estimator to be more efficient than the design based estimator when the data is multivariate normal and the best estimator in terms of minimum mean square error across all the sample designs when the normality model assumptions are violated.As in chapter 4 the parametric model based estimators were found not to be robust in some circumstances. We also found that the NW estimator is approximately unconditionally unbiased across all the sample designs considered for the real population,but is severely unconditionally biased when the data is linear.Thus the Nadaraya-Watson estimation procedure is a model based procedure which is robust and and efficient when the data is not normal.In comparison with the design based procedures which were found to be robust in the unconditionally distribution framework but can be inefficient in some cirmumstances ,the NW estimator is definitely a better procedure,if one is worried about the efficiency of the estimation procedure.Among all the estimators of the regression coefficient so far studied in the literature ,our proposed model based nonparametric estimator is the most efficient and robust estimator across various sample designs.

CHAPTER 7

CONCLUSIONS AND RECOMMENDATIONS

7.1 SUMMARY OF THE CONCLUSION

The aim of this thesis has been to evaluate various procedures for estimating the regression coefficient in the marginal distribution of the survey variables $\tilde{\underset{\sim}{y}}$ for data collected from complex surveys.We have assumed that the aggregate model is the best fit for the population data,thus fitting one regression equation to the whole data set.We have also assumed that the sample design used to select samples from the population is noninformative,i.e the selection mechanism depends on the design values only,which are assumed known apriori for the whole population.

Our main interest in the evaluation of these estimation procedures for analysing data from complex surveys has been to look for a 'robust' procedure which does not depend on the assumptions made about the distribution describing the population structure.

In this thesis;

(1) We have shown in chapter one that all the various approaches to statistical inferences for analysing data f rom complex surveys can be obtained as special cases of one big model,we called it the *global model*.Thus we have linked all the various statistical approaches to inference together.

(2) In chapter 2,we extended Skinner's(1983) work on predictors of the covariance matrix of the survey variables by deriving consistent predictors of the

covariance matrix under the linear homoscedastic model assumptions, which relaxes the multivariate normality assumptions made by Skinner(1983).

(3) In chapter 3 we derived the asymptotic properties of the estimators proposed by Fuller(1982) under the following models;

(i) Linear homoscedastic model,

(ii) Quadratic homoscedastic model

and the

(iii) linear heteroscedastic model.

We verified the validity of these asymptotic results derived in chapter 3, in a simulation study reported in chapter 4. In this simulation study, we found that;

(i) All the estimation procedures studied in this work were found to be robust for equal probability designs. We however noted that if we decide to use only this class of designs, we will be excluding more efficient sample designs in this class. For example under the quadratic homoscedastic model assumptions, the maximum likelihood estimator is more efficient for the U-shaped design than the equal probability design.

(ii) The ordinary least squares estimator was found to be severely biased for the unequal probability designs both conditionally and unconditionally. We do not recommend the use of this estimator for analysing data from complex surveys.

(iii) The design based procedures advocated by Kish and Frankel (1974) were unconditionally robust but their conditional properties may be poor in some circumstances. We also found that in some circumstances, these procedures are very inefficient. However if one is not worried about efficiency and conditional properties, then these procedures are the best to use for analysing sample survey data.

(iv) The parametric model based adjusted procedures like the maximum likelihood estimator ,were found to be conditionally and unconditionally approximately unbiased under the linear homoscedastic model.
However when only the linearity assumption is violated,then these procedures are severely conditionally and unconditionally biased for the increasing allocation designs and for all unequal probability design when only the homoscedasticity assumption is violated.We can therefore conclude that these procedures are not robust to the violations of the underlying model assumptions in some circumstances.Only under those circumstances where these procedures are robust can we recommend their use.

(v) The weighted model based procedures were found to be unconditionally robust but their conditional properties when the underlying model assumptions were violated,may be poor.We also did not find any significant gain in efficiency using the weighted model based procedures over the design based procedures.

(vi) The equally weighted Fuller estimator was found to be robust conditionally and unconditionally when only the homoscedasticity assumption was violated.However when only the linearity assumption is violated, then this estimator was found to be biased conditionally and unconditionally for the unequal probability designs.Our empirical results indicates that this estimator was the most efficient and robust adjusted estimator to use when it is known that the structure of the population is linear and heterocedastic.We also found that it compromises the efficiency properties of the design based procedures and the model based procedures for the unequal probability designs,when only the linearity assumption is violated.

(4) Since we found that neither the design based procedures nor the parametric model based procedures are best in all circumstances,we proposed a nonparametric model based procedure which does not make assumptions about the distribution of the population structure,in chapter 5.In chapter 6,we carried out a simulation study to compare these nonparametric procedures with the parametric and design based procedures.We found that one nonparametric model based procedure,namely Nadaraya -Watson kernel estimator was efficient and approximately unconditionally unbiased when the linear homoscedastic assumptions were violated.However when the population is linear and homoscedastic,this estimator was found to be unconditionally biased.We therefore recommend the use of these nonparametric procedures for analysing data from complex surveys,if efficient robust estimation procedures are required for analysis.

Thus this work did not end up with a best procedure in all circumstances,but the nonparametric model based procedures looks more promising than the parametric procedures,and may be with further improvements we might be able to make them universally acceptable in all situations.

In the course of this study we detected some areas for improving these nonparametric model based procedures.In the next section we list some areas for future research.

7.3 <u>RECOMMENDATIONS FOR FURTHER WORK</u>

Below we give possible extensions to this work;

(1) In this work we have developed a non-parametric model based procedure for estimating regression coefficients. This can be extended to other methods of multivariate data analysis like principal components, correlation analysis etc.

(2) We also have considered only the univariate case where we had only one design variable. Theoretically it is possible to extend this work to cases where q>1 and p>1, for nonparametric model based procedures.

(3) The nonparametric model based procedure we proposed was based on fixed bandwidth for the kernel function. It will be interesting to investigate whether kernels with variable bandwidth, which takes into account the sparcity of the data will improve this procedure.

(4) Scott and Smith (1969), Royall(1976b) extended the parametric model based procedures to clustered populations, it is also possible to extend these nonparametric model based procedures to clustered populations.

(5) In our empirical studies we did not look at the coverage properties of the nonparametric estimators. To be able to do this we need to derive the estimators of the variance of these nonparametric estimators. Thus this work can be extended to the estimation of variance of these estimators.

(6) We derived the nonparametric model based procedures only for linear smoothers, these methods can be extended to other classes of smoothers.

In a preliminary study we extended the nonparametric model based procedures to another class of smoothers, the smoothing splines. *splines* as defined by Eubank[1988] and others, are piecewise polynomials subjected to a maximum number of continuity constraints. Thus a spline is a piecewise polynomial whose different segements have been joined (tied) together at the knots $\zeta_1,...,\zeta_k$ in a fashion which insures certain continuity properties.

To conform these spline smoothers with our proposed estimators given in equation [5.17] and [5.18] we used Wahbas'(1975) kernel approximation formulae for spline smoothers.

Wahba [1975] has shown that the smoothing spline estimator can be represented approximately as a linear function of the data values y_j. Thus there exists a weight function $w_\lambda(z,z_j)$ such that

$$\hat{f}(z)=n^{-1}\sum w_\lambda(z,z_j)y_j. \qquad [7.1]$$

let $w_\lambda(u)=\lambda w(u/\lambda)$

where λ denotes the bandwidth parameter and the function w(t) is defined as

$$w(t)= 0.5 \exp(-|t|/1.41)\sin((|t|/1.41)+\pi/4) \qquad [7.2]$$

and has the following properties;

$$\left.\begin{aligned} &\int w(t)dt=1. \\ &\int tw(t)dt=0. \\ &\int w^2(t)dt<\infty. \\ &\int t^2w(t)dt=\int t^3w(t)dt=0. \\ &\text{and } \int t^4w(t)dt=-1. \end{aligned}\right\} \qquad [7.3]$$

We see that the properties of the function w(.) in [7.2] are similar to those given for the kernel function in [5.10] but can take negatives values as well.

Thus the smoothing splines estimator corresponds approximately to a kernel type estimator of order 4. Eubank(1988) have shown that if the function f(.) is assumed to be periodic then $\hat{f}(z)$ corresponds to a spline estimator with fixed bandwidth parameter k and weights

$$w_\lambda(z,z_j)=kw(u/k) \quad \text{where } k=\lambda^{1/4}. \qquad [7.4]$$

i.e the estimator corresponding to the periodic smoothing splines is

$$\hat{f}_p(z)=n^{-1}\sum_j kw(z-z_j/k)y_j.$$

where w(.) is as defined in [7.2].We note that the sum of the weight $n^{-1}w_\lambda(z,z_j)$ does not sum to one.To remove this anomaly we divide the weights by their sum so that if we denote the modified weights by $w_{m\lambda}(z,z_j)$ then

$$w_{m\lambda}(z,z_j)=w(z-z_j/k)/\sum_j w(z-z_j/k). \qquad [7.5]$$

therefore the periodic spline estimator of the function f(z) is given by;

$$\hat{f}_{mp}(z)=\sum_j w(z-z_j/k)y_j/\sum_j w(z-z_j/k).$$

Since the spline estimators corresponds approximately to the kernel estimators,their asymptotic properties can be deduced from those of the kernel estimators derived in section [5.4].

In a preliminary investigation of the splines we carried out a simulation study with real data given in section [6.3.2] chapter 6.Below we present the unconditional results of the approximate spline estimator of the regression coefficients,together with the maximum likelihood and the Nadaraya-Watson estimator,for the three probability designs defined in chapter 6.We used the same bandwidth parameters as those for the Nadaraya-Watson estimator as optimal bandwidths for the spline estimator.

TABLE 7.1

UNCONDITIONAL MEANS OF THE THREE ESTIMATORS OF REGRESSION COEFFICIENT OVER 1000 REPLICATIONS FOR EACH DESIGN.

N=10000, n=100 TRUE VALUE B_{12}=0.595

SAMPLE DESIGN	MEAN SQUARE ERRORS $\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,spline}$	$\hat{\beta}_{12,nw}$
D1	0.0245	0.1518	0.0056
D2	0.0260	0.0922	0.0060
D3	0.0128	0.1309	0.0072

TABLE 7.2

UNCONDITIONAL STANDARD DEVIATIONS OF THE THREE ESTIMATORS OF REGRESSION COEFFICIENT OVER 1000 REPLICATIONS FOR EACH DESIGN.

N=10000, n=100

SAMPLE DESIGN	STANDARD DEVIATIONS $\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,spline}$	$\hat{\beta}_{12,nw}$
D1	0.111	0.777	0.111
D2	0.106	0.330	0.108
D3	0.111	0.679	0.112

TABLE 7.3

UNCONDITIONAL MEAN SQUARE ERRORS OF THE THREE ESTIMATORS OF REGRESSION COEFFICIENT OVER 1000 REPLICATIONS FOR EACH DESIGN.

N=10000,n=100

SAMPLE DESIGN	MEAN SQUARE ERRORS $\hat{\beta}_{12,ml}$	$\hat{\beta}_{12,spline}$	$\hat{\beta}_{12,nw}$
D1	0.0130	0.6265	0.0120
D2	0.0120	0.1175	0.0117
D3	0.0125	0.4775	0.0123

From tables [7.1]-[7.3] we see that the approximate spline kernel estimator is severely biased and very inefficient.One reason why this approximate spline estimator is biased may be because the kernel function assign negative weights to some observations thus introducing severe bias in their prediction.

These results indicates that the kernel approximation for splines may not be appropriate.Further work is required to check how the spline estimators behave in their own right.

REFERENCES

Anscombe,F.J (1960)
Rejection of outliers.Technometrics,2 123-147

Aitkin,A.C.(1934)
Note on selection from multivariate normal population.Proc.Edin.Math.Soc. 4 106-110

Aitkin,A.C.(1935)
A further note on multivariate selection. Proc.Edin.Math.Soc. 5 263-269

Aitchison,J and Dunsmore,I.R (1975)
Statistical prediction analysis. Cambridge university press,Cambrigde.

Anderson,T.W.(1957)
Maximum likelihood estimates for a Multivariate normal distribution when some observations are missing. J.Amer.Statist.assoc.52 200-203

Anderson,T.W.(1971)
An introduction to Multivariate Statistical Analysis.Wiley,New York

Bebbington,A.C.and Smith,T.M.F.(1977)
The effect of survey design on multivariate analysis.In O'Muircheartaigh and Payne (1977b)

De Boor,C.(1978)
A practical guide to splines. New York, Springer-Verlag.

Brillinger,D.R.(1979)
A continous form of post stratification. Ann.Inst.Statist.Math 31 Part A 271-277

Brewer,K.R.W.and Mellor,R.W.(1973)
The effect of sample structure on analytical surveys.Aust.J.Statist.15 145-152

Brewer,K.R.W.(1979)
A class of Robust designs for large scale surveys.J.Amer.Statist.assoc.74 911-915

Binder,D.A.(1983)

On the variances of asymptotically normal estimators from complex surveys. Int.Statist.rev.51 279-292

Benedetti,J.(1974)
Kernel estimation of regression functions. Doctoral dissertation University of Washington Seattle.

Benedetti,J.(1975)
Kernel estimation of regression functions. Proc.Comp.Sc.and Statist.8th annual Symposium on the interface 405-412

Benedetti,J.(1977)
On nonparametric estimation of regression functions.J.Roy.Statist.Soc.B. 39 248-253

Buckley,M.J.and Eagleson,G.K.(1988)
The estimation of the variance in nonparametric regression.Biometrika 75 189-199

Buja,A,Hastie,T and Tibshirani R (1989).
Linear smoothers and additive models. Ann.statist.17 453-555.

Basu,D.(1969)
Role of sufficiency and likelihood principles in sample survey theory.Sankhya A 31 441-454.

Bartlett,M.S.(1963)
Statistical estimation of density functions. Sankhya A 25 245

Box,G.E.P (1953)
Non normality and test on variances. Biometrika 31 318-335

Butler,R.W (1986)
Predictive Likelihood inference with applications. J.Roy.Statist.Soc.B.48 1-38

Cochran,W.G.(1977)
Sampling Techniques 3rd edition Wiley New York.

Cassel,C.M.,Sardal,C.E.,(1972)
A model for studying robustness of estimators and informativeness of labels

in sampling with varying probabilities. J.Roy.Statist.Soc.B.34 279-289

Cassel,C.M.,Sardal,C.E.,Wretman,J.H.(1977)
Foundations of inference in survey sampling. Wiley New York.

Cassel,C.M.,Sardal,C.E.,Wretman,J.H.(1979)
Prediction theory for finite populations when model based and design based principles are combined.Scand.J.Statist. 6. 97-106

Campbell,C (1977)
Properties of the ordinary and weighted least squares estimators of regression coefficients for two stage sampling. Proc.Social.Statist.sect.Amer.Statist.Assoc .800-805.

Chambers,R.L (1986)
Design adjusted parameter estimation. J.Roy.Statist.Soc.A.149 161-173.

Cheng,K.F and Lin,P.E (1981)
Nonparametric estimation of regression function.Zeitschrift für wahrscheinlichkeits theorie und verwandte gebiete 57 223-233

Cleveland,W.S (1979)
Robust locally weighted regression and smoothing scatterplots. J.Amer.Statist.Assoc 74 829-836.

Cohen,A.C (1951)
Restriction and selection in multinormal distributions. Ann.of Math.Statist.28 731-741

Cumberland,W.G and Royall,R.M.(1981)
Prediction models and unequal probability sampling.J.Roy.Statist.Soc.B.43.353-367

Demets,D.and Halperin,M.(1977)
Estimation of a simple regression coefficient in samples arising from a subsampling procedure.Biometrics 33 47-56

Dempster,A.P.,Laird N.M.,Rubin,D.B.,(1977)
Maximum likelihood estimation from incomplete data via EM algorithm.

J.Roy.Statist.Soc.B 39 1-38

Dawid,A.P (1977)
Conditional independence in statistical theory.
J.Roy.Statist.Soc.B 41 1-31.

Eubank,R.L.(1988)
Spline smoothing and nonparametric regression. Marcel Dekker inc.New.York.

Ericson,W.A.(1969)
Subjective Bayesian models in sampling from finite population.J.Roy.Statist.Soc. B 31 195-225

Friendman,J.A and Stuelzle,W (1982)
Projection pursuit regression.
J.Amer.Statist.Assoc.76 817-823

Fuller,W.A (1973)
Regression analysis for sample surveys.
Int.assoc.of survey statisticians.1st meeting. Viena,August,1973.

Fuller,W.A (1975)
Regression analysis for sample surveys.
Sankhya C 37 117-132

Fuller,W.A (1982)
Regression estimation of regression equations. manuscript.

Gasser,T.Muller,H.G.and Mammitzsch(1985)
Kernels for nonparametric curve estimation.J.Roy.Statist.Soc. B 47 238-252.

Gasser,T,Sroka and Jennen-steinmetz,c (1986)
Residual variance and residual pattern in nonlinear regression.Biometrika 73 625-633

Gasser,T. and Muller H.G (1979)
Kernel estimation of regression functions, Smoothing techniques for curve estimation. eds.T.Gasser and M.Rosenblatt,New York; Springer-Verlag, 23-68

Gasser,T. and Engel,J.(1990)
The choice of weights in kernel regression estimation.Biometrika 77 377-381.

Godambe,V.P.(1966)

A new approach to sampling from finite population.J.Roy.Statist.Soc.B.28 310-319

Gross,A.L and Perry,P (1983)
Validating a selection test,a predictive probability approach. Psychometrika 48 113-127

Godambe,V.P.(1982)
Estimation in survey sampling:Robustness and optimality.J.Amer.Statist.Assoc.77 393-403

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

Hardle,W and Gasser,T (1984)
Robust nonparametric function fitting. J.Roy.Statist.Soc.B.46 42-51.

Hartley,H.O. and Sielken R.L. (1975)
A superpopulation viewpoint for finite population sampling.Biometrics 31 411-422

Hall,P. and Marron,J.S.(1990)
On variance estimation in non-parametric regression.Biometrika 77 415-419

Hampel,F.R (1971)
A general qualitative definition of robustness.Ann.Math.Statist.42 1887-1896

Holt,D. and Smith,T.M.F.(1976)
The design of surveys for planning purposes.Aust.J.Statist.18 37-44

Holt,D.,Smith,T.M.F. and Winters,P.D (1980)
Regression analysis of data from complex surveys.J.Roy.Statist.Soc.A 143 474-487.

Hausman,J.A and Wise,D.A (1977)
Social experimentation,truncated distributions and efficient estimation. Econometrica 45 319-339

Holmes,D (1987)
The effect of selection on the robustness of multivariate methods.

Unpublished Doctoral thesis, University of Southampton, U.K.

Huber, P.J (1964)
Robust estimation of a location parameter. Annals.Math.Statist.35 73-101

Huber, P.J (1981)
Robust statistics, New York, Wiley.

Hutchison, M.F, and De hoog, F.R (1985)
Smoothing noisy data with spline functions. Numeische Mathematik, 47 99-106.

Hinkley, D.V (1979)
Predictive likelihood. Ann.Statist.7 718-728

Isaki, C.T. and Fuller, W.A (1982)
Survey design under the regression superpopulation model J.Amer.Statist.Assoc. 77 89-96

Jennen-steinmetz, C. and Gasser, T. (1988)
A unifying approach to nonparametric regression estimation. J.Amer.Statist.Assoc.83 1084-1089.

Jonnrup, H and Rennermalm, B (1976)
Regression analysis in samples from finite populations. Scand.J.Statist. 3 33-37

Kendall, M.G and Buckland W.R (1981)
A dictionary of statistical terms. Prepared by International Statistical Institute, Oliver and Boyd, Edinburgh 4th edition.

Kussmaul, K (1969)
Protection against assuming the wrong degree in polynomial regression. Technometrics 11 677-682

Konijn, H.S (1962)
Regression analysis in sample surveys. J.Amer.Statist.Assoc.57 590-606.

Koch, G.K, Freeman, D.H and Freeman, J.L (1975)
Strategies in the Multivariate analysis of data from complex surveys. Int.Statist.Rev.43 59-78

Kish,L.and Frankel,M.R. (1974)
Inference from complex surveys.
J.Roy.Statist.Soc.B.36 1-36

Little,R.J.A (1982)
Models for nonresponse in sample surveys.
J.Amer.Statist.Assoc.77 237-250.

Lauritzen,S.L (1974)
Sufficiency,prediction and extreme models.
Scand.J.Statist.1 128-134

Lawley,D.N.(1943)
A note on Karl Pearson's selection formulae.
Proc.Roy.Soc.Edin.A 62 28-30

Lederman,W.(1938)
Sampling distribution and selection in a normal population.Biometrika 30 295-304.

Mardia,K.V.,Kent,J.T and Bibby,J.M (1979)
Multivariate Analysis Academic press London.

Mckennell,A.C (1969)
Methodological Problems in a survey of aircraft noise annoyance.The Statistician 19 1-29.

Muller,H.G (1984)
Smooth optimal kernel estimators of densties, regression curves and modes.Ann.Statist.12 766-774

Mukhopadhay P.(1978)
Estimation of the variance of a finite population under a superpopulation model.
Metrika 25 115-122

Muller,H.G and Stadtmuller,U.(1987)
Estimation of heteroscedasticity in regression analysis.Ann.Statist.15 610-635

Nadaraya,E.A (1964)
On estimating regression.
Theory of probability application 9 141-142

Nadaraya,E.A (1964)
Remarks on nonparametric estimates for for density functions and regression curves.
Theory of probability application 15 134-137

Nadaraya,E.A (1965)
On nonparametric estimation of density functions and regression curves. Theory of probability application 10 186-190.

Nathan,G. and Holt,D (1980)
The effect of survey design on regression analysis.Journal of the Royal Statistical Society B 42 377-386

Pfefferman,D.J and Holmes D.(1985)
Robustness consideration in the choice of a method of inference for regression analysis of survey data.J.Roy.Statist.Soc A 148 268-278.

Pfefferman,D.J and Smith T.M.F.(1985)
Regression analysis for populations in cross sectional surveys. Int.Statist.Rev.53 37-59.

Pearson,K.(1903)
On the influence of natural selection on the variability and correlation of organs. Phil.Trans.A.200 1-66

Priestley,M.B.and Chao,M.T (1972)
Nonparametric function fitting. J.Roy.Statist.Soc.B.34 385-392

Parzen,E.(1962)
On the estimation of the probability density function and mode.Ann.Math.Statist.33 1065-1076

Quenouille,M.H (1956)
Notes on bias in estimation. Biometrika 43 353-360

Rao,C.R.(1973)
Linear statistical inference and its applications. 2nd edition Wiley,New York.

Rao J.N.K (1975)
Comment on the paper by A.P Dempster "A Subjective look at robustness" Bulletin of the Int.Statist.Inst.Proc.of the 40th

session pp387.

Rice,J (1984)
Boundary modification for kernel regression. Comm.Statist. A 13 893-900

Rosenblatt,M (1956)
Remarks on some nonparametric estimate of density functions. Annals.Statist.27 832-837.

Royall,R.M.(1970)
On finite population sampling theory under certain linear regression models.Biometrika 57 377-389.

Royall,R.M.(1971)
Linear regression models in finite population sampling theory.In V.P.Godambe and D.A Sprott (Eds) Foundations of Statistical inference. Holt,Rinehard and Winston,Toronto.259-274

Royall,R.M.(1976a)
Likelihood functions in finite population sampling theory..Biometrika 63 605-614.

Royall,R.M.(1976b)
The linear least squares prediction approach to two stage sampling.J.Amer.Statist.Assoc.71 657-664.

Royall,R.M.and Herson,J.(1973a)
Robust estimation in finite populations I .J.Amer.Statist.Assoc.68 880-889.

Royall,R.M.and Herson,J.(1973b)
Robust estimation in finite populations II .J.Amer.Statist.Assoc.68 890-893.

Rubin,D.B (1985)
Iteratively reweighting least squares. Entry in Encyclopedia of statistical sciences,Vol 4 Kolz,s,Johnson,N.L and Read,C.B eds,New York.

Singh,R.S (1979)
Mean squared errors of the estimates of a

density and its derivatives.Biometrika 66, 177-180.

Schuster,E.F,(1972)
Joint asymptotic distribution of the estimated regression function at a finite number of distinct points.
Annals of math.statist. 43 84-88

Scott,A.J.(1977)
On the problem of randomization in survey sampling.Sankhya C,39 1-9

Scott,A.J. and Smith,T.M.F.(1969)
Estimation in Multistage surveys.
J.Amer.Statist.Assoc 64 830-840 .

Skinner,C.J.(1981)
Estimation of the variance of a finite population for cluster samples.Sankhya B 392-398.

Skinner,C.J.(1982)
Multivariate analysis of sample survey data.Unpublished Doctoral thesis University of Southampton.

Skinner,C.J.(1983)
Multivariate prediction from selected samples.Biometrika 70 289-292

Skinner,C.J.,Holt,D.and Smith,T.M.F (1989)
Analysis of complex surveys.Wiley,New York.

Silverman,B.W (1985)
Some aspects of the spline smoothing approach to nonparametric curve fitting.
J.Roy.Statist.Soc.B 47 1-52

Silverman,B.W (1986)
Density estimation for statistics and data analysis. Chapman and Hall,New York.

Smith,T.M.F.(1976)
The foundations of Survey sampling.
J.Roy .Statist.Soc.A 139 183-195.

Smith,T.M.F.(1978)
A model building approach to survey analysis.European meeting of statisticians. Oslo,August,1978.

Smith,T.M.F.(1979)
Statistical sampling in Auditing;A statistician viewpoint. The statistician 28 267-280.

Smith,T.M.F.(1981)
Regression analysis for complex surveys.In Current topics in survey sampling.D.Krewski ,J.N.K.Rao and R.Platek(Eds) Academic Press, New York, 267-292.

Smith,T.M.F.(1983)
On the validity of inference from nonrandom samples.J.Roy .Statist.Soc.A 146 394-403.

Smith,T.M.F.(1988)
To weight or not to weight,that is the question.In Bayesian statistics 3rd edition J.M Bernardo,M.H.Degroot,D.V Lindley and A.F.M Smith,Oxford,Oxford University Press.

Stone,C.J (1977)
Constant nonparametric regression. Annals.Statist.5 505-545.

Sugden,R.A (1979)
Inference on symmetric functions of exchangeable populations.J.Roy.Statist.Soc.B 41 269-273.

Sugden,R.A and Smith,T.M.F (1984)
Ignorable and informative designs in survey sampling inference. Biometrika 71 495-506.

Tapia,R.A and Thompson,J.R (1978)
Nonparametric probability density estimation. John Hopkins University Press,Baltimore and London.

Tallis,G.M (1978)
Note on robust estimation in finite population. Sankhya C 40 136-138.

Watson,G.S (1964)
Smooth regression analysis. Sankhya A 359-372.

Wahba,G (1975)

Smoothing noisy data with spline functions.
Numer.Math. 24 383-393.