A MODEL - BASED APPROACH TO ESTIMATION OF FINITE POPULATION TOTAL USING LOCAL LINEAR POLYNOMIAL REGRESSION ESTIMATOR

BY

KITHIKII KASUNGO

A Dissertation Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Statistics of Kenyatta University.

JULY 2002

DECLARATION

This is my original research work and has not been presented for a degree award or any other award either in part or in whole in any other University.

SIGNATURE: ..... Date: 13. 08. 2002

NAME: KITHIKII KASUNGO

This dissertation has been submitted for examination with our approval as University supervisors:

SIGNATURE: ..... Date: 13/8/2002

NAME: DR. ROMANUS ODHIAMBO, Ph.D.

DEPARTMENT OF MATHEMATICS AND STATISTICS,

JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY.

SIGNATURE: ..... Date: 14/8/02

NAME: DR. CHARLES WAFULA, Ph.D.

DEPARTMENT OF MATHEMATICS,

KENYATTA UNIVERSITY.

DEDICATION

I dedicate this work to my beloved wife **Agnes Kalekye** whose care, encouragement and professional acumen spurred me to work with unflagging zeal; my daughters **Mary Muli** and **Flora Kalimi,** my sons **Daniel Kivoto** and **Tony Kasungo** who inspired me to treat my academic work as worship. Their continued patience, tolerance, inspiration, support and commitment to my education enabled me to finish this work successfully. Given that the course was self-sponsored, they had to ceaselessly sacrifice a lot to make the course a success. May **God** bless them all in abundance.

ACKNOWLEDGEMENTS

This work marks a significant milestone in my journey of life. My sincere thanks first go to the Almighty God for the gifts of Life, Piety, Fortitude, Wisdom, Knowledge, Understanding and His provision at all times of need.

Many people have helped me in many ways in making this work what it is. I mention a few whose pivotal assistance went directly into realizing the final document. To my supervisors **Dr. Romanus Odhiambo** and **Dr. Charles Wafula**, for their scholarly and professional guidance, tireless efforts, availability, patience and interest, particularly at the formative stage, when everything looked strange to me. Their continued support guided me to the very end. Their pieces of advice, critical insights, devotion of energy and readiness to discuss my work at length were inestimable and very inspiring to me, particularly when things seemed to hit a dead end. I will be forever grateful.

I am equally grateful to all members of the academic staff of Mathematics Department, Kenyatta University. My special and invaluable appreciation go to **Dr. Njenga**- the Chair Mathematics Department, **Dr. Odongo**, **Dr. Kahiri**, **Mr. Ruto** and the late **Mr. Githu** who taught me and also provided relevant materials as well as useful comments.

Special thanks go to **Mr. Tony Gacheru** of the Kenyatta Virtual University at Kenyatta University. His technical support enabled me to complete this work in time. I wish to express my gratitude for the extremely valuable help I received from **John Nderitu** of Central Bureau of Statistics, Nairobi – Kenya and from all those who have been so kind as to

check for the accuracy of the data included in this research. To **Mr. Samson Kaplelach, Mr. Kennedy Asembo** and **Mr. George Maina** and for typing this work from its manuscript stage, I wish to express my most profound gratitude for their support and assistance.

Special mention goes to my colleagues **Muriithi Karuku** and **Benjamin Muema**. They were of good company throughout the journey. Their support, co-operation, criticisms and suggestions contributed immensely to the successful completion of this course. I owe them a very special gratitude.

I wish to acknowledge the support I received from my family members during the entire period of the research. My parents, **Mr. Kasungo Mukunu** and **Mrs. Kalimi Kasungo** for their prayers. Their unwavering perseverance in nurturing me to maturity and teaching me the essence of hard work contributed positively to the success of this work; my brothers **Boniface, Moses** and **Ambrose**; my sisters **Rose**, **Ruth** and **Stella** and my mother-in-law, **Mrs. Mary Muli** were of immense moral support and help. I thank them so much for the selfless and ceaseless sacrifices they made in my favour. To my late sisters **Juliana**, **Florence** and my late father-in-law, **Mr. Daniel Kivoto** for their love and kindness, I will never cease to be grateful. *Immortality lies not in the things you leave behind, but in the people that your life touched* (Anon, 1949).

I am indebted to all persons whose names do not appear here but who in one way or the other contributed to this study, I deeply express my sincere appreciation to them and say my sincere gratitude is no less.

Finally, I wish to stress that any errors, gaps, omissions, weaknesses, obscurities or selection of topics which remain in this research work are entirely my own responsibility.

May the Grace of our **Lord God** be upon you all … abundantly.

ABSTRACT

Estimation of finite population totals in the presence of auxiliary information is considered. A class of estimators based on local polynomial regression is proposed. Like generalized regression estimators, these estimators are weighted linear combination of study variables, in which the weights are calibrated to known control totals, but the assumptions on the super population model are considerably weaker. The estimators are shown to be asymptotically model-unbiased and consistent under mild assumptions.

Simulation experiments indicate that the local polynomial regression estimator is more efficient than regression estimators when the model regression function is incorrectly specified, while being approximately as efficient when the parametric specification is correct.

TABLE OF CONTENTS

CHAPTER ONE

1.0 OVERVIEW

1.1 Introduction

Sample surveys are concerned with obtaining desired information from a population. In practice finite populations are of interest to researchers in Medicine, Economics, Politics and many other aspects of life.

Ideally, total information in the population is obtained from a census where every individual in the population is involved in giving information. However, most of the time due to certain constraints like cost, time, literacy level and other geographical factors, it is not always possible to carry out a census effectively. Statistical methods of extracting information from the finite population have been developed. In these methods, part of the population, referred to as a sample, is used and the information about the population is inferred from the sample.

The theory of sample survey aims at developing sampling strategies that result in selection of a sample that is a good representation of the whole population. It provides procedures for making statistical inferences about the study variable otherwise referred to as the survey variable. It is also used in determining the criteria for comparing different strategies in order to obtain optimal results from a sample survey.

1.2 Terminologies used

1.2.1 Infinite population

This is a list or frame denoted by $U=(U_1,U_2\ldots,U_N)$ of N identifiable units. In infinite population, N is unknown since it is infinitely large.

1.2.2 Finite population

This is a list or frame denoted by $U=(U_1, U_2, \ldots, U_N)$ of N identifiable units. In finite population, N is less than infinity and is usually known.

1.2.3 Target population

This is a population whose information is required. It can also be defined as a set of all individuals whose information is required.

1.2.4 Sampled population

This is the set of all the individuals in the sampling frame.

1.2.5 Sample

We assume that there exists a population frame or list denoted by $U=(U_1, U_2, \ldots, U_N)$ of N identifiable units. A sample is then defined as a subset s of U selected according to a rule, P(s). If the selection rule employs randomisation, then we have a random sampling design, but the notation embraces purposive design where P(s) is unit for $s = s_0$ and zero otherwise. The canonical survey design is simple random sampling where,

$$P(s) = \begin{cases} \binom{N}{n}^{-1} & \text{for all samples of size n} \\ 0 & \text{otherwise} \end{cases}$$

We note that $n \leq N$, $s = \{i_1, i_2, \ldots, i_n\}$ where n is the sample size and i_i denotes the ith unit in sample.

1.2.6 Sampling design

If $S = \{i_1, i_2, \ldots, i_n\}$ is a sample of size n distinct units and $p(s \mid z, w)$ is the probability of selecting sample s where w may include survey variables, $z = (z_1, z_2, \ldots, z_N)$ is any prior information about the units, then we define a probability sampling design to consist of:

(a) A sampling plan such that each member of the population has a known probability greater than zero of inclusion in the sample, and

(b) Procedures for inference such that for reasonably large samples the correctness of inference does not depend on an assumed model.

1.2.7 Sampling Units

In sample surveys, the population individuals are denoted by $(U_1, U_2, \ldots, U_N)$ whereas the individuals in the sample are denoted by $(u_1, u_2, \ldots, u_N)$. Therefore the sampling units can be seen as subsets or sub-divisions of the whole population such that $U_i \cap U_{j \neq i} = \emptyset$.

This implies that $\bigcup_{i=1}^{N} U_i = U$ where U is a collection of N units U_i'S

1.2.8 Identifiable units

Units of a finite population are said to be identifiable if they can be uniquely labelled from 1 to N and the label of each unit is unique. The labelling can be conceptual rather than real as in areas on a map, line transect or a compilation of separate lists. There is a 1-1 correspondence between the units and the indices 1, 2, ..., N such that the population comprises of $U=(U_1, U_2, \ldots, U_N)$.

1.2.9 Sampling frame

This is the list of all individuals in the finite population. It is from the sampling frame that a surveyor selects a sample. The list should be clear and concise so that the sampling units could be identified unambiguously. Sometimes the formation of a frame is not achievable in some populations such as in birds and fish for they are ever mobile and are not labelled. Thus sampling these populations requires special treatment. Since these populations are not fixed, and there are no frames, sampling cannot be controlled, and selection is by assumption rather than by design, the standard assumption being either simple random sampling or stratified simple random sampling. Seber (1982) argues that the absence of the labels is tackled by repeatedly sampling the target population and tagging, or labelling, the sampled animals. The number of the marked animals and when they were marked are then recorded in the second subsequent surveys. The information from this partially labelled sample can be employed to estimate demographic characteristics of the population, especially its size, under certain

assumptions such as equi-catchability, and constant immigration and death rates. If there are s samples then there are 2^s-1 possible records for sampled animals that are never captured.

The problem is to estimate properties of the missing group. If the population is closed, with no immigration or emigration, then the population size N is fixed but unknown value, which is a key target parameter. If the population is open, then the population size at the onset may be fixed but subsequently varies over sample occasions according to the migration and emigration processes.

1.2.10 Inclusion probabilities

For all individuals in the finite population, define

$$I_i(s) = \begin{cases} 1 & \text{if unit i in the population is in the selected sample} \\ 0 & \text{otherwise} \end{cases}$$

Then we define $$\Pi_i = P\{i \in s\} = \sum_{s \ni i} P(s)$$

and call Π_i the first order inclusion probability.

$$\text{Let} \quad \Pi_{ij} = P\{i, \ j \in s\} = \sum_{s \ni i,j} P(s)$$

then Π_{ij} is referred to as the second order inclusion probability. It is the joint inclusion probability of the i^{th} and the j^{th} units of the target population. A sample obtained by such a design is called a probability sample. In order for a sampling design to be called a probability

sampling design, Π_i must be strictly positive. A sampling design is then said to be measurable if

(a) $\Pi_i > 0$ and (b) $\Pi_{ij} > 0$.

In probability proportional to size sampling without replacement Πps, we find that a good Πps design requires that

(i) $\Pi_i \propto x_i$ where $x_i > 0$ is a measure of size.

(ii) $\Pi_{ij} > 0$

(iii) $\Pi_{ij} < \Pi_i \Pi_j$, $i \neq j$

(iv) The inclusion probabilities Π_i and Π_{ij} are easily computed.

(v) The joint inclusion probabilities should not depend on the order of the unit in U.

Condition (i) defines a Πps sample, and the corresponding Horvitz-Thompson estimator of the population total is

$$\hat{T}_{HT} = \sum_s \frac{y_i}{\Pi_i} \quad \text{.....(1.0)}$$

When conditions (ii) and (iii) hold, non-negative unbiased estimators exist, and so these are the key statistical conditions. Condition (i) uses the auxiliary information at the design stage and will yield an efficient sampling scheme when $Y_i = kx_i$ where k is some constant. For fixed sample size designs these conditions follow directly from the Yates-Grundy-Sen form for the variance of the Horvitz-Thompson estimator, namely

$$V_{\Pi ps} = \sum_{i \neq j} \sum (\Pi_i \Pi_j - \Pi_{ij}) \left(\frac{Y_i}{\Pi_i} - \frac{Y_j}{\Pi_j} \right)^2 \quad \text{.....(1.1)}$$

1.2.11 Characteristic of interest

A finite population is a collection of N identifiable units labelled $U=(U_1, U_2, \ldots, U_N)$. Attached to each unit $U_i \in U$ is a vector of survey variables with values y_i for $i=1,2,\ldots,N$. We let $\underline{Y}=(y_1, y_2, \ldots y_N)^T$. We also write $\underline{Y} = (y_s, \bar{y}_s)^T$ the partition of $\underline{Y}$ induced by the selection of the units in s. The sample data is the set of labels and associated values, which we denote by

$$d_s = \{(i, y_i); i \in s\} \quad \ldots\ldots(1.2)$$

We thus call the survey variable $\underline{Y} = (y_1, y_2, \ldots, y_N)^T$ the study variable. Using the survey variable $\underline{Y} = (y_1, y_2, \ldots, y_N)^T$ we estimate some well-defined descriptive functions of $\underline{Y} = (y_1, y_2,, y_N)^T$ the so-called characteristics of interest. The common characteristics of interests are:

i) The finite population total

$$Y = \sum_{i=1}^{N} y_i$$

ii) The finite population mean

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

iii) The finite population variance

$$V(Y) = \frac{1}{N}\sum_{i=1}^{N} (y_i - \bar{y})^2$$

1.2.12 Auxiliary information

Associated with each unit of $U=(U_1, U_2, \ldots, U_N)$ is a certain known characteristic vector $\underline{X} = (x_1, x_2, \ldots, x_N)^T$ referred to as auxiliary information and is positively correlated to the study variable $\underline{Y}=(y_1, y_2, \ldots, y_N)^T$. The auxiliary information is known before hand. Thus the technique is to use the obtained sample, plus the auxiliary information to make inference about $\underline{Y}=(y_1, y_2, \ldots, y_N)^T$ or a function of $\underline{Y}$.

Auxiliary information is the additional information for improving sample survey processes. It may be information on the characteristic of whole units of the population from which the sample is drawn or on the ancillary variables in relation to the study variable. For example, if the study variable is income, the auxiliary information may be expenditure on food, rent, transport etc, or previous census, if the study variable is the current census.

When auxiliary information is available for each and every unit of the population, it can be used in several ways to improve the efficiency of the estimators of the study variables. One way of using auxiliary information is found in the sample selection through unequal probability sampling, population stratification, double sampling and clustering of units. It is also used in estimation, for example in ratio, regression, nonparametric and Horvitz-Thompson methods of estimation.

There could also exist q characteristics positively correlated to $Y=(y_1, y_2, \ldots, y_N)^T$, giving us a N x q matrix of covariates; $X_{nxq}=((x_{ij}))_{nxq}$

1.3 Selection of the sample

A realized sample s can be regarded as an outcome of a random variable s which takes as its elements any of the 2^N subsets. If we let S denote the set of all possible samples from a finite population $U=(U_1, U_2,, U_N)$ and p(s) the probability that a sample s is drawn, then we say that a probability sampling design assign each $s \in S$ a probability $p(s) \geq 0$ such that

$$\sum_s p(s) = 1$$

Samples are selected by simple random sampling with or without replacement in which case each unit in the population has an equal probability of being selected.

If the units of the population vary considerably in size, the simple random sampling may not be appropriate since it does not take into account the possible importance of the larger units in the population. There are various ways of solving this problem. One is to assign an unequal probability of selection to the different units of the finite population. If units vary in size and the variable under study is proportional to size, then the probabilities of selection may be assigned in proportion to the size of the unit.

A selection procedure in which units are selected with probability proportional to the measure of size is known as sampling with probability proportional to size (pps).
The procedure of selecting a sample of size n with unequal probabilities is as follows.

If X_i is an integer proportional to size of the i^{th} unit, i=1, 2, ..., N, we form successive cumulative totals

$y_1 = x_1$

$y_2 = x_1 + x_2$

.
.
.

$y_N = x_1 + x_2 + \ldots + x_N$

we then draw a random number R not exceeding $\sum_{i=1}^{N} x_i = Y_i$ from a table of random numbers. If $Y_{i-1} < R < Y_i$, the i^{th} unit is selected. The procedure is repeated n times until a sample of size n is obtained.

1.4 Sample survey estimation problem

In many cases the sample is used to describe the real population from which it was selected by estimating population totals and other descriptive statistics such as odd ratios or correlations. This will be referred to as descriptive inferences. In other cases inferences may explore properties of the processes that generate the finite population values. This is a classical statistical inference analysis and will be termed analytic inferences.

1.4.1 Descriptive inferences

This involves the estimation of some well-defined descriptive functions of $Y = (y_1, y_2, \ldots, y_N)^T$ the so called 'parameters' of the finite population.

The common functions of interest are:

(i) The finite population total $Y = \sum_{i=1}^{N} y_i$

(ii) The finite population mean $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$

(iii) The finite population variance $V(Y) = \frac{1}{N}\sum_{i=1}^{N} (y_i - \bar{y})^2$

1.4.2 Analytic inference

In this case, inferences may explore properties of the process that generate the population values. One approach is to assume that the finite population has been generated by a super population model, f (y, x, θ) say, and the interest centers on estimation of parameters (θ) of the super population model. This is classical statistical analysis and will be termed analytic inference.

The super-population model can be employed to predict the unobserved values y_i, $i \in \bar{s}$, and this mode of inference is a form of predictive inference. If say $f(x; y; \theta) = \alpha + \beta x_i$, where $\theta = (\alpha, \beta)^T$ then the problem would be to estimate the super-population parameters α and β. Having obtained the estimates of α and β i.e $\hat{\alpha}$ and $\hat{\beta}$ respectively and using the known auxiliary information x_i, i = 1,2,..., N we predict the unobserved values y_i, $i \in \bar{s}$.

1.5 Approaches to sample survey estimation problem

The twin problems of sample survey design and of finite population inference is tackled in two ways, viz:

i. The classical approach, otherwise referred to as the randomisation approach

ii. The predictive approach otherwise referred to as the super-population approach.

There are, however, fundamental differences in the way these two approaches view the population units.

1.5.1 The Classical Approach or Randomisation Inference

The framework for randomisation inference is the distribution of results of all the possible samples that could have been drawn using the random sampling scheme, p(s).

The distribution depends on the $\underline{Y}$, the population matrix of values of the survey variable, which in general, is unknown. In practice, sample sizes are large and it is asserted that a normal distribution will be a good approximation to the underlying randomisation distribution of point estimators. This distribution is determined by the first and second order inclusion probabilities of p(s).

In this approach, each population unit U_i, i=1,2,…, N is associated with a fixed but unknown real number which is the value of the variable under study. Inference in this is thus based on the observed quantities $\underline{Y}= (y_1, y_2, \ldots, y_N)^T$ which were initially chosen to the sample

through the design, p(s). The implication here is that inference will be tied to the chosen design p(s). Hence the classical approach is design-based. The set-up in the classical approach assumes that the population units are labelled and the statistician has access to the true and fixed value of y_i of the unit U_i. This is quite demanding and unjustifiable.

1.5.2 The Predictive Approach or Model-Based Approach or Super-Population Approach

The model based approach starts from the assumption that the measurement of interest in the finite population can be treated as a realized value of a random variable Y. The randomness is introduced directly into the y-values.

In this approach, a super-population model is inherent in any given finite population. The model employed characterize the actual values, both the observed and the unobserved which are considered as a realization of random variables $y_1 y_2, \ldots, y_N$. The relationship among the variables is expressed as a model of the joint distribution of the random variables.

In this approach, the first problem is to construct a stochastic model for $\underline{Y}$ that incorporates within it all the population information contained in the known prior values of $\underline{Y}$. If $\underline{Y}$ contains indicators for strata or clusters then this information should be incorporated into the model for $\underline{Y}$. Once a model is chosen, a sample selection scheme consistent with that model and taking into account practical considerations such as costs will be employed to draw a sample s.

As before, we denote the sampling rule of this type by P (s/y). The problem is then to use the model, the sample data and the information in the sampling scheme to make a predictive inference about the unobserved random variables y_i, $i \in \bar{s}$, in order to estimate functions of the finite population values such as the mean of $\underline{Y}$, which is now a random variable. The choice of the model and its robustness to misspecification is the major issue. Small deviations from a chosen model may lead to serious errors in an inference. So the model based approach depends on assumptions and used naively will not be robust.

A super-population model traditionally used in sample survey is:

$E(y_i)=\beta x_i$, $i=1, 2, \ldots, n$

$Var(y_i)=\sigma^2(x_i)$, $i=1, 2, \ldots, n$

$Cov(y_i, y_j)=0$, $i \neq j$, $j=n+1, n+2, \ldots, N$ ....(1.3)

Under this model an optimal estimator of the population mean is the ratio estimator defined by:

$$\hat{T}(y) = \overline{Y}_R = \frac{\bar{y}}{\bar{x}} . \overline{X} \qquad \text{.....(1.4)}$$

where $\bar{x}, \bar{y}$ and $\overline{X}, \overline{Y}$ are the sample and population means respectively.

1.6 Classical Approach

1.6.1 An estimator $\hat{T}(y)$

In a classical approach, an estimator $\hat{T}(y/s)$ is seen as a real valued function defined on $Sx\Re^N$ where $s \in S$ depends on $\underline{Y}$ through $y_{i's}$ for which unit u_i, that is, i^{th} unit in the population occurs in the sample s. We thus calculate the proposed estimator $\hat{T}(y/s)$ based on the observed quantities $\underline{Y} = (y_1, y_2, \ldots, y_n.)^T$

The estimators of the population mean $\bar{Y}$ and total Y are given by

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad \text{and}$$

$y = N\bar{y}$ (i.e expansion estimator)

$$= \frac{N\sum_{i=1}^{n} y_i}{n} \quad \text{respectively}$$

The key point here is that the estimators are basically functions of the $y_{i's}$ in the sample s.

1.6.2 Unbiasedness of $\hat{T}(y)$

An estimator $\hat{T}(y)$ based on the design P(s) is said to be design unbiased for T(Y) if

$$E_p(\hat{T}(y)/s) = \sum_{s \in S} \hat{T}(y)p(s)$$

$$= T(Y)$$

where $E_p(\hat{T}(y)/s)$ denote the conditional expectation of $\hat{T}(y)$ given that the sample s is chosen through some design P(s).

Theorem 1

An estimator $\hat{T}(y) = \sum_{s \in S} l_{s_i} y_i$ for the population total is unbiased for T(Y) where $\sum_{i \in s} l_{s_i} p(s) = 1$, $i \leq 1 \leq N$, l_{s_i}'s do not depend on the y_i's although they may be functions of x_i's and i is the ith unit (B.K. Sinha 1991, sec 2.4)

Proof

$$E_p[\hat{T}(y)/s] = \sum_{s \in S} \hat{T}(y)p(s)$$

$$= \sum_{s \in S} p(s) \sum_{i \in s} l_{s_i} y_i$$

$$= \sum_{s \in S} Y_i \sum_{i \in s} l_{s_i} p(s)$$

$$= \sum_{i=1}^{N} Y_i, \text{ since } \sum_{i \in s} l_{s_i} p(s) = 1$$

$$= T(Y)$$

Hence $\hat{T}(y)$ is an unbiased estimator for T (Y). However, if $\hat{T}(y)$ is biased for T(Y)

then we have

$E_p[\ \hat{T}(y)/s] = T(Y) +$ bias term which we can write as

$E_p[\ \hat{T}(y)/s] = T(Y) + B_p[\ \hat{T}(y)/s]$ and as such the bias will be given by

$B_p\ [\hat{T}(y)/s] = E_p[\ \hat{T}(y)/s] - T(Y)$

1.6. 3 The Variance and Mean Squared Error (MSE) of $\hat{T}(y)$

The variance of $\hat{T}(y)$ is given by

$$\text{Var}\left[\hat{T}(y)/s\right] = E_p\left\{\left[\hat{T}(y)/s\right] - E_p\left[\hat{T}(y)/s\right]\right\}^2 \quad \text{.....(1.5)}$$

However, if the estimator is not unbiased, then its mean squared error is given by

$$MSE_p[\hat{T}(y)/s] = E_p[\{\hat{T}(y)/s\} - T(Y)\ ]^2 \quad \text{.....(1.6)}$$

$$= E_p[\{\hat{T}(y)/s\} - E_p\left\{\hat{T}(y)/s\right\} + E_p\ \left\{\hat{T}(y)/s\right\} - T(Y)]^2$$

$$= E_p[\{\hat{T}(y)/s\} - E_p\left\{\hat{T}(y)/s\right\}]^2 + E_p[E_p\left\{\hat{T}(y)/s\right\} - T(Y)]^2$$

$$= \text{Var}_p[\hat{T}(y)/s] + [B_p\left\{\hat{T}(y)\right\}]^2 \quad \text{.....} \quad (1.7)$$

1.6.4 The Criterion for comparing competing strategies

In the randomisation approach, the pair $\left[P,\hat{T}(y)\right]$ denotes a sampling strategy where $\hat{T}$(y) depends on P. The performance of any strategy $\left[P,\hat{T}(y)\right]$ is judged through the minimization of (1.6) and (1.7) stated in the previous page.

The problem that arises immediately here is if we opt to use the minimization criterion then we shall have difficulties in choosing between an unbiased estimator with a small variance and the biased estimator with a small mean square error.

1.7 The Prediction Approach

1.7.1 An estimator $\hat{T}$(y)

Let $\hat{T}$(y) be an unbiased estimator of T(Y). Then an estimator $\hat{T}$(y) is said to be model unbiased for T (Y) if E_m [$\hat{T}$(y)/s, $\underline{Y}$]= E_m [$\hat{T}$(y)] where E_m [$\hat{T}$(y) / s, $\underline{Y}$] denotes the conditional expectation of $\hat{T}$(y) given sample (s, $\underline{Y}$) with respect to a given model. Suppose $\hat{T}$(y) is biased then the bias of $\hat{T}$(y) is given by $B_M\left[\hat{T}(y)\right]$.

Thus $B_M\left[\hat{T}(y)\right] = E_m\left[\hat{T}(y) - T(Y)\right]$ .....(1.8)

1.7.2 The Variance and MSE of $\hat{T}(y)$

Let $\hat{T}$ (y) be the unbiased estimator of T (Y) . Under prediction approach, variance of $\hat{T}$ (y)is given by

$$Var_m[\hat{T}(y)/s, \underline{Y}] = E_m\{[\hat{T}(y)/s, \underline{Y}] - E_m[\hat{T}(y)/s, \underline{Y}]\}^2, \quad \ldots\ldots(1.9)$$

whereas the mean square error (MSE) is given by

$$MSE_m[\hat{T}(y)/s, \underline{Y}] = Var_m[\hat{T}(y)/s, \underline{Y}] + \{B_M[\hat{T}(y)/s, \underline{Y}]\}^2 \ldots\ldots(1.10)$$

If however, the bias is Zero, then

$$MSE_m[\hat{T}(y)/s, \underline{Y}] = Var_m[\hat{T}(y)/s, \underline{Y}] \ldots\ldots(1.11)$$

1.8 Towards a Compromise

The desire to reconcile the classical approach and the super-population approach is based on the fact that these two approaches are not necessarily opposing. Thus a blend of the two can be used to produce optimal results.

Godambe and Thompon (1977) suggested the quantity

$$E_M E_P[\hat{T}(y) - T(Y)]^2 \quad \ldots\ldots(1.12)$$

which could be used for minimization purposes in our search for optimal strategies. Sundberg (1994) advocated the "mean square error" of predicted squared errors as an universal criterion for the choice of variance estimator, V, say. He argued that this choice should be based on the predictive criterion of minimizing

$$E_{P,m}[\{\hat{T}(y) - T(Y)\}^2 - V]^2. \ldots\ldots(1.13)$$

which is analogous to the use of the anticipated variance proposed by Isaki and Fuller (1982) for choosing point estimators.

Expanding (1.12) above, we have

$$E_P E_m [\hat{T}(Y) - T(Y)]^2 = E_P E_m ([\hat{T}(y) - E_m \{\hat{T}(y)\} + E_m \{\hat{T}(y)\} - T(Y)])^2$$

$$= E_P [E_m \{\hat{T}(y) - E_m [\hat{T}(y)]\}^2 + \{E_m [\hat{T}(y)] - T(Y)\}^2]$$

$$= E_P [Var_m \{\hat{T}(y)\}] + E_P [B_m \{\hat{T}(y)\}]^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1.14).$$

Remark

(i) E_p and E_m can be interchanged since p does not depend on $Y_{i's}$.

(ii) We note that (1.14) provides a measure of uncertainty based on model and design based approaches. This makes a good expression for searching for optimal strategies.

If the problem of non-response and measurement errors is ignored, then the survey inference is based on two processes: the process that generates the population values and the process that selects the sample.

Randomisation approach is based on the assumption that the population values are unknown fixed constants so that there is no model generating the population values. In this case the only probabilistic information resides in the sample selection probabilities under random sampling.

Model based inferences assume that there is a generation process for the population values and then bases inference on a model for that process. In addition modellers must consider the sample selection process and then justify ignoring it in every case.

If selection cannot be ignored, then both processes must be used for making inferences. Thus the model-assisted inference is a form of randomisation inference that employs models to determine the point estimators. This model-assisted approach to inference appears to achieve a working compromise between model-based and p-based inference for finite population parameters and is being widely adopted in some areas of survey analysis. In the light of this remark, we adopt the prediction approach in this study. In chapter two, we review model-assisted approach under local polynomial regression method in survey sampling as discussed in Breidt and Opsomer (2000).

1.9 Objectives and Outline of the Project

1.9.1 Objectives of the project

This study was undertaken with the following objectives:

(i) To make use of the model- based approach to estimate finite population total in the presence of auxiliary information based on local linear polynomial smoothing.

(ii) Theoretical results have shown that local linear polynomial smoothing is applicable to a wide range of problems, Breidt and Opsomer (2000). Also shown by the theoretical results are statistical properties in the regression context including design adaptability, consistency and asymptotic unbiasedness. This

study therefore employs this approach in the context of survey sampling to study the properties of the proposed local polynomial regression estimator under conditions applicable in model-based surveys.

(iii) To carry out an empirical study on simulation experiments to compare the performance of our proposed estimator with those of the already established ones both parametric and non-parametric.

1.9.2 Outline of the Project

In chapter one, we review the sample survey estimation problem. We mention that there are two main different approaches to the sample survey estimation problem. As a conclusion of the chapter we mention that it is possible to marry these two approaches as a means of obtaining optimal sampling strategies.

In chapter two, we review the work done by Breidt and Opsomer (2000) and a class of estimators based on local linear polynomial regression proposed is examined. We then show that the estimators are weighted linear combinations of study variables, in which the weights are calibrated to known control totals.

In chapter three, we consider the nonparametric regression estimator based on a weighted local linear polynomial regression in estimation of finite population total. Here we point out that the method has advantages over other popular kernel methods. We in particular, emphasize that this method has the ability of design adaptation and that the local linear regression smoothers have high asymptotic efficiency (that is, can be 100% with a

suitable choice of kernel and bandwidth) among all possible linear smoothers, including those produced by kernel, orthogonal series and spline methods.

In chapter four, we carry out an empirical study to compare the performance of estimators studied in chapter two and three. Seven populations and four different estimators are considered. The first estimators are parametric corresponding to constant and linear whereas the last two are nonparametric. The nonparametric ones prove to be more robust to model violations. This seems to be a major step in solving the robustness problem that has haunted sample survey researchers for a long time.

In chapter five we give suggestions for areas of further study and give concluding remarks of the research work.

CHAPTER TWO

2.0 LOCAL POLYNOMIAL REGRESSION ESTIMATORS IN SURVEY SAMPLING

2.1 Introduction

This chapter describes the use of a new type of model – assisted nonparametric regression estimator (Breidt and Opsomer (2000)) for the finite population total, T(y) based on local polynomial smoothing. It also looks at the theoretical properties of such an estimator.

In many survey problems, auxiliary information is available for all the elements of the population of interest. Indeed, use of auxiliary information in estimating parameters of a finite population of study variable is a central problem in surveys.

One approach to this problem is the super-population approach in which a working model, ξ describing the relationship between the auxiliary variable X and the study variable Y is assumed. Estimators are sought which have good efficiency if the model is true but maintain desirable properties like the asymptotic design – unbiasedness (ADU) i.e. unbiasedness over repeated sampling from the finite population and design consistency if the model is false. Typically, the assumed models are linear, leading to the familiar ratio and regression estimators [e.g., Cochran (1977)], the best linear unbiased estimators [Brewer (1963), Royall (1970)], the generalized regression estimators [Cassel, Särndal and Wretman (1977)], and related estimators [Wright (1983), Isaki and Fuller (1982)].

2.2 The Local Polynomial Regression Estimator

Local polynomial regression is a generalization of kernel regression. Cleveland (1979) and Cleveland and Delvin (1988) showed that these techniques are applicable to a wide range of problems. Theoretical work by Fan (1993) and Ruppert and Wand (1994) showed that these estimators have many desirable theoretical properties including adaptation to the design of the covariates, consistency and asymptotic unbiasedness. Wand and Jones (1995) provide a clear explanation of the asymptotic theory for kernel regression and local polynomial regression. The monograph by Fan and Gijbels (1996) explores a wide range of application areas of local polynomial regression techniques. However, the applications of these techniques to model – assisted survey sampling is new.

2.2.1 The General Framework

Consider a finite population $U=(U_1, U_2, \ldots, U_N)$. For each $U_i \in U$, $i=1, 2, \ldots, N$, an auxiliary variable X_i is available.

Let $T(x)=\sum_{i\in U} x_i$

A sample s is drawn from U, according to a fixed –size sampling design P(s), where p(s) is the probability of drawing the sample s. Let n be the size of s. Assume

$$\Pi_i=P_r\{i \in s\}=\sum_{s:i\in s} P(s) > 0 \text{ and } \Pi_{ij}=P_r\{i, j \in s\}=\sum_{S:i,j\in S} P(s) > 0 \text{ for all } i, j \in U.$$

The study variable y_i is observed for each $i \in s$. The goal is to estimate the finite population total

$$T(Y) \text{ i.e. } T(Y)=\sum_{i\in u} y_i \quad , i = 1,2,3,\ldots,N$$

Let $I_i = \begin{cases} 1 & if\ i \in s\ , i = 1,2,...n \\ 0 & otherwise \end{cases}$

Then $E_p(I_i) = \Pi_i$ where $E_p(.)$ denotes expectation with respect to the sampling design P(s).

Using this notation, an estimator $\hat{T}(y) of\ T(Y)$ is said to be design – unbiased estimator of T (Y) if $E_p\left[\hat{T}(y)\right] = T(Y)$. A well-known design – unbiased estimator of T(Y) is the Horvitz – Thompson estimator,

$$\hat{T}(y) = \sum_{i \in s}\left(\frac{y_i}{\Pi_i}\right),\ i = 1, 2, ..., n \quad \text{.....(2.1)}$$

$$= \sum_{i \in U}\left(\frac{y_i I_i}{\Pi_i}\right) \quad \text{.....(2.2)}$$

(Horvitz – Thompson (1952).

Averaging over all the possible sample values from the finite population we obtain

$$E_p\left(\hat{T}(y)\right) = E_p\left\{\sum_{i \in s}\left(\frac{y_i}{\Pi_i}\right)\right\}$$

$$= E_p\left\{\sum_{i \in U}\left(\frac{y_i I_i}{\Pi_i}\right)\right\}$$

$$= \sum_{i \in U} E_p\left(\frac{y_i I_i}{\Pi_i}\right)$$

$$= \sum_{i \in U} \frac{y_i}{\Pi_i} E_p(I_i)$$

$$= \sum_{i \in U} \frac{y_i}{\Pi_i} \Pi_i$$

$$= \sum_{i \in U} y_i$$

$$= T(Y)$$

This implies that the estimator $\hat{T}(y)$ is design – unbiased for the finite population total T (Y).

The variance of the Horvitz – Thompson estimator under the sampling design is defined as

$$Var\ \hat{T}(y) = \sum_{i,j \in U} (\Pi_{ij} - \Pi_i \Pi_j) \frac{y_i}{\Pi_i} \frac{y_i}{\Pi_j} \quad \text{.....(2.3)}$$

The equation (2.3) may be written as

$$Var\ \hat{T}(y) = \sum_{i=1}^{N} \frac{y_i^2}{\Pi_i}(1 - \Pi_i) + \sum_{i \neq} \sum_{j} (\Pi_{ij} - \Pi_i \Pi_j) \left(\frac{y_i}{\Pi_i} \cdot \frac{y_j}{\Pi_j} \right) \quad \text{.....(2.4)}$$

The equation (2.4) may also be written as

$$Var\ \hat{T}(y) = \sum_{i \neq} \sum_{j} (\Pi_i \Pi_j - \Pi_{ij}) \left(\frac{y_i}{\Pi_i} - \frac{y_j}{\Pi_j} \right)^2 \quad \text{.....(2.5)}$$

We now show that equation (2.4) and (2.5) are equivalent.

$$\hat{T}(y)=\sum_{i\in s}\frac{y_i}{\Pi_i}, \qquad i=1,2,3,...,n$$

$$=\sum_{i\in U}\frac{Y_i}{\Pi_i}I_i, \qquad i=1,2,3,...,N$$

$$Var\,\hat{T}(y)=\text{var}\sum_{i\in U}\left(\frac{Y_iI_i}{\Pi_i}\right), \text{ where } I_i \text{ is an indicator variable}$$

$$=\sum_{i\in U}\frac{Y_i^2}{\Pi_i^2}\text{var}(I_i)+\sum_{i\neq}^{N}\sum_{j}^{N}\frac{Y_i}{\Pi_i}\frac{Y_j}{\Pi_j}\text{cov}(I_i,I_j)$$

$$=\sum_{i\in U}\frac{Y^2}{\Pi^2}(1-\Pi_i)\Pi_i+\sum_{i\neq}^{N}\sum_{j}^{N}\frac{Y_i}{\Pi_i}\frac{Y_j}{\Pi_j}(\Pi_{ij}-\Pi_i\Pi_j)$$

$$=\frac{1}{2}\left[\sum_{i\in U}\frac{Y_i^2}{\Pi_i^2}(1-\Pi_i)\Pi_i+\sum_{j\in U}\frac{Y_j^2}{\Pi_j^2}(1-\Pi_j)\Pi_j+2\sum_{i\neq}^{N}\sum_{j}^{N}\frac{Y_i}{\Pi_i}.\frac{Y_j}{\Pi_j}(\Pi_{ij}-\Pi_i\Pi_j)\right]$$

$$But\ \Pi_i(1-\Pi_i)=\sum_{j(j\neq i)}(\Pi_i\Pi_j-\Pi_{ij})$$

Therefore

$$Var\hat{T}(y)=\frac{1}{2}\left\{\sum_{i\neq}\sum_{j}\frac{Y_i^2}{\Pi_i^2}(\Pi_i\Pi_j-\Pi_{ij})+\sum_{i\neq}\sum_{j}\frac{Y_j^2}{\Pi_j}(\Pi_i\Pi_j-\Pi_{ij})-2\sum_{i\neq}^{N}\sum_{j}^{N}\frac{Y_i}{\Pi_i}\frac{Y_j}{\Pi_j}(\Pi_i\Pi_j-\Pi_{ij})\right\}$$

$$=\sum_{i\neq}^{N}\sum_{j}^{N}\frac{Y_i^2}{\Pi_i^2}(\Pi_i\Pi_j-\Pi_{ij})-2\sum_{i\neq}^{N}\sum_{j}^{N}\frac{Y_i}{\Pi_i}\frac{Y_j}{\Pi_j}(\Pi_i\Pi_j-\Pi_{ij})+\sum_{i\neq}\sum_{j}\frac{Y_j^2}{\Pi_j^2}(\Pi_i\Pi_j-\Pi_{ij})$$

$$=\sum_{i\neq}^{N}\sum_{j}^{N}(\Pi_i\Pi_j-\Pi_{ij})\left(\frac{Y_i}{\Pi_i}-\frac{Y_j}{\Pi_j}\right)^2$$

We note that $\hat{T}(y)$ does not depend on the $\{x_i\}$, that is, x_i's are not incorporated at the estimation stage. It is of interest to improve upon the efficiency of the Horvitz—Thompson estimator by using this auxiliary information.

2.2.2 Motivation

The estimator we use is motivated by modelling the finite population of the $Y_{i's}$ conditioned on the auxiliary variable x_i, a realization from an infinite super-population ξ, in which

$Y_i = m(x_i) + \varepsilon_i$, i=1, 2, …, N

where ε_i are independent and identically distributed random variables with mean zero and variance $v(x_i)$, $m(x_i)$ is a smooth function of $x_{i's}$ and $v(x_i)$ is also smooth and strictly positive. Given x_i, $m(x_i) = E_\xi[Y_i]$ and so is called the regression function while $v(x_i) = Var_\xi(Y_i)$ and so is called the variance function.

Definition: The kernel function, K(.) is defined as a continous, bounded and symmetric real function which integrates to one, i.e. $\int_{-\infty}^{\infty} K(u)du = 1$.

Let $K(u / h_N)$ denote a continuous kernel function and let h_N denote a bandwidth parameter, which controls the amount of smoothing to be done. We then begin by defining the local polynomial kernel estimator of degree q based on the entire finite population. Let $Y_U = [y_i]$, $i \in U$ be the N-vector of $y_{i's}$ in the finite population.

Define the N x (q + 1) matrix,

$$X_U = \begin{pmatrix} 1 & (x_1 - x_i) & . & . & (x_1 - x_i)^q \\ . & & & & \\ . & & & & \\ . & & & & \\ 1 & (x_N - x_i) & . & . & (x_N - x_i)^q \end{pmatrix}$$

$= (1, (x_j - x_i), \ldots, (x_j - x_i)^q)$, where $j \in U$

and define the N x N matrix

$$W_U = diag\left\{\frac{1}{h_N}K\left(\frac{x_j - x_i}{h_N}\right)\right\},\ j \in U$$

Let e_r represent a p+1 vector with 1 in the r^{th} position and 0 elsewhere. The local polynomial kernel estimation of the regression function at x_i, based on the entire finite population is given by

$$\begin{aligned} m_i &= e_i^T\left(X_u^T W_u X_u\right)^{-1} X_u^T W_u Y_u \\ &= W_u^{*T} Y_u \end{aligned} \quad \ldots\ldots (2.6)$$

which is well defined as long as $X^T{}_u W_u X_u$ is invertible and

$$W_u^{*T} = e_i^T\left(X_u^T W_u X_u\right)^{-1} X_u^T W_u$$

If these m_i's were known, then a design-unbiased estimator of the finite population total, T(y) would be the generalized difference estimator

$$T^*(y) = \sum_{i\in s}\frac{y_i - M_i}{\Pi_i} + \sum_{i\in U} M_i \quad \ldots\ldots (2.7)$$

[Särndal, Swensson and Wretman (1992), page 221]. The design variance of the estimator (2.7) is

$$Var\,T^*(y) = \sum_{i,j\in U}\left(\Pi_{ij} - \Pi_i\Pi_j\right)\left\{\left(\frac{y_i - m_j}{\Pi_i}\right)\left(\frac{y_j - m_j}{\Pi_j}\right)\right\} \quad \ldots\ldots (2.8)$$

Which we would expect to be smaller than (2.3); the deviations $\{y_i - m_i\} = \{[m(x_i) - m_i]$

$+ \varepsilon_i\}$ will typically have smaller variation than the $\{y_i\}$ for any reasonable smoothing procedure under the model ξ.

The population estimator m_i is the traditional local polynomial regression estimator for the unknown function m(.), widely discussed in the nonparametric regression literature. In the present context, it cannot be calculated, because only the y_i in $s \subset U$ are known. Therefore, we will replace each m_i by a sample-based consistent estimator.

Let $y_s = [y_i]$, $i \in s$ be the n vector of y_i's obtained in the sample. Define the n x (q+1) matrix

$X_s = [1, x_j - x_i, \ldots, (x_j - x_i)^q]$, $j \in s$.

And define the n x n matrix

$$W_s = diag\left\{\frac{1}{\Pi_j h_N} K\left(\frac{x_j - x_i}{h_N}\right)\right\}, \quad j \in s$$

A sample design-based estimator of m_i is then given by

as long as $X^T_s W_s X_s$ is invertible. and $W_s^{0^T} = e_i^T \left(X_s^T W_s X_s\right)^{-1} X_s^T W_s$

$$\hat{m}_i = e_i^T \left(X_s^T W_s X_s\right)^{-1} X_s^T W_s Y_s$$
$$= W_s^{0^T} Y_s \quad \ldots\ldots (2.9)$$

When we substitute the $\hat{m}_i$ into (2.7), we have the local polynomial regression estimator for the finite population total, T(Y);

$$\hat{T}_{lp}(y) = \sum_{i \in s} \frac{y_i - \hat{m}_i}{\Pi_i} + \sum_{i \in U} \hat{m}_i \quad \ldots\ldots (2.10)$$

The sample estimator in (2.9) differs in one important way from the traditional local polynomial regression estimator. The presence of the inclusion probabilities in the "smoothing weights" W^{o}_{s} makes our sample-based estimator $\hat{m}_i$ a design-consistent estimator of the finite population smoother m_i, which is based on some (not necessarily optimal) bandwidth h_N, considered fixed here for any N. In real survey problems, h_N will rarely be optimal because a single bandwidth would be chosen and used to compute weights to be applied to all study variables. Regardless of the choice of h_N, m_i is a well-defined parameter of the finite population. Specifically, m_i is a function of finite population totals each of which can be estimated consistently by their corresponding Horvitz-Thompson estimators. That is, we have included probability weights in the smoothing weights in order to construct asymptotically design unbiased (ADU) and design-consistent estimators of the finite population smoother m_i. This is consistent with development of the generalized regression estimator (GREG), which our procedure reverts to as the bandwidth becomes large.

In principle, the estimator (2.9) can be undefined for certain $i \in U$, even if the population estimator in (2.6) is defined everywhere: if for some sample s there are less than (q + 1) observation in the support of the kernel at some x_i, then the matrix $X^{T}_{s}W_{s}X_{s}$ will be singular. This is not a problem in practice because it can be avoided by selecting a bandwidth that is sufficiently large to make $X^{T}_{s}W_{s}X_{s}$ invertible at all locations x_i. However, that situation cannot be excluded theoretically as long as the bandwidth is considered fixed for given population. Therefore, for the purpose of the theoretical derivations, we consider an adjusted

sample estimator that is guaranteed to exist for any sample s⊂U. The adjusted sample estimator for m_i is given by

$$\hat{M}_i^* = e_i^T \left(X_s^T W_s X_s + diag\left\{ \frac{\delta}{N^2} \right\}_{j=1}^{q+1} \right)^{-1} X_s^T W_s Y_s$$

$$= W_s^{*T} Y_s \qquad \text{..... (2.11)}$$

for some $\delta>0$ and

$$W_s^{*T} = e_i^T \left(X_s^T W_s X_s + diag\left\{ \frac{\delta}{N^2} \right\}_{j=1}^{q+1} \right)^{-1} X_s^T W_s \ .$$

The term δN^{-2} in the denominator are small order adjustments that ensure the estimator is well defined for all s⊂U. The adjustment was also used by Fan (1993) for the same reason when the x_i are considered random. Another possible adjustment would consist of replacing the usual choice of a kernel with compact support by one with infinite support such as the Gaussian kernel. In practice, however, such kernels have been found to increase the computational complexity of local polynomial fitting and result in less satisfactorily fits compared to those obtained with compactly supported kernels. The adjustment proposed here maintains the sparseness of the smoothing vector w_s, and its effect can be made arbitrarily small by choosing δ accordingly. We let

$$\hat{T}_{lp}^*(y) = \sum_{i \in s} \left(\frac{y_i - \hat{m}_i^*}{\Pi_i} \right) + \sum_{i \in U} \hat{m}_i^* \qquad \text{.....(2.12)}$$

denote the local polynomial regression estimator that uses the adjusted sample smoother in (2.11). The development of the model-assisted local polynomial regression estimator could clearly be followed for other kinds of smoothing procedures. We focus on the local polynomial methodology because it is of considerable practical interest. In the case q = 0, the estimator relies on kernel regression, and behaves like a classical post-stratification estimator, but mixed over different possible stratum boundaries. As the bandwidth becomes large, the estimator reverts to the Hajek estimator,

$$\frac{N}{\hat{N}}\hat{T}(y), \quad where \ \hat{N}=\sum_s \frac{1}{\Pi_k}$$

In local linear regression (q=1) case, the estimator looks like a post-stratified regression estimator, and the estimator reverts to the classical regression estimator as the bandwidth becomes large.

2.3 Notation and Assumptions

In studying the design and model properties of the estimators, our basic approach is to use a Taylor linearization for the sample smoother $\hat{m}_i$. Firstly we note that we can write

$m_i = f(N^{-1}t_i, 0)$ and $\hat{m}_i = f(N^{-1}\hat{t}_i, \delta)$ for some function f, where the δ comes from the adjustment in (2.11) and variables in the population fit (2.6),

$$t_i = [t_{ig}]_{g=1}^{G}$$

$$= \left[\sum_{k \in U} \frac{1}{h_N} K\left(\frac{x_K - x_i}{h_N} \right) Z^*_{igk} \right]_{g=1}^{G}$$

$$= \left[\sum_{k \in U} Z^{**}_{igk} \right]_{g=1}^{G}$$

and $\hat{t}_i = [\hat{t}_{ig}]_{g=1}^{G}$

$$= \left[\sum_{k \in U_N} \frac{1}{h_N} K\left(\frac{x_k - x_i}{h_N} \right) Z^*_{igk} \frac{I_k}{\Pi_k} \right]_{g=1}^{G}$$

$$= \left[\sum_{k \in U} Z^{***}_{igk} \right]_{g=1}^{G}$$

for suitable Z^*_{igk}

For local polynomial regression of degree q, $G = 3q + 2$. If we let $G_1 = 2q + 1$, we can write the Z^*_{igk} as

$$Z^*_{igk} = \begin{cases} (x_k - x_i)^{g-1} & g \le G_1. \\ (x_k - x_i)^{g-G_1-1} \quad y_k & g > G_1. \end{cases}$$

Example

The kernel regression (q = 0) and local linear regression (q =1) cases are of particular interest. In the case q=0,

$$t_{i1} = \sum_{k \in U} \frac{1}{h_N} K\left(\frac{x_K - x_i}{h_N} \right) \quad and$$

$$t_{i2} = \sum_{k \in U} \frac{1}{h_N} K\left(\frac{x_K - x_i}{h_N} \right) y_K.$$

so that (2.6) is the Nadaraya-Watson estimator, based on the entire finite population of the model regression function: $m_i = t^{-1}{}_{i1} t_{i2}$

Ignoring the δ-adjustment the corresponding sample-based estimator is then

$$\hat{m}_i = \hat{t}_{i1}^{-1} \hat{t}_{i2}$$

In the case of local linear regression (q=1),

$$t_{i1} = \sum_{k \in U} \frac{1}{h_N} K\left(\frac{x_K - x_i}{h_N}\right),$$

$$t_{i2} = \sum_{k \in U} \frac{1}{h_N} K\left(\frac{x_K - x_i}{h_N}\right)(x_k - x_i),$$

$$t_{i3} = \sum_{k \in U} \frac{1}{h_N} K\left(\frac{x_K - x_i}{h_N}\right)(x_k - x_i)^2,$$

$$t_{i4} = \sum_{k \in U} \frac{1}{h_N} K\left(\frac{x_K - x_i}{h_N}\right) y_k, \quad \textit{and}$$

$$t_{i5} = \sum_{k \in U} \frac{1}{h_N} K\left(\frac{x_K - x_i}{h_N}\right)(x_k - x_i) y_K.$$

Then

$$m_i = \frac{t_{i3} t_{i4} - t_{i2} t_{i5}}{t_{i1} t_{i3} - t_{i2}^2}$$

With the corresponding sample-based estimator

$$\hat{m}_i = \frac{\hat{t}_{i3} \hat{t}_{i4} - \hat{t}_{i2} \hat{t}_{i5}}{\hat{t}_{i1} \hat{t}_{i3} - \hat{t}_{i2}^2}$$

Using a Taylor approximation, define

$$R_{iN} = \hat{m}_i - m - \frac{1}{N}\sum_{k\in U} z_{ik}\left(\frac{I_k}{\Pi_k} - 1\right) - \left.\frac{\partial \hat{m}_i}{\partial \delta}\right|_{\hat{t}_i = t_i,\ \delta = 0} \frac{\delta}{N^2} \quad \ldots\ldots\ldots \quad (2.13)$$

where

$$z_{ik} = \sum_{g=1}^{G} \frac{\partial \hat{M}_i}{\partial (N^{-1}\hat{t}_{ig})_{\hat{t}_i = t_i,\ \delta = 0}} z^{**}_{igk}$$

To prove our theoretical results, we make the following assumptions and lemmas:

(A1): Distribution of the errors under ξ.

The errors ε_i are independent and have mean zero, variance $v(x_i)$ and compact support uniformly for all N.

(A2): For each N, the x_i's are considered fixed with respect to the super-population model ξ.

The x_i's are independent and identically distributed as

$$F(x) = \int_{-\infty}^{x} f(t)\, dt$$

Where f(.) is a density with compact support $[a_x, b_x]$ and $f(x) > 0$ for all $x \in [a_x, b_x]$.

(A3): Mean and Variance functions m, v on $[a_x, b_x]$.

The mean function m(.) is continuous and has q + 1 continuous derivatives, and the variance function v(x) is continuous and strictly positive.

(A4): Kernel K

The kernel k (.) has compact support [-1, 1], is symmetric and continuous, and satisfies

$$\int_{-1}^{1} K(u)\,du = 1$$

(A5): Sampling rate $n_N N^{-1}$ and bandwidth h_N.

As $N \to \infty$, $n_N N^{-1} \to \Pi \in [0, 1]$,

$$h_N \to 0, \text{ and } \frac{N h_N^2}{\log\log N} \to \infty$$

(A6): Inclusion probabilities Π_i and Π_{ij}

$$For\ \ all\ \ N,\ \min_{i \in U_N} \Pi_i \geq \lambda > 0,\ \min_{i,j \in U_N} \Pi_{ij} \geq \lambda^* > 0$$

$$and\ \ \lim_{N \to \infty} \sup\ n_N \max_{i,j \in U_N} \left| \Pi_{ij} - \Pi_i \Pi_j \right| < \infty$$

$$\lim_{N \to \infty} n_N^2 \max_{(i_1,i_2,i_3,i_4) \in D_4,N} \left| E_p \left\{ (I_{i_1} - \Pi_{i_1})(I_{i_2} - \Pi_{i_2})(I_{i_3} - \Pi_{i_3})(I_{i_4} - \Pi_{i_4}) \right\} \right| < \infty$$

(A7): Additional assumptions involving higher-order inclusion probabilities

where $D_{t,n}$ denotes the set of all distinct t-tuples ($i_1, i_2, \ldots, i_t$) from U_N.

Thus,

$$\lim_{N\to\infty} \underset{(i_1,i_2,i_3,i_4\in D4,N)}{\text{Max}} \left|E_p\left\{(I_{i_1}I_{i_2}-\Pi_{i_1}I_2)(I_{i_3}I_{i_4}-\Pi_{i_3}I_4)\right\}\right| = 0$$

and

$$\lim_{N\to\infty} \text{Sup } n_N \underset{(i_1,i_2,i_3)\in D_{3,N}}{\text{Max}} \left|E_p\left\{(I_{i_1}-\Pi_{i_1})^2(I_{i_2}-\Pi_{i_2})(I_{i_3}-\Pi_{i_3})\right\}\right| < \infty$$

Lemmas:

Lemma 1: Assume (A2) and (A5). Then

$$\left|\frac{F_N(x+h_N)-F_N(x-h_N)}{2h_N} - f(x)\right| \to 0$$

as $N \to \infty$, *uniformly for all* x.

Lemma 2:

Under (A1) – (A6):

(i) For $k \geq 0$,

$$\lim_{N\to\infty} Sup \frac{1}{N}\sum_{i\in U_N}\left(\frac{1}{2Nh_N}\sum_{j\in U_N} I_{\{x_i-h_N\leq x_j\leq x_i+h_N\}}\right)^k < \infty$$

(ii) There exists N* independent of x, such that $N \geq N^*$.

(iii) $N^{-1}t_{ig}$ are uniformly bounded in *i* and $N^{-1}\hat{t}_{i\,g}$ are uniformly bounded in *i* and s.

$$\text{implies } \sum_{k\in U_N} I_{\{|x-x_k|\leq h_N\}} \geq q+1$$

(iv) m_i are uniformly bounded in *i* and the $\hat{m}_i$ are uniformly bounded in *i* and s.

(v) The first, second, third and fourth order mixed partials of $\hat{m}_i$ with respect to $N^{-1}t_{ig}$ and δ, evaluated at $\hat{t}_i = t_i$ $\delta = 0$, are uniformly bounded in *i*.

(vi) R^2_{iN} are uniformly bounded in *i* and s.

Lemma 3:

Assume (A1) –(A7). For the Taylor linearization remainders of the samples local polynomial residuals in (2.10),

$$\frac{n_N}{N}\sum_{i\in U_N} E_p\left[R_{iN}^2\right]=0\left(\frac{1}{n_N h_N^2}\right)$$

Lemma 4:

Assume (A1) – (A7). Then

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i\in U_N} E_p\left(\hat{m}_i - m_i\right)^2 = 0$$

Lemma 5:

Assume (A1) – (A7), then,

$$\lim_{N\to\infty}\frac{n_N}{N^2}E_p\left[\sum_{i,j\in U_N}\left(\hat{m}_i - m_i\right)\left(\hat{m}_j - m_j\right)\left(1-\frac{I_i}{\Pi_i}\right)\left(1-\frac{I_j}{\Pi_j}\right)\right]=0$$

Lemma 6:

Under (A1) – (A5)

Lemma 7:

Assume (A1) – (A7). Then

$$\lim_{N\to\infty}\frac{n_N}{N}\sum_{i\in U_N} E\left(R_{iN}^2\right)=0$$
$$\lim_{N\to\infty} N^{-1}\sum_{i\in U_N} E\left(m_i - m(x_i)\right)^2 = 0$$

Lemma 8:

Assume (A1) – (A7) hold. Then,

$$\lim_{N\to\infty}\frac{n_N}{N^2}E\left[\sum_{i,j\in U_N}\left(\hat{m}_i-m_i\right)\left(\hat{m}_j-m_j\right)\left(1-\frac{I_i}{\Pi_i}\right)\left(1-\frac{I_j}{\Pi_j}\right)\right]=0$$

2.4 Theoretical Properties of Model Assisted Local Polynomial Regression Estimator for the Finite Population Total

2.4.1 Weighting and Calibration

From equation (2.10),

$$\begin{aligned}\hat{T}(y)&=\sum_{i\in s}\frac{y_i-\hat{m}_i}{\Pi_i}+\sum_{i\in U}\hat{m}_i\\&=\sum_{i\in s}\frac{y_i}{\Pi_i}-\sum_{i\in s}\frac{\hat{m}_i}{\Pi_i}+\sum_{i\in U}\hat{m}_i\\&=\sum_{i\in s}\frac{y_i}{\Pi_i}-\sum_{i\in U}\frac{\hat{m}_i I_i}{\Pi_i}+\sum_{i\in U}\hat{m}_i\\&=\sum_{i\in s}\frac{y_i}{\Pi_i}+\sum_{i\in U}\left(\hat{m}_i-\frac{\hat{m}_i I_i}{\Pi_i}\right)\\&=\sum_{i\in s}\frac{y_i}{\Pi_i}+\sum_{i\in U}\left(1-\frac{I_i}{\Pi_i}\right)\hat{m}_i\end{aligned}$$

where I_i is the indicator variable for unit i =1, 2, …, N in a finite population, such that

$$I_i = \begin{cases} 1, & if\ i \in s \\ 0, & otherwise \end{cases}$$

and S = ($i_1, i_2, \ldots, i_n$) is the set of labels selected by a sampling mechanism. For samples of fixed size n we have

$$\sum_{i=1}^{N} I_i = n$$

and that the $P_r\{I_i = 1\} = \Pi_i$ which is the inclusion probability for the i^{th} unit when randomization is employed. We assume that $0 < \Pi_i < 1$ for all i.

Let $j \in U$ be in the non-sample.

Then

$$\hat{T}(y) = \sum_{i \in s} \frac{y_i}{\Pi_i} + \sum_{j \in U} \left(1 - \frac{I_j}{\Pi_j}\right) \hat{m}_j$$

But $\hat{m}_j = \sum_{i \in s} w_i^T y_i,$

Therefore

$$\hat{T}(y) = \sum_{i \in s} \frac{y_i}{\Pi_i} + \sum_{j \in U} \left(1 - \frac{I_j}{\Pi_j}\right) \sum_{i \in s} w_i^T y_i$$

$$= \sum_{i \in s} \frac{y_i}{\Pi_i} + \sum_{i \in s} \sum_{j \in U} \left(1 - \frac{I_j}{\Pi_j}\right) w^T y_i$$

$$= \sum_{i \in s} \left\{ \frac{1}{\Pi_i} + \sum_{j \in U} \left(1 - \frac{I_j}{\Pi_j}\right) w_i^T e_i \right\} y_i \quad \ldots\ldots\ldots\ldots \text{(2.14).}$$

$$= \sum_{i \in s} w_i y_i \quad \text{.....(2.15)}$$

$$\text{where} \quad w_i = \left\{ \frac{1}{\Pi_i} + \sum_{j \in U} \left(1 - \frac{I_j}{\Pi_j} \right) w_i^T e_i \right\}$$

and w_i, the weighting is the inverse of unit selection probabilities on the basis of randomisation inference whereas e_i is a q +1 vector with a 1 in the i^{th} position and 0 elsewhere. In a model-based framework, probability designs are ignorable and so probability weights have no obvious role.

Thus from (2.14), $\hat{T}_{lp}(y)$, is a linear combination of the sample $y_{i's}$ where the weights are the inverse inclusion probabilities, suitably modified to reflect the information in the auxiliary variable X_i. The same reasoning applies directly to $\hat{T}^*_{lp}(y)$. Because the weights are independent of $y_{i's}$, they can be applied to the auxiliary variables 1, x_i, $x_i^2, \ldots, x_i^q$. Thus for the local linear polynomial regression estimator $\hat{T}_{lp}(y)$,

$$\sum_{i \in s} w_i X_i^{\,l} = \sum_{i \in U} X_i^{\,l} I_i \quad \text{.....(2.16)}$$

for $l = 0, 1, 2, \ldots, q$ and I_i is the indicator variable previously defined. That is, the weights are exactly calibrated to the q+1 known control totals N, $T_x, \ldots, T_x^q$. Calibration is a highly desirable property for survey weights and in fact motivates the class of estimators considered by Deville and Särndal (1992). Part of the desirability of the calibration property comes from the fact that if $m(x_i)$ is exactly a q^{th}

degree polynomial function of X_i, then $\hat{T}_{lp}(y)$ is exactly model-unbiased. In addition, the control totals are often published in official tables or otherwise widely disseminated as benchmark values, so reproducing them from the sample is reassuring to the user. While the local polynomial regression estimator $\hat{T}^*_{lp}(y)$ is no longer exactly calibrated, it is approximately so, in the sense that its weights reproduce the control totals to terms of $0(\delta N^{-1})$.

2.4.2 Asymptotic Design Unbiasedness (ADU) and Consistency

The price for using $\hat{m}_{i's}$ in place of $m_{i's}$ in the generalized difference estimator (2.7) is design bias. The estimator $\hat{T}^*_{lp}(y)$ is, however, asymptotically design unbiased and design consistent under mild conditions, as the following theorem demonstrates.

Theorem 1: Assume (A1) – (A7):

Then the local polynomial regression estimator

$$\hat{T}_{lp}(y)=\sum_{i\in U}\left\{\left(y_i-\hat{m}_i\right)\frac{I_i}{\Pi_i}+\hat{m}_i\right\} \text{ is}$$

Asymptotically design unbiased (ADU) in the sense that

$$\lim_{N\to\infty} E_p\left(\frac{\hat{T}_{lp}-T}{N}\right)=0, \textit{ with } \xi-\textit{probability } 1,$$

and is design consistent in the sense that

$$\lim_{N\to\infty} E_p\left[I_{\left\{\left|\hat{T}_{HT}-T\right|>N\eta\right\}} \right]=0$$

With ξ- probability 1 for all $\eta > 0$

Proof: By Markov's inequality, it suffices to show

$$\lim_{N\to\infty} E_p\left|\frac{\hat{T}-T}{N}\right|=0$$

We write

$$\frac{\hat{T}-T}{N}=\sum_{i\in U}\frac{y_i-m_i}{N}\left(\frac{I_i}{\Pi_i}-1\right)+\sum_{i\in U}\frac{\hat{m}_i-m_i}{N}\left(1-\frac{I_i}{\Pi_i}\right).$$

Then,

$$E_p\left|\frac{\hat{T}-T}{N}\right|\le E_p\left|\sum_{i\in U}\frac{y_i-m_i}{N}\left(\frac{I_i}{\Pi_i}-1\right)\right|+E_p\left[\left|\sum_{i\in U}\frac{\left(\hat{m}_i-m_i\right)^2}{N}\right|E_p\left[\left|\sum_{i\in U}\frac{1-\Pi_i^{-1}}{N}I_i\right|\right]\right] \quad \text{.....(2.17)}$$

Under (A1)–(A6) and using the fact that

$$\lim_{N\to\infty} Sup\frac{1}{N}\sum_{i\in U}(y_i-m_i)^2<\infty$$

by Lemma 2 (iv), the first term on the right hand side of (2.17) converges to zero as $N\to\infty$, following the argument of Theorem1in Robinson and Särndal (1983).

Under (A6),

$$E_p\left\{\sum_{i\in U}\frac{\left(1-\Pi_i^{-1}I_i\right)^2}{N}\right\}=\sum_{i\in U}\frac{\Pi_i(1-\Pi_i)}{N\Pi_i^2}\le\frac{1}{\lambda}$$

Combining this with Lemma 4, the second term on the right hand side of (2.17) converges to zero as

$N \to \infty$, and the Theorem follows.

2.4.3 Asymptotic Mean Squared Error (AMSE)

The asymptotic mean squared error of the local polynomial regression estimator is equivalent to the variance of the generalized difference estimator, given in (2.8).

Theorem 2:

Assume (A1) – (A7). Then,

$$n_N E_p\left(\frac{\hat{T}^* - T}{N}\right)^2 = \frac{n_N}{N^2}\sum_{i,j\in U_N}(y_i - m_i)(y_j - m_j)\frac{\Pi_{ij} - \Pi_i\Pi_j}{\Pi_i\Pi_j} + o(1) \quad \ldots\ldots\ldots\ldots \quad (2.18).$$

Proof

Let $a_N = n_N^{\frac{1}{2}}\sum_{i\in U}\left(\frac{y_i - m_i}{N}\right)\left(\frac{I_i}{\Pi_i} - 1\right)$

and

$$b_N = n_N^{\frac{1}{2}}\sum_{i\in U}\left(\frac{m_i - \hat{m}_i}{N}\right)\left(\frac{I_i}{\Pi_i} - 1\right)$$

then

$$E_p(a_N^2) = \frac{n_N}{N^2}\sum_{i,j\in U}(y_i - m_i)(y_j - m_j)\frac{\Pi_{ij} - \Pi_i\Pi_j}{\Pi_i\Pi_j}$$

$$\leq \left(\frac{1}{\lambda} + \frac{n_N \underset{i,j\in U: i\neq j}{Max}\left|\Pi_{ij} - \Pi_i\Pi_j\right|}{\lambda^2}\right)\sum_{i\in U}\frac{(y_i - m_i)^2}{N}$$

So that

$\lim_{N\to\infty} Sup\, E_p(a_N^{\,2}) \le \infty$, by (A6). By Lemma 5,

$E_p(b_N^{\,2}) = 0(1)$, so that

$$E_p[a_N b_N] \le \left\{E_p(a_N^{\,2})\, E_p(b_N^{\,2})\right\}^{\frac{1}{2}} = 0(1).$$

Hence,

$$n_N E_p\left(\frac{\hat{T}^* - T}{N}\right)^2 = E_p(a_N^{\,2}) + 2E_p(a_N b_N) + E_p(b_N^{\,2})$$

$$= E_p(a_N^{\,2}) + 0(1)$$

and the result is proved.

The next result shows that the asymptotic mean squared error in (2.18) can be estimated consistently under mild assumptions.

Theorem 3

Assume A1-A7. Then

$$\lim_{N\to\infty} n_N E_p\left|\hat{V}(N^{-1}\hat{T}^*(y)) - AMSE(N^{-1}\hat{T}^*(y))\right| = 0,$$

where

$$\hat{V}\left(\frac{\hat{T}^*(y)}{N}\right) = \frac{1}{N^2}\sum_{i,j\in U}(y_i - \hat{m}_i)(y_j - \hat{m}_j)\frac{\Pi_{ij} - \Pi_i\Pi_j}{\Pi_i\Pi_j}\frac{I_i I_j}{\Pi_{ij}} \quad \text{.....(2.19)}$$

And

$$AMSE\left(\frac{\hat{T}^*(y)}{N}\right)=\frac{1}{N^2}\sum_{i,j\in U}(y_i-m_i)(y_j-m_j)\frac{\Pi_{ij}-\Pi_i\Pi_j}{\Pi_i\Pi_j}$$

Therefore $\hat{V}\left(\frac{\hat{T}^*(y)}{N}\right)$ is asymptotically design unbiased and design consistent

for $AMSE\frac{\hat{T}^*(y)}{N}$

Proof:

Let $A_N = n_N E_p\left|\frac{1}{N^2}\sum_{i,j\in U_N}(y_i-m_i)(y_j-m_j)\frac{\Pi_{ij}-\Pi_i\Pi_j}{\Pi_i\Pi_j}.\frac{I_iI_j-\Pi_{ij}}{\Pi_{ij}}\right|$

Now

$$n_N{}^2 E_p\left\{\frac{1}{N^2}\sum_{i,j\in U_N}(y_i-m_i)(y_j-m_j)\left(\frac{\Pi_{ij}-\Pi_i\Pi_j}{\Pi_i\Pi_j}\right)\left(\frac{I_iI_j-\Pi_{ij}}{\Pi_{ij}}\right)\right\}^2$$

$$=n_N{}^2\sum_{i,j\in U}\left(\frac{1-\Pi_i}{\Pi_i}\right)\left(\frac{1-\Pi_K}{\Pi_K}\right)\frac{(y_j-m_i)^2(y_K-m_K)^2}{N^4}.\frac{\Pi_{iK}-\Pi_i\Pi_K}{\Pi_i\Pi_K}$$

$$+2n_N{}^2\sum_{i\in U_N}\sum_{k,l\in U_N:k\neq l}\frac{1-\Pi_i}{\Pi_i}.\frac{\Pi_{kl}-\Pi_k\Pi_l}{\Pi_k\Pi_l}\frac{(y_i-m_i)^2(y_k-m_k)(y_l-m_l)}{N^4}X$$

$$E_p\left[\frac{I_i-\Pi_i}{\Pi_i}.\frac{I_kI_l-\Pi_{kl}}{\Pi_{kl}}\right]+n_N{}^2\sum_{i,j\in U_N:i\neq j}\sum_{k,l\in U_N:k\neq l}\frac{\Pi_{ij}-\Pi_i\Pi_j}{\Pi_i\Pi_j}\frac{\Pi_{kl}-\Pi_k\Pi_l}{\Pi_k\Pi_l}X$$

$$\frac{(y_i-m_i)(y_j-m_j)(y_k-m_k)(y_l-m_l)}{N^4}E_p\left[\frac{I_iI_j-\Pi_{ij}}{\Pi_{ij}}.\frac{I_kI_l-\Pi_{kl}}{\Pi_{kl}}\right]$$

$$=a_{1N}+a_{2N}+a_{3N}.$$

But

$$a_{1N} \le n_N^{\,2} \sum_{i \in U_N} \frac{(y_i - m_i)^4}{\lambda^3 N^4} + n_N^{\,2} \sum_{i,k \in U_N : i \ne k} \frac{(y_i - m_i)^2 (y_k - m_k)^2 \left| \Pi_{ik} - \Pi_i \Pi_k \right|}{\lambda^4 N^4}$$

$$\le \left(\frac{1}{N\lambda^3} + \frac{n_N \underset{i,k \in U_N : i \ne k}{Max} \left| \Pi_{ik} - \Pi_i \Pi_k \right|}{N\lambda^4} \right) \sum_{i \in U_N} \frac{(y_i - m_i)^4}{N},$$

which goes to zero as $N \to \infty$, and

$$a_{3N} \le \frac{(n_N \underset{i,k \in U_N : i \ne k}{Max} \left| \Pi_{ik} - \Pi_i \Pi_k \right|)^2}{\lambda^4 \lambda^{*2}} \sum_{i,j \in U_N : i \ne j} \sum_{k,l \in U_N : k \ne l} X$$

$$\frac{\left| (y_i - m_i)(y_j - m_j)(y_k - m_k)(y_l - m_l) \right|}{N^4} E_p \left[\frac{I_i I_j - \Pi_{ij}}{\Pi_{ij}} \frac{I_k I_l - \Pi_{kl}}{\Pi_{kl}} \right]$$

$$\le 0(N^{-1}) + \frac{\left(n_N \underset{i,k \in U_N : i \ne k}{Max} \left| \Pi_{ik} - \Pi_i \Pi_k \right| \right)^2}{\lambda^4 \lambda^{*2}} X$$

$$\underset{(i,j,k,l) \in D_4, N}{\text{Max}} \left| E_p \left[\frac{I_i I_j - \Pi_{ij}}{\Pi_{ij}} \frac{I_k I_l - \Pi_{kl}}{\Pi_{kl}} \right] \right| \sum_{i \in U} \frac{(y_i - m_i)^4}{N},$$

which converges to zero as $N \to \infty$ by (A7).

The Cauchy-Schwartz inequality may then be applied to show that $a_{2N} \to 0$ as $N \to \infty$, and it follows that $A_N \to 0$ as $N \to \infty$.

Next, we write

$$B_N = n_N E_p \left| \frac{1}{N^2} \sum_{i,j \in U_N} \left\{ 2(y_i - m_i)(m_j - \hat{m}_j) + (m_i - \hat{m}_i)(m_j - \hat{m}_j) \right\} \frac{\Pi_{ij} - \Pi_i \Pi_j}{\Pi_i \Pi_j} \frac{I_i I_j}{\Pi_{ij}} \right|$$

$$\leq \left(\frac{2n_N \underset{i,j \in U: i \neq j}{Max} \left| \Pi_{ij} - \Pi_i \Pi_j \right|}{\lambda^2 \lambda^*} + \frac{2n_N}{\lambda^2 N} \right) X \left\{ \frac{\sum_{i \in U_N} (y_i - m_i)^2}{N} \frac{\sum_{i \in U_N} E_p[(m_i - \hat{m}_i)^2]}{N} \right\}^{\frac{1}{2}}$$

$$+ \left(\frac{n_N \underset{i,j \in U: i \neq j}{Max} \left| \Pi_{ij} - \Pi_i \Pi_j \right|}{\lambda^2 \lambda^*} + \frac{n_N}{\lambda^2 N} \right) \frac{\sum_{i \in U_N} E_p[(m_i - \hat{m}_i)^2]}{N} \rightarrow 0$$

as $N \rightarrow \infty$ using (A6) and Lemma 4. The result then follows because

$$n_N E_p \left| \hat{V}(N^{-1} \hat{T}^*(y)) - AMSE(N^{-1} \hat{T}^*(y)) \right| \leq A_N + B_N.$$

Clearly, $\hat{V}(N^{-1} \hat{T}^*(y))$ is ADU of $AMSE(N^{-1} \hat{T}^*(y))$

An alternative variance estimator could be constructed by replacing the term $\Pi_i^{-1} \Pi_j^{-1}$ in (2.18) with the product of weights $w_i w_j$, $i \in s$, $j \in s$ from (2.16). This is the analogue of the weighted residual technique [Särndal, Swenson and Wretman (1989)] for estimating the variance of the general regression estimator, which they propose to improve the conditional and small sample properties of the variance estimator.

2.4.4 Bandwidth Selection

We define the bandwidth estimator as

$$\hat{h}_{cv} = \arg\min\ cv(h_N),\ where$$

$$cv(h_N) = \frac{1}{N^2}\sum_{i,j\in s}\left(y_i - \hat{m}_i^{(-i)}\right)\left(y_j - \hat{m}_j^{(-j)}\right)\left(\frac{\Pi_{ij} - \Pi_i\Pi_j}{\Pi_i\Pi_j}\right)\cdot\frac{1}{\Pi_{ij}}$$

In bandwidth selection technique, we consider a sample-based bandwidth selection method, which aims to minimize the design MSE based on the results of Theorem 2. We can justify this criterion by following the theorem proved in Breidt and Opsomer (1999b).

Theorem 4:

Under given conditions (A1) – (A7)

$$\lim_{n\to\infty} E_p\left\{\frac{cv(h_N)}{E_p\left(\frac{\hat{T}^*(y) - T}{N}\right)^2} - 1\right\} = 0$$

It is possible to re-write cv(h) in a computationally more tractable form as

$$cv(h_N) = \frac{1}{N^2}\sum_{i,j\in s}\left[\left(\frac{y_i - \hat{m}_i}{1 - [w_{si}]_i}\right)\cdot\left(\frac{y_j - \hat{m}_j}{1 - [w_{sj}]_j}\right)\cdot\left(\frac{\Pi_{ij} - \Pi_i\Pi_j}{\Pi_i\Pi_j}\right)\cdot\frac{1}{\Pi_{i,j}}\right]$$

so that the points with the individually removed observations can be computed directly based on the smoother vector W_{si}.

In practice, the same regression weights are often used for several different sets of y_i as most surveys measure many variables. Hence, trying to find the "best" bandwidth for a specific variable is not always the best approach. The result of selection method can still be useful because they provide a sample-based measure of goodness of fit with which to evaluate alternative bandwidth choices.

2.4.5 Asymptotic Normality

The local polynomial regression estimator inherits the limiting distribution properties of the generalized difference estimator as we now demonstrate in the following Theorem.

Theorem 5:

Assume that (A1) – (A7) hold and let T*(y) and Var T*(y) be as defined in (2.7) and (2.8), respectively.

Then

$$\frac{N^{-1}[T^*(y) - T(y)]}{Var_p^{\frac{1}{2}}[N^{-1}T^*(y)]} \xrightarrow{\xi} N(0,1) \text{ as } N \to \infty \text{ implies}$$

$$\frac{N^{-1}[\hat{T}^*(y) - T(y)]}{\hat{V}^{\frac{1}{2}}[N^{-1}\hat{T}^*(y)]} \xrightarrow{\xi} N(0,1) \text{ as } N \to \infty, \text{ where } \hat{V}[N^{-1}\hat{T}^*(y)] \text{ is given in (2.19).}$$

Proof:

From the proof of Theorem 2,

$$N^{-1}[\hat{T}^*(y) - T(y)] = \sum_{i \in U} \frac{y_i - m_i}{N}\left(\frac{I_i}{\Pi_i} - 1\right) + 0_p(n_N^{-\frac{1}{2}})$$

$$= \frac{1}{N}\left[\hat{T}^*(y) - T(y)\right] + 0_p(n_N^{-\frac{1}{2}})$$

Further,

$$\frac{\hat{V}[N^{-1}\hat{T}^*(y)]}{AMSE[N^{-1}\hat{T}^*(y)]} \xrightarrow{P} 1, \text{ by Theorem3, so the result is established.}$$

Thus, establishing a central limit theorem (CLT) for the local polynomial regression estimator is equivalent to establishing a CLT for the generalized difference estimator, which in turn is essentially the same problem as establishing a CLT for the Horvitz-Thompson estimator. Additional conditions on the design beyond those of Theorem 3 are generally needed; for example, conditions which ensure that the design is well approximated by unequal probability Bernoulli sampling conditioned to the fixed sample size n_N, or by successive sampling with stable draw-to-draw selection probabilities [e.g., Sen (1988), Thompson (1997), page 62]. These conditions can be verified on a design-by design basis. In the following Corollary, we establish a central limit theorem for the pivotal statistic under simple random sampling.

Corollary 1

Assume that the design is simple random sampling without replacement, and assume that (A1) – (A7) hold.

Then

$$\frac{N^{-1}[\hat{T}^*(y)-T(y)]}{\hat{V}^{\frac{1}{2}}[N^{-1}\hat{T}^*(y)]} \xrightarrow{\xi} N(0,1), \text{ as N} \to \infty, \text{ where } \hat{V}(N^{-1}\hat{T}^*(y)) \text{ can be}$$

written as

$$\hat{V}[N^{-1}\hat{T}^*(y)] = \left(1-\frac{n_N}{N}\right)\frac{\sum_{i\in s}(y_i-\hat{m}_i)^2 - n_N^{-1}\left[\sum_{i\in s}\left(y_i-\hat{m}_i\right)\right]^2}{n_N(n_N-1)}$$

Proof:

From the assumptions and Lemma 2 (iv)

$$\lim_{N\to\infty} Sup\frac{1}{N}\sum_{i\in U_N}(y_i-m_i)^4 < \infty,$$

from which the Lyapunov condition of Thompson (1997) can be deduced.

Noting that

$$Var_p[N^{-1}T^*(y)] = \left(1-\frac{n_N}{N}\right)\frac{\sum_{i\in U_N}(y_i-m_i)^2 - N^{-1}\left[\sum_{i\in U_N}(y_i-m_i)\right]^2}{n_N(N-1)},$$

then from Theorem 3 of Thompson (1997),

$$\frac{N^{-1}[T^*(y)-T(y)]}{[Var_p N^{-1}T^*(y)]^{\frac{1}{2}}} \xrightarrow{p} N(0,1).$$

so that the result follows from Theorem 4.

2.4.6 Robustness

Robustness is usually understood to mean that inferences made from a sample are insensitive to violations of the assumptions that have been made. In principle, and ordinarily in fact, robustness is achieved in probability-sampling surveys by the use of sampling with known probabilities (i.e., the randomisation) and consistent estimators, and using a large enough sample that the central limit theorem applies, so that estimators can be regarded as approximately normally distributed.

In addition to defining robustness, we consider the behaviour of the anticipated variance,

$$Var\left\{N^{-1}[\hat{T}^*(y)-T(y)]\right\}=E\left\{N^{-1}[\hat{T}^*(y)-T(y)]\right\}^2-E^2\left\{N^{-1}[\hat{T}^*(y)-T(y)]\right\},\ldots\ldots\ldots\ldots(2.20)$$

where the expectation is taken over both design, P_N, and model, ξ. From the previous results

$$E^2\left[N^{-1}\left\{\hat{T}^*(y)-T(y)\right\}\right]=0(n_N{}^{-1}),$$

So that the model-averaged design mean squared error and the anticipated variance are asymptotically equivalent in this case.

Godambe and Joshi (1965) showed that for any estimator $\hat{T}$ (y) satisfying

$$E\left[N^{-1}\left\{\hat{T}(y)-T(y)\right\}\right]=0,$$

the following inequality holds :

$$E\left(\frac{\hat{T}(y)-T(y)}{N}\right)^2\geq\frac{1}{N^2}\sum_{i\in U}V(x_i)\left(\frac{1-\Pi_i}{\Pi_i}\right)$$

The right hand-side of the above expression is the Godambe-Joshi lower bound (GJ lower bound) which attains its minimum value when $\Pi_i \propto V^{1/2}(x_i)$. Conditions under which generalized regression estimators asymptotically attain this lower bound have been studied by Wright (1983), Tam (1988) and others. In what follows, we prove that the local polynomial regression estimator is robust in the sense that it asymptotically attains the Godambe-joshi lower bound.

Theorem 6:

Under (A1) – (A7), $\hat{T}^*(y)$ asymptotically attains the Godambe-Joshi lower bound, in the sense that

$$n_N E\left\{\frac{\hat{T}(y)-T(y)}{N}\right\}^2 = \frac{n_N}{N^2}\sum_{i\in U_N} V(x_i)\left(\frac{1-\Pi_i}{\Pi_i}\right)+0(1).$$

Proof

Write:

$$b_N = \frac{n_N^{\frac{1}{2}}}{N}\sum_{i\in U_N}(m_i-\hat{m}_i)\left(\frac{I_i}{\Pi_i}-1\right)$$

$$c_N = \frac{n_N^{\frac{1}{2}}}{N}\sum_{i\in U_N}\{y_i-m(x_i)\}\left(\frac{I_i}{\Pi_i}-1\right)$$

$$d_N = \frac{n_N^{\frac{1}{2}}}{N}\sum_{i\in U_N}\{m(x_i)-m_i\}\left\{\frac{I_i}{\Pi_i}-1\right\}$$

Then

$$n_N E\left[\frac{\hat{T}^*(y)-T(y)}{N}\right]^2 = E(b_N^2)+E(c_N^2)+E(d_N^2)+2E(b_Nc_N)+2E(b_Nd_N)+2E(c_Nd_N)$$

By Lemma 8, $E(b_N{}^2) \to 0$ as $N \to \infty$.

Next,

$$E(d_N{}^2) = \frac{n_N}{N^2} \sum_{i,j \in U_N} E\left[(m_i - m_i(x_i))(m_j - m_j(x_j))\right] \frac{\Pi_{ij} - \Pi_i \Pi_j}{\Pi_i \Pi_j}$$

$$\leq \left(\frac{n_N \underset{i,j \in U_N : i \neq j}{Max} \left| \Pi_{ij} - \Pi_i \Pi_j \right|}{\lambda^2} + \frac{1}{\lambda} \right) \sum_{i \in U_N} \frac{E(m_i - m_i(x_i))^2}{N} \to 0$$

as $N \to \infty$ by Lemma 6.

Note that

$$E(c_N{}^2) = \frac{n_N}{N^2} \sum_{i \in U_N} V(x_i) \frac{1 - \Pi_i}{\Pi_i}$$

So that

$$\lim_{N \to \infty} Sup \; E(c_N{}^2) \leq \lim_{N \to \infty} Sup \; \frac{1}{N\lambda} \sum_{i \in U_N} V(x_i) < \infty$$

by (A3). The cross-product terms go to zero as $N \to \infty$ by application of the Cauchy-Schwartz inequality, and the result is proved.

In this chapter we have examined the use and theoretical properties of the new type of model-assisted nonparametric regression estimator for the finite population total based on local polynomial smoothing. In what follows in the next chapter, we study the desirable theoretical properties including adaptation to designs of covariates, consistency, asymptotic

unbiasedness and its application to finite population total estimation under conditions applicable in model-based surveys.

CHAPTER THREE

3.0 DESIGN-ADAPTIVE NONPARAMETRIC REGRESSION

3.1 Introduction

Kernel estimators for smooth curves require modification when estimation near end-points of the support both for practical and asymptotic reasons. The construction of such boundary kernels as solutions of variational problem is a difficult task.

For estimating the finite population total, we suggest an alternative estimation procedure using the theory of local linear regression. The proposed estimator adapts robustly to both interior and boundary points.

Consider a finite population of N identifiable units $U=(U_1, U_2, \ldots, U_N)$. Suppose that to each of these units there exists two numbers (x_i, y_i) which are positively correlated and are such that $(x_i, y_i) > (0, 0)$, for all $i \in U$, where $x_{i's}$ are known for all $i \in U$ but y_i is only known if $i \in s$, s is a subset of U, chosen using a probability selection plan, P, which assigns a probability, p(s), to a given s such that $p(s) \geq 0$; $\sum_{s \in S} p(s) = 1$, and $s = U_{\{s\}}$.

Given s, we can compute a statistic $\hat{T}(y)$ based on the observed $y_{i's}$, $i \in s$ and all the prior values $x_{i's}$. Let T(y) be the finite population function (i.e. census value) of interest. The problem thus considered here is thus of estimating T(y), its bias and the error variance of such an estimator.

3.2 A Model-Based Approach

A standard method or approach to estimating T(y) assumes that the values of y can be looked upon as realization of some unknown random variable $Y = (y_1, y_2, \ldots, y_N)$ whose conditional distribution can be specified with $X = (x_1, x_2, \ldots, x_N)$ being a conditioning parameter. The distribution is generally described by a probability model ξ. If we assume that a particular model relating the variables holds, then an appropriate estimator can be based on this model. For example, under the simple regression model

$$Y_i = \alpha + \beta x_i + \sigma(x_i)e_i, \; i=1, 2, \ldots, N \quad \ldots\ldots (3.1)$$

With α, and β unknown, $\sigma(x_i)$ known and e_i identically and independently distributed with mean zero and unknown variance then the minimum variance unbiased linear estimator of T is the ratio estimator, or regression estimator.

$$\hat{T}_{lin} = \sum_{j\in s} Y_j + \sum_{i\in U-s} Y_i, \; i = 1,2,\ldots n, \; j = n+1, n+2, \ldots, N$$

$$= \sum_{j\in s} Y_j + \sum_{i\in U-s} \{E(Y_i|X_i = x_i)\}$$

$$= \sum_{j\in s} Y_j + \sum_{i\in U-s} \left(\hat{\alpha} + \hat{\beta}\, x_i\right) \quad \ldots\ldots (3.2)$$

where $\hat{\alpha}$, $\hat{\beta}$ are the appropriate weighted least squares estimates of α, β. It needs to be noted that the parameters α and β are essentially nuisance parameters since T is of interest. Many sample survey practitioners are uncomfortable with this approach due to the uncertainties in the choice of the model, that is, the robustness problem.

The alternative is to insist that the sample be selected according to a probability design p(s), and assure robustness within the sampling framework by incorporating the inclusion

probabilities into the estimator. For example we can stratify on the auxiliary, employ stratified equal probability random sampling without replacement and use the expansion estimator.

Naturally, there are many other possibilities for design and model-based estimators including design-based estimators that incorporate the model, for example, the combined regression estimator Cochran (1977), Hansen et al (1983) and Royal and Cumberland (1981) give further discussion on design-based versus model-based approaches. For this reason, a new estimator based on nonparametric regression approach is suggested. Here, we weaken the assumptions concerning the relationship between Y_i and X_i. In particular we consider the model

$$E(y_i/x_i) = m(x_i)$$

$$Var(y_i/x_i) = \sigma^2(x_i) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.3)$$

$$Cov(y_i, y_j) = 0, \quad i \neq j, \; i = 1, 2, \ldots, n, \; j = n+1, n+2, \ldots, N$$

Further we assume that the functions $m(x_i)$, $\sigma(x_i)$ are Lipschits continuous (i.e. smooth). Under this model several nonparametric procedures can be used to estimate the population total, T(y). The widely used smoothing procedures are:

(1) Smoothing splines [Wahba (1975)]

(2) K-nearest Neighbour [K-N-N]

(3) Kernel smoothers i.e

- (i) Priestly-Chao [Priestly-Chao (1972), Gasser and Müller (1979)]
- (ii) Nadaraya-Watson [Nadaraya (1964), Watson (1964)]

None of these smoothing functions is uniformly best. Kernel smoothers have been found to have optimal minimax properties [Gasser and Engel (1990)]. As such, in this study we shall use the kernel functions of the Nadaraya-Watson type to develop an estimator of the finite population total.

An alternative standpoint would be to embed nonparametric smoothing in a design-based framework, as in Kuk (1993) and Jones and Bradbury (1993).

3.2.1 A Nonparametric estimator of the Total-a Review

The idea of non-parametric regression goes back to Nadaraya (1964) and Watson (1964). A recent reference is Hardle (1991). There exist many types of nonparametric regression approaches such as kernels, spline and orthogonal series methods. In this section we consider the simple Nadaraya-Watson kernel estimator.

Consider the model

$$Y_i = m(x_i) + \sigma(x_i)e_i,\ i=1, 2, \ldots, N \quad \ldots\ldots(3.4)$$

where $m(x_i)$ is a smooth function and $e_{i's}$ are identically and independently distributed with mean zero and constant variance. In this case the population generated by this model is homoscedastic. Suppose we wish to estimate m(x). To estimate m(x), several methods have been proposed and these are kernel, spline and orthogonal series approaches. Among these are two popular kernel methods proposed by Gasser and Müller (1979), Nadaraya (1964) and Watson (1964).

With k being a kernel and h_N being a bandwidth, Table 1 summarizes the asymptotic behaviour of the Nadaraya-Watson estimator and the Gasser-Müller estimator the local linear regression smoother to be introduced in section 3.3.

Table1: Point-wise Bias and Variance of Kernel Regression Smoothers

Method	Bias	Variance
Nadaraya- Watson	$\frac{1}{2}m''(x)+\frac{m'(x)f_x^{1}(x)}{f_x(x)}\int_{-\infty}^{\infty}u^2k(u)duh_n^2$	$\frac{\sigma^2(x)}{f_x(x)nh_n}\int_{-\infty}^{\infty}k^2(u)du$
Gasser- Müller	$\frac{1}{2}m''(x)\int_{-\infty}^{\infty}u^2k(u)duh_n^2$	$\frac{3\sigma^2(x)}{2f_x(x)nh_n}\int_{-\infty}^{\infty}k^2(u)du$
Local Linear smoother	$\frac{1}{2}m''(x)\int_{-\infty}^{\infty}u^2k(u)duh_n^2$	$\frac{\sigma^2(x)}{f_x(x)nh_n}\int_{-\infty}^{\infty}k^2(u)du$

The bias of the Nadaraya-Watson estimator depends on the intrinsic part $m^{11}(x)$ interplaying with the artifact $m'(x)\frac{f_x'(x)}{f_x(x)}$ due to the local constant fit. Keeping $m^{11}(x)$ fixed, we first remark that in the highly clustered design where $|f^{1}_{x}(x)/f_x(x)|$ is large, the bias of the Nadaraya-Watson estimator is large. This implies that the estimator cannot adapt to highly clustered designs. We also note that when $|m^{1}(x)|$ is large, so is the bias of that estimator. Thus even in the situation of linear regression m(x) = a + bx with a large coefficient b, the bias of the estimator is also large. In other words, the Nadaraya-Watson estimator is not good at testing linearity.

Odhiambo and Mwalili (2000) used the Nadaraya-Watson smoother to derive a nonparametric regression estimator for the finite population total, T(y).

From the model

$$Y = m(x_i) + \sigma(x_i)\, e_i,\ i = 1, 2, \ldots, N$$

Where m (.) is a smooth function and the e_i independently distributed with mean zero and constant variance, suppose we wish to estimate m (x_i). One possibility is to average the nearby values of Y_i, where "nearby" is measured by the distance $|x_i - x|$. Let k(u) be a symmetric density function, for example the standard normal function. For a given scaling factor ("bandwidth") b,

$$define\ k_b(u) = \frac{1}{b} k\left(\frac{u}{b}\right),\ and\ weights$$

$$w_i = \frac{\frac{1}{b} k_b[(X_i - x)/b]}{\frac{1}{b}\sum_{i=1}^{n} k_b[(X_i - x)/b]}$$

The larger b is, the more equal the weights. The Nadaraya-Watson estimator of m (x) is

$$\hat{m}(x) = \sum_{j=1}^{n} w_j(x) y_j \qquad \ldots\ldots(3.4)$$

Hence $\hat{T}_{np} = \sum_{j \in s} y_j + \sum_{i \in r} \hat{m}(x_i)$ for j =1,2,...,n

where $\hat{T}_{np}$ is the estimated finite population total, j is in the sample s and i is in the non - sample r

Under reasonable conditions on m(.) and the design point x, $\hat{m}$ (x) will be consistent for m(x) as b$\rightarrow$ 0, nb$\rightarrow \infty$ when n$\rightarrow \infty$. However, as observed in 3.2.1, an estimator based on Nadaraya-Watson smoother lacks local adaptability. In the subsequent section, we make use of local polynomial regression approach to obtain an estimator for the finite population total, T(y).

3.3 Local Polynomial Regression

This is a design-adaptive regression method based on a weighted local linear regression and the estimator based on this approach repairs the drawbacks of the two popular kernel smoothers described in 3.2. Looking at the Table 1, it is clear that such a method adapts to various design densities and to both interior and boundary points. Because of these adaptations, we sometimes refer to it as a design-adaptive regression estimator.

The local linear smoother not only is superior to two popular kernel regression estimators, but is also the best among all linear smoothers including those produced by orthogonal series and the spline methods.

In local polynomial regression method, we suppose that the second derivative of m(x) exists. In a small neighbourhood of a point x, $m(y) \approx m(x) + m^{1}(x)(y-x) \equiv a+b(y-x)$.

Thus the problem of estimating m(x) is equivalent to a local linear regression problem of estimating the intercept a.

If we consider a weighted local linear regression, we find "a" and "b" to minimize,

$$\sum_{j=1}^{n}\{Y_j - a - b(x_j - x)\}^2 \frac{1}{nh_n} k\left(\frac{x - X_j}{h_n}\right) \quad \text{.....(3.5)}$$

Let â and $\hat{b}$ be the solution to the weighted least square problem (3.5). Simple calculation yields that:

$$\hat{a} = \frac{\sum_{j=1}^{n} w_j y_j}{\sum_{j=1}^{n} w_j} \quad \text{.....(3.6)}$$

$$\text{where} \quad w_j = \frac{1}{nh_n} K\left(\frac{x - X_j}{h_n}\right)\left[S_{n,2} - (x - X_j)S_{n,1}\right] \quad \text{.....(3.7)}$$

$$\text{with } S_{nl} = \frac{1}{nh_n}\sum_{j=1}^{n} K\left(\frac{x - X_j}{h_n}\right)(x - X_j)^l,\ l = 1,2$$

Thus we define the local regression smoother by

$$\hat{a} = \hat{M}_{lp}(x)$$

$$= \frac{\sum_{j=1}^{n} w_j y_j}{\sum_{j=1}^{n} w_j} \quad \text{.....(3.8)}$$

The proof of (3.7) and (3.8) is as follows

$$Let \quad Q=\sum_{j=1}^{n}\{Y_j-a-b(X_j-x)\}^2\frac{1}{nh_n}K\left(\frac{x_i-X_j}{h_n}\right) \qquad \text{.....(3.9)}$$

Differentiating (3.9) with respect to a, we have

$$\frac{\partial Q}{\partial a}=\sum_{j=1}^{n}-2\{Y_j-a-b(X_j-x)\}\frac{1}{nh_n}K\left(\frac{x-X_j}{h_n}\right)$$

For least value of a, we have

$$\sum_{j=1}^{n}\{Y_j-a-b(X_j-x)\}\frac{1}{nh_n}K\left(\frac{x-X_j}{h_n}\right)=0$$

$$\sum_{j=1}^{n}\frac{1}{nh_n}K\left(\frac{x-X_j}{h_n}\right)Y_j=a\sum_{j=1}^{n}\frac{1}{nh_n}K\left(\frac{x-X_j}{h_n}\right)+b\sum_{j=1}^{n}\frac{1}{nh_n}K\left(\frac{x-X_j}{h_n}\right)(X_j-x)$$

$$\sum_{j=1}^{n}\frac{1}{nh}K\left(\frac{x-X_j}{h_n}\right)Y_j=aS_{n,0}+bS_{n,1} \qquad \text{..... (3.10)}$$

Differentiating (3.9) with respect to b we have

$$\frac{\partial Q}{\partial b}=\sum_{j=1}^{n}-2\{Y_j-a-b(X_j-x)\}\frac{1}{nh_n}K\left(\frac{x-X_j}{h_n}\right)(X_j-x)$$

For least value of b we have:

$$\sum_{j=1}^{n}\{Y_j - a - b(X_j - x)\}\frac{1}{nh_n}K\left(\frac{x-X_j}{h_n}\right)(X_j - x)=0$$

$$\sum_{j=1}^{n}\frac{1}{nh_n}K\left(\frac{x-X_j}{h_n}\right)(X_j - x)Y_j = a\sum_{j=1}^{n}\frac{1}{nh_n}K\left(\frac{x-X_j}{h_n}\right)(X_j - x)+b\sum_{j=1}^{n}\frac{1}{nh_n}K\left(\frac{x-X_j}{h_n}\right)(X_j - x)^2$$

$$\sum_{j=1}^{n}\frac{1}{nh_n}K\left(\frac{x-X_j}{h_n}\right)(X_j - x)Y_j = aS_{n,1} + bS_{n,2} \quad \text{..... (3.11)}$$

$$S_{n,2}\sum_{j=1}^{n}Y_j\frac{1}{nh_n}K\left(\frac{x-X_j}{h_n}\right) = aS_{n,o}S_{n,2} + bS_{n,1}S_{n,2} \quad \text{..... (3.12)}$$

$$S_{n,1}\sum_{j=1}^{n}Y_j\frac{1}{nh_n}K\left(\frac{x-X_j}{h_n}\right)(X_j - x) = aS_{n,1}^2 + bS_{n,1}S_{n,2} \quad \text{..... (3.13)}$$

Equation 3.12 and 3.13 gives

$$\hat{a} = \frac{S_{n,2}\sum_{j=1}^{n}\frac{1}{nh_n}K(\frac{x-X_j}{h_n})Y_j - S_{n,1}\sum_{j=1}^{n}\frac{1}{nh_n}K(\frac{x-X_j}{h_n})(X_j - x)Y_j}{S_{n,0}S_{n,2} - S_{n,1}^2}$$

$$= \frac{1}{nh_n}\sum_{j=1}^{n}\frac{\left\{S_{n,2}Y_jK\left(\frac{x-X_j}{h_n}\right) - S_{n,1}Y_jK\left(\frac{x-X_j}{h_n}\right)(X_j - x)\right\}}{\left[S_{n,2}S_{n,0} - S_{n,1}^2\right]}$$

$$= \frac{1}{nh_n} \sum_{j=1}^{n} \frac{\{S_{n,2} - (X_j - x)S_{n,1}\} K\left(\frac{x - X_j}{h_n}\right) Y_j}{S_{n,2} S_{n,0} - S_{n,1}^2}$$

$$= \frac{1}{nh_n} \sum_{j=1}^{n} \frac{[S_{n,2} - (X_j - x)S_{n,1}] K\left(\frac{x - X_j}{h_n}\right)}{S_{n,2} S_{n,0} - S_{n,1}^2} . Y_j \quad \text{.....(3.14)}$$

Which is the analogue of (3.7).

This idea is an extension of Stone (1977) who used a kernel function

$$K(x) = \frac{1}{2} L_{[|x| \le 1]},$$

and was studied by Cleveland (1979), Lejene (1985), Müller (1987) and Tsybakov (1986).

We note that $\hat{m}(x)$ is a weighted average of the responses and is called a linear smoother in this literature. The intuition at the beginning of this section suggests that b estimates $m^1(x)$. The bandwidth h_n can be chosen either subjectively by data analysts or objectively by data. A frequently used bandwidth selection technique is the cross-validation method (Stone), which chooses h_n to minimize

$$\sum_{j=1}^{n} \left\{ Y_j - \hat{m}_{-j}(x_j) \right\}^2 \quad \text{.....(3.15)}$$

Where $m_{-j}(.)$ is the regression estimator of (3.7) without using the j^{th} observation (x_j, Y_j). An alternative method is the plug-in approach [Hall, Sheather, Jones and Marron (1991)], that offers a faster rate of convergence in the density estimation.

Let $X=x_j$ be any point in the non-sample. Then in analogy to Dorfman (1992) we suggest

$$\hat{T}_{lp}=\sum_{i\in s}Y_i+\sum_{j\in U-s}\hat{m}_{lp}(x_j) \qquad \text{.....(3.16)}$$

as an estimator of the finite population total, T(y). As with model-based estimator, generally, this estimator ignores the sampling probabilities.

3.3.1 Extension to a Case when Derivative exists up to Order P

We assume that the population is generated by a model:

$$Y_i=m(x_i)+\varepsilon_i,\ i=1, 2, \ldots, N \qquad \text{.....(3.17)}$$

Where ε_i is an independent sequence of random variables with mean zero and variance $\sigma^2(x)$. m(.) is a smooth function and $\sigma^2(.)$ is also smooth and positive.

To obtain the local polynomial regression estimator, we assume a polynomial regression model locally around x. Suppose the regression function m(.) has derivatives up to a certain order P, by a Taylor approximation we then have

$$m(z)\approx\sum_{j=0}^{P}\frac{m^{(j)}(x)}{j!}(z-x)^j \qquad \text{.....(3.18)}$$

$$\equiv\sum_{j=0}^{P}\beta_j(x)(z-x)^j,\ \text{where}\ \beta_j(x)=\frac{m^{(j)}(x)}{j!}\ \text{and for } z \text{ in a neigbourhood of } x.$$

We then use, locally, the weighted least squares method to obtain estimators for m(x), that is, with $\beta_x=(\beta_o(x), \ldots, \beta_p(x))^T$,

$$\min_{\beta_x} imize \sum_{i=1}^{n} \left\{ Y_i - \sum_{j=0}^{P} \beta_j(x)(x_i - x)^j \right\}^2 W_{n,i}(x) \quad \text{..... (3.19)}$$

We then consider a local polynomial regression model with kernel weights

$$W_{n,i}(x) = \frac{1}{h} K\left(\frac{x_i - x}{h}\right),$$

where k(.) is the kernel function and $h > 0$ is the bandwidth. The minimization problem then becomes:

$$\underset{\beta_x}{\text{minimize}} \sum_{i=1}^{n} \left\{ Y_i - \sum_{j=0}^{P} \beta_j(x)(x_i - x)^j \right\}^2 \frac{1}{h} K\left(\frac{x_i - x}{h}\right) \quad \text{..... (3.20).}$$

Let $\hat{\beta}_x = \left(\hat{\beta}_0(x), \ldots, \hat{\beta}_p(x)\right)^T$ denote the minimizer of (3.20). The estimator m(x) then becomes $\hat{m}(x; p, h) = \hat{\beta}_0(x)$. The other parameters $\hat{\beta}_j(x)$, $j = 1, 2, \ldots, p$ provide estimators for the derivatives of the regression function m(.) at x up to order p. Since the local polynomial approximation only applies locally, the estimation procedure is also local and must be redone when estimating m(.) at another point. Because of this local modelling, the degree p of the polynomial approximation should be kept small, in contrast with the global modelling, where higher order polynomial are required to control the bias (illustrations in Fan and Gijbels (1996, pg.2-5)). The solution to the minimization problem (3.20) is obtained from weighted least squares theory.

3.3.2 Linear Representation of the Local Polynomial Smoother

Let $\underline{Y} = (Y_1, Y_2, \ldots, Y_n)^T$ be the vector of Y_i's in the sample and

$$W_x = diag_{1\le i\le n}\left\{\frac{1}{h}K\left(\frac{x_i - x}{h}\right)\right\}.$$

$$Define\ \ X_x = \begin{pmatrix} 1 & x_1 - x & \dots & (x_1 - x)^P \\ \cdot & & & \\ \cdot & & & \\ 1 & x_n - x & \dots & (x_n - x)^P \end{pmatrix}$$

the n x (p+1) design matrix. With this notation, the least squares problem (3.20) can be re-written as;

$$\min_{\beta_x} imize (Y - X_x\beta_x)^T W_x (Y - X_x\beta_x) \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \quad (3.21).$$

To minimize with respect to β_x, we have to let

$$F_x(x) = (Y - X_x\beta_x)^T W_x (Y - X_x\beta_x)$$

Hence equation (3.21) becomes

$$F_x(x) = \min_{\beta_x} imize \left[Y^T W_x Y - Y^T W_x X_x\beta_x - (X_x\beta_x)^T W_x Y + (X_x\beta_x)^T W_x X_x\beta_x\right]$$

To minimize with respect to β_x we have

$$\frac{\partial F_x(.)}{\partial \beta_x} = 0$$

$$\Rightarrow \quad -2X_x^T W_x Y + 2X_x^T W_x \beta_x = 0$$

$$\Rightarrow \quad \hat{\beta}_x = \left(X_x^T W_x X_x\right)^{-1} X_x^T W_x Y \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \quad (3.22)$$

The estimator for m(x) is

$$\hat{m}(x; p, h) = \hat{\beta}_0(x)$$

$$= e_1^T \left(X_x^T W_x X_x\right)^{-1} X_x^T W_x Y, \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3.23)$$

where e_1^T is the $(p+1)$ vector $(1, 0, \dots, 0)$.

We now show that

$$\hat{m}(x;p,h) = \hat{\beta}_0(x)$$

$$= e_1^T \left(X_x^T W_x X_x\right)^{-1} X_x^T W_x Y$$

is of the form (3.21) for $l = 1$.

Since e^T_1 is of order 1 x (p+1)

X^T is of order (p+1) x n

W_x is of order n x n

X_x is of order n x (p+1)

Then $X^T{}_x W_x X_x$ is of order (p+1) by (p+1) and therefore $(X^T{}_x W_x X_x)^{-1}$ is of order (p+1) by (p+1). Thus $e_1^T (X^T{}_x W_x X_x)^{-1}$ is of order 1 x p+1.

$X^T{}_x W_x$ is of order (p+1) by n, hence $e^T_1 (X^T{}_x W_x X_x)^{-1}$ is of order 1 x n.

Let $e^T_1 (X^T{}_x W_x X_x)^{-1} = W$. Then

$$\hat{m}(x;p,h) = \hat{\beta}_0(x)$$

$$= WY$$

$$= (W_1, W_2, \ldots, W_n) \begin{pmatrix} Y_1 \\ Y_2 \\ . \\ Y_n \end{pmatrix}$$

$$= W_1 Y_1 + W_2 Y_2 + \ldots, + W_n Y_n$$

$$= \sum_{j=1}^{n} W_j Y_j$$

From this expression, it is clear that local polynomial estimators are linear smoothers of the form (3.7) for l =1. The expression also shows that the local polynomial regression estimators are linear smoothers of the form given in (3.5). The coefficients in the linear combination also depend on the degree p of the polynomial approximation. For p=0, the estimator reduces to the Nadaraya-Watson estimator, that is, the Nadaraya-Watson estimator can be seen as a local constant approximation to the regression function. For p=1, a more explicit formula for (3.23) is given by;

$$\hat{m}(x;1,h)=\frac{1}{nh}\sum_{i=1}^{n}\frac{S_2(x,h)-S_1(x,h)(x_i-x)}{S_2(x,h)S_o(x,h)-S_1^2(x,h)}K\left(\frac{x_i-x}{h}\right)Y_i \quad \text{..... (3.24).}$$

where $S_r(x,h)=(nh)^{-1}\sum_{i=1}^{n}(x_i-x)^r K\left(\frac{x_i-x}{h}\right)$

Similar to the Nadaraya-Watson estimator, the local linear estimator first uses the transformed kernel function;

$K_{x,h}(u)=\frac{1}{h}K\left(\frac{(u-x)}{h}\right)$, that is, compute $K_{x,h}(x_i)$. These weights are then rescaled by

$\frac{S_2(x,h)-S_1(x,h)(x_i-x)}{S_2(x,h)S_0(x,h)-S_1^2(x,h)}$, not by a constant factor as for the Nadaraya – Watson estim

Hence the estimator for the finite population total, T(y), based on the local linear polynomial regression estimator becomes

$$\hat{T}_{lp}=\sum_{i\in s}Y_i+\sum_{j\in r}\hat{m}_{lp}(x_j) \quad \text{.....(3.25)}$$

For i=1, 2, …, n, j=n+1, n+2, …, N and x = x_j is in the non-sample.

But $\hat{m}(x_j)=\hat{m}(x_j,1,h)$

$$=\frac{1}{nh}\sum_{i=1}^{n}\frac{S_2(x,h)-S_1(x,h)(x_i-x_j)}{S_2(x,h)S_0(x,h)-S_1(x,h)}K\left(\frac{x_i-x_j}{h}\right)Y_i \quad \text{.....(3.26)}$$

Therefore

$$\hat{T}_{lp}=\sum_{i\in s}Y_i+\sum_{j\in r}\left\{\frac{1}{nh}\sum_{i\in s}\frac{S_2(x,h)-S_1(x,h)(x_i-x_j)}{S_2(x,h)S_0(x,h)-S_1^2(x,h)}K\left(\frac{x_i-x_j}{h}\right)Y_i\right\} \quad \text{..... (3.27)}$$

$$=\sum_{i\in s}Y_i+\frac{1}{nh}\sum_{j\in r}\sum_{i\in s}\frac{S_2(x,h)-S_1(x,h)(x_i-x_j)}{S_2(x,h)S_0(x,h)-S_1^2(x,h)}K\left(\frac{x_i-x_j}{h}\right)Y \quad \text{.....(3.28)}$$

3.4 The Asymptotic properties of the Local linear Polynomial Regression Estimator for the finite population total

Having derived the local linear polynomial regression estimator for the finite population total, T(y), we now study its asymptotic properties. Consider Bivariate data that can be thought of as a random sample from a certain population. Let $(X_1,Y_1),\ldots,(X_n,Y_n)$ be a random sample from a population (X,Y) with regression function defined as

$$m(x) = E(Y_i|X_i = x_i) \quad \text{.....(3.29)}$$

We need the following conditions (Ruppert and Wand (1994)) to study the asymptotic properties of the local linear polynomial regression estimator:

(i) The regression function m(x) has a bounded and continuous second derivative.

(ii) The conditional variance $\sigma^2(x) = var(Y_i \mid X_i=x_i)$ is bounded and continuous and that $\sigma^2(.) > 0$.

(iii) The marginal density f_x of the covariate X is continuous and bounded away from zero in the interval (a_0, b_0) and that $f_x'(.)$ is also continuous at x.

(iv) The kernel function K is a bounded density function with

$$\int_{-\infty}^{\infty} xK(x)dx = 0 \text{ and } \int_{-\infty}^{\infty} K^2(x)dx < \infty$$

$$c_k = \int_{-\infty}^{\infty} u^2 K(u)du$$

$$d_k = \int_{-\infty}^{\infty} K^2(u)du$$

In the sequel we always denote

(v) $h \to 0$, $nh \to 0$, as $n \to \infty$.

(vi) All averages are bounded as n and N become large.

3.4.1 The Conditional Mean (Bias) of the Prediction error in a Finite Population Total Estimation

Suppose $X = x_j$ is any point in the non-sample, then we can estimate $m(x_j)$.

We have suggested that

$$\hat{T}_{lp} = \sum_{i \in s} Y_i + \sum_{j \in r} \hat{m}_{lp}(x_j); \qquad i = 1,2,...,n$$

$$j = n+1, n+2, ..., N$$

as an estimator of the finite population total, T(y). Hence the prediction error is readily expressed as

$$\hat{T}_{lp} - T = \sum_{j \in r} \hat{m}_{lp}(x_j) - \sum_{j \in r} y_j$$

However,

$$\hat{m}_{lp}(x_j) = \frac{1}{nh} \sum_{i=1}^{n} \frac{S_2(x,h) - S_1(x,h)(x - x)}{S_2(x,h)S_0(x,h) - S_1^2(x,h)} K\left(\frac{x_i - x_j}{h}\right) Y_i$$

Where $S_r(x,h) = \frac{1}{nh} \sum_{i=1}^{n} (x_i - x_j)^r K\left(\frac{x_i - x_j}{h}\right)$, r=1,2.

Thus

$$\hat{T}_{lp} - T = \sum_{j \in r} \left\{ \frac{1}{nh} \sum_{i=1} \frac{S_2(x,h) - S_1(x,h)(x_i - x_j)}{S_2(x,h)S_0(x,h) - S_1^2(x,h)} K\left(\frac{x_i - x_j}{h}\right) Y_i \right\} - \sum_{j \in r} Y_j$$

Therefore the conditional mean of the prediction error otherwise known as the bias is,

$$E\left\{(\hat{T}_{lp} - T)/X_u\right\} = E\left[\sum_{j \in r} \left\{ \frac{1}{nh} \sum_{i \in s} \frac{S_2(x,h) - S_1(x,h)(x_i - x_j)}{S_2(x,h)S_0(x,h) - S_1^2(x,h)} K\left(\frac{x_i - x_j}{h}\right) Y_i \right\} - \sum_{j \in r} Y_j \right]$$

$$= \sum_{j \in r} E\left\{ \frac{1}{nh} \sum_{i \in s} \frac{S_2(x,h) - S_1(x,h)(x_i - x_j)}{S_2(x,h)S_0(x,h) - S_1^2(x,h)} K\left(\frac{x_i - x_j}{h}\right) Y_i \right\} - \sum_{j \in r} E(Y_j / X_u)$$

Thus,

$$E\left\{(\hat{T}_{lp} - T)/X_u\right\} = \sum_{j \in r} E\left\{ \frac{1}{nh} \sum_{i \in s} \frac{S_2(x,h) - S_1(x,h)(x_i - x_j)}{S_2(x,h)S_0(x,h) - S_1^2(x,h)} K\left(\frac{x_i - x_j}{h}\right) Y_i \right\} - \sum_{j \in r} m(x_j)$$

$$= \sum_{j \in r} E\left\{\hat{m}_{lp}(x_j)\right\} - \sum_{j \in r} m(x_j)$$

$$= \sum_{j \in r} \left\{ E(\hat{m}_{lp}(x_j) - m(x_j)) \right\}$$

But $E\left\{\hat{m}_{lp}(x_j) - m(x_j)\right\} \approx \frac{1}{2} m''(x) h^2 \int_{-1}^{1} u^2 K(u) du + 0(h^2)$ (Ruppert and Wand (1994))

Hence;

$$E\left\{(\hat{T}_{lp}-T)/X_u\right\} \approx \sum_{j\in r}\left\{\frac{1}{2}m''(x)h^2\int_{-1}^{1}u^2K(u)du + 0(h^2)\right\}$$

$$\approx \sum_{j\in r}\left\{\frac{1}{2}m''(x)h^2c_k + 0(h^2)\right\}.$$

Therefore,

$$E\left\{\frac{(\hat{T}_{lp}-T)/X_u}{N}\right\} \approx \frac{1}{N}\sum_{j\in r}\left\{\frac{1}{2}m''(x)h^2c_k + 0(h^2)\right\}$$

$$\approx \frac{N-n}{N}\sum_{j\in r}\left\{\frac{\frac{1}{2}m''(x)\,h^2c_k + 0(h^2)}{N-n}\right\}$$

$$\approx \left(1-n/N\right)\sum_{j\in r}\left\{\frac{\frac{1}{2}m''(x)h^2c_k + 0(h^2)}{N-n}\right\}$$

$$\approx \frac{h^2}{2}\left\{\left(1-\frac{n}{N}\right)\sum_{j\in r}\left[\frac{m''(x)c_k + 0(h^4)}{N-n}\right]\right\}$$

Clearly;

$$E\left\{\frac{(\hat{T}_{lp}-T)/X_u}{N}\right\} \xrightarrow{P} 0, \text{ as } h\to 0,\ N\to\infty \text{ with all averages being bounded.}$$

Thus the local linear polynomial smoother $\hat{T}_{lp}$ is an asymptotically unbiased estimator of the finite population total, T (Y).

3.4.2 The Conditional Variance of the Prediction Error in a Finite Population Total Estimation

The prediction error has already been readily expressed as $\hat{T}_{lp}-T$. Therefore the variance of the prediction error is readily expressed as

$$\mathrm{Var}\left(\{\hat{T}_{lp}-T\}/X_u\right)=\mathrm{Var}\left\{\sum_{j\in r}\hat{m}_{lp}(x_j)-\sum_{j\in r}m(x_j)\right\}$$

$$=\mathrm{Var}\sum_{j\in r}\hat{m}_{lp}(x_j)+\mathrm{Var}\sum_{j\in r}m(x_j)$$

$$=\sum_{j\in r}\mathrm{Var}\,\hat{m}_{lp}(x_j)+\sum_{j\in r}\mathrm{Var}\,m(x_j)$$

But $\mathrm{Var}\,\hat{m}_{lp}(x_j)\approx\dfrac{\sigma^2(x_j)}{nhf_x(x)}\int_{-1}^{1}K^2(u)du+0\left(\dfrac{1}{nh}\right)$ (Ruppert and Wand (1994))

Hence

$$\mathrm{Var}\left(\{\hat{T}_{lp}-T\}/X_u\right)\approx\sum_{j\in r}\left\{\frac{\sigma^2(x_j)}{nhf_x(x)}\int_{-1}^{1}K^2(u)du+0\left(\frac{1}{nh}\right)\right\}+\sum_{j\in r}\sigma^2(x_j)$$

$$\approx\frac{1}{nh}\left\{\sum_{j\in r}\frac{\sigma^2(x_j)}{f_x(x)}\int_{-1}^{1}K^2(u)du\right\}+0\left(\frac{1}{nh}\right)+\sum_{j\in r}\sigma^2(x_j)$$

$$\approx\frac{1}{nh}\left\{\sum_{j\in r}\frac{\sigma^2(x_j)}{f_x(x)}d_k\right\}+0\left(\frac{1}{nh}\right)+\sum_{j\in r}\sigma^2(x_j).$$

But under deterministic framework for X, $f_x(x)=1$.

Then;

$$Var\left(\{\hat{T}_{lp}-T\}/X_u\right)\approx\frac{1}{nh}\left\{\sum_{j\in r}\sigma^2(x_j)\right\}d_k+0\left(\frac{N-n}{nh}\right)+\sum_{j\in r}\sigma^2(x_j)$$

But for large n and under stable conditions,

$\sum_{j\in r}\sigma^2(x_j)$ can be ignored.

Hence

$$var\left(\{\hat{T}_{lp}-T/\}X_u\right)\approx\frac{1}{nh}\left[\sum_{j\in r}\sigma(x_j^2)\right]d_k+0\left(\frac{N-n}{nh}\right)$$

3.4.3: The Mean Squared Error in a Finite population Total Estimation

Under the prediction approach, the relationship between mean square error, variance and bias of the local polynomial regression estimator for the finite population total T(y) is given by

$$MSE_m\left[\hat{T}_{lp}(\underline{Y})/s,\underline{Y}\right]=Var_m\left[\hat{T}_{lp}/s,\underline{Y}\right]+\left\{B_m\left[\hat{T}_{lp}/s,\underline{Y}\right]\right\}^2$$

Thus

$$MSE_m\left[\hat{T}_{lp}(\underline{Y})/s,\underline{Y}\right]\approx\sum_{j\in r}\left\{\frac{\sigma^2(x_j)}{nhf_x(x)}\int_{-1}^{1}K^2(u)du+0_p\left(\frac{1}{nh}\right)\right\}+$$

$$\left\{\sum_{j\in r}\left[\frac{h^2}{2}m''(x)\int_{-1}^{1}u^2K(u)du+0_p\left(h^2\right)\right]\right\}^2$$

$$\approx \sum_{j\in r}\left\{\frac{\sigma^2(x)}{nhf_x(x)}d_k + 0\left(\frac{1}{nh}\right)\right\} + \left(\sum_{j\in r}\left\{\frac{h^2}{2}m''(x)c_k + 0_p(h^2)\right\}\right)^2$$

$$\approx \sum_{j\in r}\frac{\sigma^2(x_j)}{nhf_x(x)}d_k + \left\{\sum_{j\in r}\frac{h^2}{2}m''(x)c_k\right\}^2 + 0_p\left(h^4 + \frac{1}{nh}\right)$$

$$\approx \sum_{j\in r}\frac{\sigma^2(x_j)}{nhf_x(x)}d_k + \frac{h^4}{4}\left\{\sum_{j\in r}m''(x)c_k\right\}^2 + 0_p\left(h^4 + \frac{1}{nh}\right)$$

$$\approx \sum_{j\in r}\frac{\sigma^2(x_j)}{nh}d_k + \frac{h^4}{4}\left\{\sum_{j\in r}m''(x)c_k\right\}^2 + 0_p\left(h^4 + \frac{1}{nh}\right)$$

$$\approx \frac{1}{nh}\sum_{j\in r}\sigma^2(x_j)d_k + \frac{h^4}{4}\left\{\sum_{j\in r}m''(x)c_k\right\}^2 + 0_p\left(h^4 + \frac{1}{nh}\right)$$

Since $f_x(x) = 1$ under deterministic framework for X.

But $h\to 0$, $nh\to\infty$ as $n\to\infty$. Hence MSE $(\hat{T}_{lp}) \to 0$ and $\hat{T}_{lp} \to T$. Thus $\hat{T}_{lp}$ is a consistent estimator for the finite population total, T.

So far we have proposed that the local linear smoother not only is asymptotically design unbiased(ADU) but also a consistent estimator. This study would be incomplete without an empirical study on the efficiency of our estimators. To find out which estimator is more efficient and in what circumstances we now embark on empirical study in the next chapter.

CHAPTER FOUR

4.0 AN EMPIRICAL STUDY

4.1 Introduction

In this chapter, we carry out an empirical study on some simulation experiments to compare the performance of four estimators namely:

1. Horvitz – Thompson (HT) (1952)
2. Linear Regression (REG) Cochran (1977)
3. Model based Nonparametric (KERN) Dorfman (1992)
4. Local polynomial with p=1 (LPR 1)

The first two are parametric estimators corresponding to constant and linear estimators respectively whereas the last two are nonparametric. The last two are model-based. In KERN and LPR 1, the estimated mean function from a nonparametric procedure is used to predict each non – sampled y_i. In KERN we use the Nadaraya – Watson estimator under equal probability sampling.

We take the working model to be $m(x_i) = \beta x_i$, $v(x_i) = \sigma^2$ where β is some constant. We therefore consider this to be the correct model for the first of our study populations. We consider four mean functions:

Linear: $m_1(x_i) = 1+2(x_i-0.5)$,

Quadratic: $m_2(x_i) = 1+2(x_i - 0.5)^2$,

Cycle 1: $m_3(x_i) = 2+ \sin(2\pi x_i)$,

Exponential: $m_4(x_i) = \exp(-8x_i)$,

where $x_i \in [0, 1]$. These represent a range of correct and incorrect model specifications for the various estimators considered.

For m_1, linear regression estimator is expected to be the preferred estimator, since the assumed model is correctly specified. The remaining mean functions represent various departures from the linear model. It is therefore interesting to see how much efficiency, if any, is lost by only assuming that the underlying model is smooth instead of linear. The function m_2 is quadratic. The function m_3 is sinusoidal completing one full cycle on [0, 1]. For m_4, the trend is exponential, so that an assumed linear model would be misspecified over the whole range of x_k, but would be reasonable locally.

4.2 Search for the Optimal Bandwidth

The Epanechnikov Kernel or optimal Kernel $K(u)= \begin{cases} \frac{3}{4}[1-u^2] & \text{if } u^2 \le 1 \\ 0 & \text{otherwise} \end{cases}$ is used.

Two different bandwidths were considered. These were searched within the interval

$$\frac{\sigma}{4n^{1/5}} \le h \le \frac{3\sigma}{2n^{1/5}}$$

where σ is the standard deviation of the x_k's as given in Silverman (1986). We settled for h=0.1 and h=0.25.

4.3 Design of the Study Populations

The properties of the four finite population total estimators:, $\hat{T}_{HT}, \hat{T}_{REG}, \hat{T}_{NW}$ *and* $\hat{T}_{IP}$ were studied in the seven populations; six artificial and one real (natural) population. The artificial populations were constructed in the following manner:

(i) The 300 data points of x_k's were generated as independent and identically distributed (iid) uniform U[0, 1] random variables across *k*.

(ii) The e_k's were generated as normally distributed with mean zero and variance σ^2 where σ^2 is the variance of x_k's; i.e $e_k \sim N[0, \sigma^2]$, k=1, 2, ..., 300.

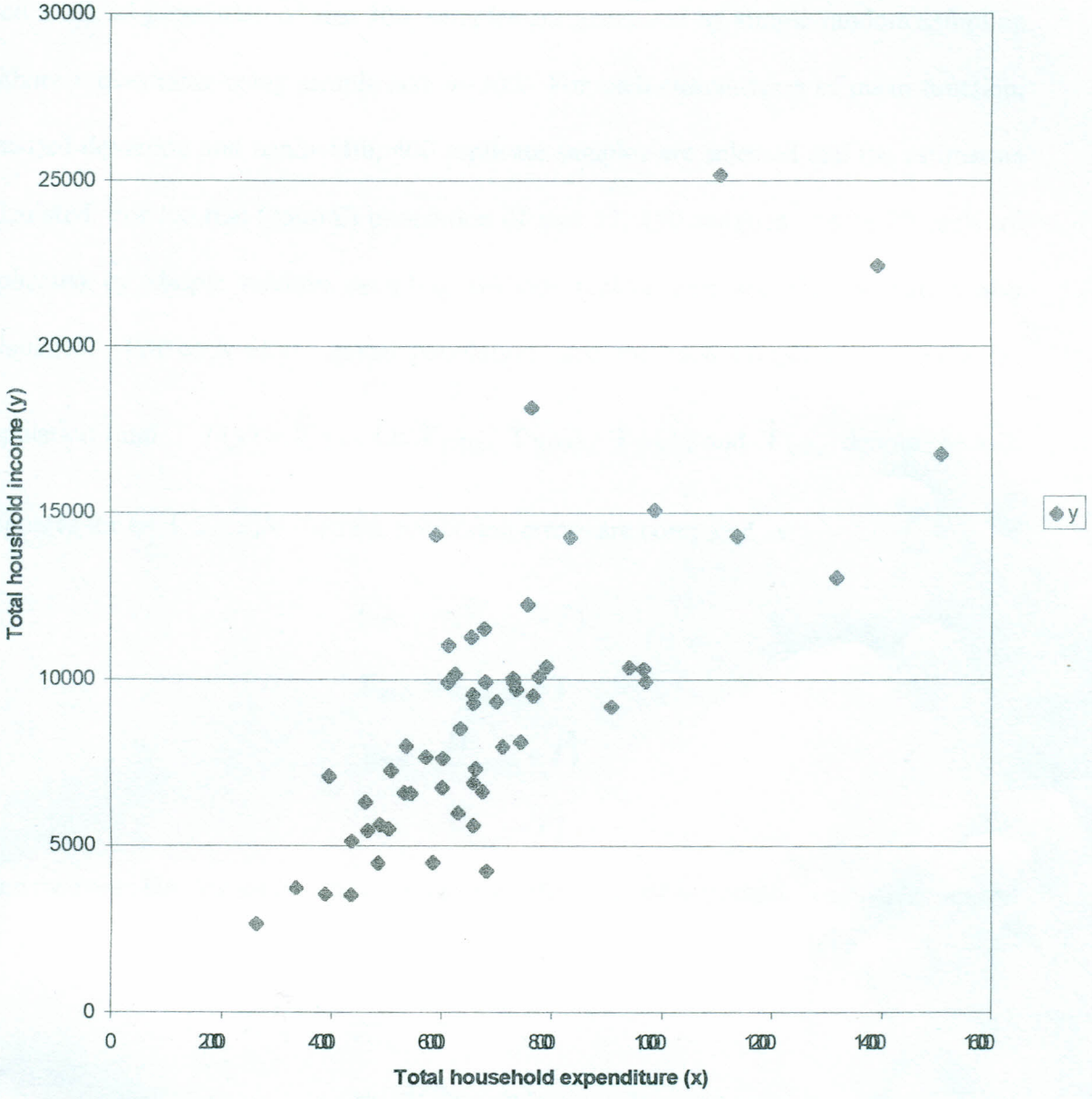
(iii) The population data points were then generated from the mean functions by adding iid $N[0, \sigma^2]$ errors in all the six cases.

Structure	Population	Model
Linear	$AP1_{(a)}$	$y_k=1+2(x_k-0.5)+\sigma_1(x_k)e_k$
	$AP1_{(b)}$	$y_k=1+2(x_k-0.5)+\sigma_2(x_k)e_k$
Exponential	$AP2_{(a)}$	$y_k=\exp(-8x_k)+\sigma_1(x_k)e_k$
	$AP2_{(b)}$	$y_k=\exp(-8x_k)+\sigma_2(x_k)e_k$
Cycle	$AP3_{(a)}$	$y_k=2+\sin 2\pi x_k+\sigma_1(x_k)e_k$
	$AP3_{(b)}$	$y_k=2+\sin 2\pi x_k+\sigma_2(x_k)e_k$

Where $\sigma_1(x_k)=0.1$ and $\sigma_2(x_k)=0.4$

The natural (real) population data points y_k's k=1, 2, 3, ..., 57 were obtained from the Central Bureau of Statistics, Basic report of 1994. Total household income by broad categories by districts was considered as the study variable y_k whereas total household expenditure by broad categories by districts was considered as the auxiliary variable x_k. The scatter diagram for the natural (real) population is given in Figure 1.

Figure 1: Scatter Diagram for the real data



4.4 Description of the Computational Procedure

We evaluate two possible values for the standard deviation of errors $\sigma = 0.1$ and $\sigma = 0.4$. For each artificial population of size 300, samples are generated by simple random sampling without replacement using sample size *n=100*. For each combination of mean function, standard deviation and bandwidth, 400 replicate samples are selected and the estimators calculated. For the real (natural) population of size 57, 150 samples of size 20 each are replicated by simple random sampling without replacement and the estimators also calculated. For each of the seven populations and for each sample we compute the population total $T(y) = \sum_{k=1}^{N} y_k$. Let $\hat{\mathbf{T}}_{\mathrm{HT(k)}}$, $\hat{\mathbf{T}}_{\mathrm{REG(k)}}$, $\hat{\mathbf{T}}_{\mathrm{NW(k)}}$, and $\hat{\mathbf{T}}_{\mathrm{lp(k)}}$ denote the k-th estimates for k= 1, 2, …,*N* then the prediction errors are computed as:

$$E_{HT} = (\hat{T}_{HT(k)} - T)$$

$$E_{REG} = \hat{T}_{REG(k)} - T$$

$$E_{NW} = \left(\hat{T}_{NW(k)} - T\right)$$

$$E_{lp} = \left(\hat{T}_{lp_{(k)}} - T\right)$$

Then the mean bias for each of the estimators for the artificial population total was computed as:

$$B(\hat{T}_{HT}) = \sum_{k=1}^{400}\left(\frac{\hat{T}_{HT(k)} - T}{400}\right)$$

$$B\left(\hat{T}_{REG}\right) = \sum_{k=1}^{400}\left(\frac{\hat{T}_{REG(k)} - T}{400}\right)$$

$$B(\hat{T}_{NW}) = \sum_{k=1}^{400}\left(\frac{\hat{T}_{NW(k)} - T}{400}\right)$$

$$B(\hat{T}_{lp}) = \sum_{k=1}^{400}\left(\frac{\hat{T}_{lp(k)} - T}{400}\right)$$

For the Natural Population total, the bias of the estimators were computed as:

$$B(\hat{T}_{HT}) = \sum_{i=1}^{150}\left(\frac{\hat{T}_{HT(k)} - T}{150}\right)$$

$$B(\hat{T}_{REG}) = \sum_{i=1}^{150}\left(\frac{\hat{T}_{REG(k)} - T}{150}\right)$$

$$B(\hat{T}_{NW}) = \sum_{k=1}^{150}\left(\frac{\hat{T}_{NW(k)} - T}{150}\right)$$

$$B(\hat{T}_{lp}) = \sum_{k=1}^{150}\left(\frac{\hat{T}_{lp(k)} - T}{150}\right)$$

The Mean Squared Error in the estimation of the artificial population total were computed as:

$$MSE\ (\hat{T}_{HT}) = \sum_{k=1}^{400}\frac{\left(\hat{T}_{HT(k)} - T\right)^2}{400}$$

$$MSE\ (\hat{T}_{REG}) = \sum_{k=1}^{400}\frac{\left(\hat{T}_{REG(k)} - T\right)^2}{400}$$

$$MSE\ (\hat{T}_{NW}) = \sum_{k=1}^{400}\frac{\left(\hat{T}_{NW(k)} - T\right)^2}{400}$$

$$MSE\ \ (\hat{T}_{lp}) = \sum_{k=1}^{400}\frac{\left(\hat{T}_{lp(k)} - T\right)^2}{400}$$

The Mean Squared Error in the estimation of the Natural Population Total were computed as:

$$MSE(\hat{T}_{HT}) = \sum_{k=1}^{150} \frac{\left(\hat{T}_{HT(k)} - T\right)^2}{150}$$

$$MSE\ (\hat{T}_{REG}) = \sum_{k=1}^{150} \frac{\left(\hat{T}_{REG(k)} - T\right)^2}{150}$$

$$MSE(\hat{T}_{NW}) = \sum_{k=1}^{150} \frac{\left(\hat{T}_{NW(k)} - T\right)^2}{150}$$

$$MSE(\hat{T}_{LP}) = \sum_{k=1}^{150} \frac{\left(\hat{T}_{lp(k)} - T\right)^2}{150}$$

The results for the biases and mean square errors for the various estimators are given in tables (2,4) and (3,5) respectively.

4.5 Results

Tabulated below is a summary of the results obtained

Table 2: Mean Bias for the seven populations for h = 0.1

	$AP1_{(a)}$	$AP1_{(b)}$	$AP2_{(a)}$	$AP2_{(b)}$	$AP3_{(a)}$	$AP3_{(b)}$	NP
$B(\hat{T}_{HT})$	-176.3468	-202.4165	63.9342	38.0299	-395.4342	-418.7050	-27.2937×10^{6}
$B(\hat{T}_{REG})$	-28.0259	-110.5815	40.0524	9.1004	-294.1365	-395.1884	-2.3194×10^{6}
$B(\hat{T}_{NW})$	-30.2774	-113.6390	11.4604	6.1209	-188.0524	-383.3592	-2.4593×10^{6}
$B(\hat{T}_{lp})$	-30.5858	-116.6171	9.4879	6.2166	-172.3042	-392.2750	-2.4935×10^{6}

Table 3: Mean Squared Error (MSE) for the seven populations for h = 0.1

	$AP1_{(a)}$	$AP1_{(b)}$	$AP2_{(a)}$	$AP2_{(b)}$	$AP3_{(a)}$	$AP3_{(b)}$	NP
$MSE(\hat{T}_{HT})$	31219.55	41086.41	4202.65	1565.03	156479.90	175431.80	279.3549×10^{12}
$MSE(\hat{T}_{REG})$	841.51	12350.43	1675.31	82.01	87088.64	156955.29	7.5385×10^{12}
$MSE(\hat{T}_{NW})$	916.40	13132.12	130.58	37.71	35730.89	147632.04	8.2086×10^{12}
$MSE(\hat{T}_{lp})$	935.12	13830.81	91.22	38.91	30033.36	153886.54	8.3761×10^{12}

Table 4: Mean bias for the seven populations for h = 0.25

	AP1 (a)	AP1 (b)	AP2 (a)	AP2 (b)	AP3 (a)	AP3 (b)	NP
B $(\hat{T}_{HT})$	-176.3468	-202.4165	63.9342	38.0299	-395.4342	-418.7050	-27.2941×10^{6}
B $(\hat{T}_{REG})$	-28.0259	-110.5815	40.0524	9.1004	-294.1365	-395.1884	-2.3194×10^{6}
B $(\hat{T}_{NW})$	-38.4742	-114.1604	24.7571	22.1765	277.7137	-399.0141	-3.3309×10^{6}
B $(\hat{T}_{lp})$	-29.5800	-113.0257	15.1267	20.6365	-178.0681	-381.3519	-2.3938×10^{6}

Table 5: Mean Squared Error (MSE) for the seven populations for h = 0.25

	AP1 (a)	AP1 (b)	AP2 (a)	AP2 (b)	AP3 (a)	AP3 (b)	NP
SE $(\hat{T}_{HT})$	31219.55	41086.41	4202.65	1565.03	156479.90	175431.80	279.3549×10^{12}
SE$(\hat{T}_{REG})$	841.51	12350.43	1675.31	82.01	87088.64	156955.29	7.5385×10^{12}
SE $(\hat{T}_{NW})$	1479.19	13251.07	658.59	537.31	78237.94	160811.49	13.2536×10^{12}
SE $(\hat{T}_{lp}$	881.66	13002.03	258.23	424.44	32063.55	146188.17	7.8890×10^{12}

Table 6 : Ratio of MSE (RMSE) of Horvitz-Thompson (HT), Linear Regression (REG) and Model- based Kernel (KERN) estimators to Local Linear Regression (LPR 1) estimator for h = 0.1

	AP1 (a)	AP1 (b)	AP2 (a)	AP2 (b)	AP3 (a)	AP3 (b)	NP
MSE ($\hat{T}_{HT}$)	33.3856	2.9706	46.0716	40.2218	5.2102	1.1400	33.3514
MSE ($\hat{T}_{REG}$)	0.8998	0.8930	18.3656	2.1077	2.8997	1.0199	0.9000
MSE ($\hat{T}_{NW}$)	0.9800	0.9495	1.4315	0.9692	1.1897	0.9594	0.9800

Table 7 : Ratio of MSE (RMSE) of Horvitz-Thompson (HT), Linear Regression (REG) and Model- based Kernel (KERN) estimators to Local Linear Regression (LPR 1) estimator for h = 0.25

	AP1 (a)	AP1 (b)	AP2 (a)	AP2 (b)	AP3 (a)	AP3 (b)	NP
RMSE ($\hat{T}_{HT}$)	35.4099	3.1600	16.2748	3.6873	4.8803	1.2004	35.4109
RMSE ($\hat{T}_{REG}$)	0.9545	0.9499	6.4877	0.1932	2.7161	1.0737	0.9556
RMSE ($\hat{T}_{NW}$)	1.6777	1.0192	2.5504	1.2659	2.4401	1.1000	1.6800

4.6 Discussion of Results

4.6.1 Bias

In all the populations, the HT was the poorest resulting in large biases as compared to the other three finite population total estimators. For all the biases considered in table 2 and 4, LPR 1 dominates HT and KERN for all the populations and essentially dominates REG for all the populations except linear where it is a strong rival.

4.6.2 Mean Squared Error (MSE)

The ratios of MSEs for the three estimators to the MSE for the local polynomial regression estimator with P=1 (LPR1) were calculated (see table 6 and 7) and it is generally found that nonparametric regression estimators perform better than the parametric estimators regardless of whether the underlying model is correctly specified or not but that effect decreases as the model variance increases.

With respect to MSE, the model based (LPR1) is found to be much better than the model-based estimator, KERN. Hence LPR1 emerges as the best among the two nonparametric estimators considered.

LPR1 estimator loses a small amount of efficiency relative to REG for the linear population but dominates REG for other populations.

4.6.3 Bandwidth

We note that LPR1 estimator at the smaller bandwidth competes with the REG estimator for the linear populations. Overall, then, the performance of LPR1 estimator is consistently good particularly at the smaller bandwidth.

We also note that as the bandwidth becomes large, the local linear polynomial regression estimator becomes less efficient. Clearly, the bandwidth has an effect on the MSE of LPR1 but Tables 2, 3, 4, 5, 6 and 7 suggest that large gains in efficiency over other estimators can be gained for a variety of bandwidth choices.

In particular, for either of the bandwidths considered here, LPR1 essentially dominates HT for all populations and essentially dominates REG for all populations except linear, where it is competitive.

The most impressive result came from the real data (Natural Population). The linear regression estimator (REG) dominates HT but became a closer competitor of the nonparametric estimators. This is due to the fact that the scatter diagram in figure 1 indicates a linear concentration of the data points. We note that as the bandwidth becomes larger, LPR1 becomes less efficient compared to REG and KERN estimators which is due to undersmoothing when the population is linear.

4.6.4 Variance

For the two values of variance considered here, we note that the MSEs for all the finite population total estimators HT, REG and KERN increase with increase in variance except in the exponential population. We also note that the increase in variance produced smaller biases for the three estimators. In particular, for either of the variances considered, smaller biases and the large MSEs for LPR 1 were observed across the populations. However, in terms of performance, LPR 1 essentially dominates HT and KERN for all the populations and dominates REG for all populations except linear where it is a strong rival.

CHAPTER FIVE

5.0 CONCLUSIONS AND AREAS FOR FURTHER RESEARCH

5.1 Introduction

In this chapter, we outline our concluding remarks and also suggest areas for further study which have emerged during the course of our study.

5.2 Conclusions

We note that the model properties of the local polynomial regression estimator are such that the estimator is a competitor to the classical survey regression estimator when the population regression function is linear but dominates the regression estimator when the regression function is not linear. However, our estimator performs well relative to other parametric and non-parametric estimators. Therefore, the linear regression estimator should be used in estimating the finite population total when the population structure is linear, but if the linearity condition is violated, then the local linear polynomial estimator is the most suitable estimator for the finite population total.

From the four approaches considered in the simulation experiments and consequently the results, it can be inferred that the local linear polynomial regresssion estimator is likely be an improvement over Horvitz-Thompson and classical survey regression estimators when the relationship between the auxiliary variable and the variable of interest is non-linear.

5.3 Areas for Further Research

In our study for the asymptotic properties of our estimator, we imposed some restrictions, which may not be realistic. One obvious and most disturbing assumption we made is that the X_i's are equispaced. Clearly this is not always the case. No wonder for the natural population the local polynomial regression estimators were highly rivalled by the classical survey regression estimators.

We assumed that the estimator we use is motivated by modelling the finite population of the Y_i's conditioned on the auxiliary variable x_i, a realization from an infinite super population ξ in which

$$y_i = m(x_i) + \epsilon_i, \; i = 1, 2, \ldots, N,$$

where ϵ_I's are i.i.d random variables with mean zero and variance $V(x_i)$. $m(x_i)$ is assumed to be smooth. $V(x_i)$ is also assumed to be smooth and positive. It would then be interesting to study the behaviour of the estimator when both $m(x_i)$ and $V(x_i)$ are assumed to be non- smooth as in the case of the data with missing information.

In our study of simulation experiments, we fixed the window smoother (Bandwidth) at 0.1 and 0.25 in estimating the finite population totals based on KERN and LPR1. It would be interesting if a variable specific bandwidth selection procedure is employed. This is an open area for further research.

In our simulation experiments, we compared the ratios of MSEs for the various estimators to the MSE for the local polynomial regression estimator with $p = 1$. It would then be

interesting to compare the performance of the estimators with local polynomial regression estimator with p = 2, 3, 4, 5, 6, ...

In this study, we have been referring to covariates (auxiliary information) as the individual characteristics of the study subject. These are factors such as age, sex, and weight among others. Although we have been mentioning covariates as a vector, the analysis was done using univariate case. A study where the covariates are considered as a vector can be undertaken.

REFERENCES

BREIDT, F. J. and OPSOMER, J. D. (2000). *Local Polynomial Regression Estimators in Survey Sampling.* Working paper, Iowa State University. Department of Statistics. **28** 1026 – 1053.

BREIDT, F.J. AND OPSOMER, J.D. (1999 b). *Implementing the local polynomial regression estimators.* Working paper, Iowa State University. Department of Statistics.

BREWER, K.R.W. (1963). *Ratio estimation in finite population: Some results deductible from the assumption of an underlying stochastic process.* Austral. J. Statistic.**5** 93-105.

CASSEL, C. M., S Ä RNDAL, C.E. and WRETMAN, J. H. (1977). *Foundations of inference in Survey Sampling.* Wiley, New York.

CENTRAL BUREAU OF STATISTICS, BASIC REPORT (1994).

CLEVELAND, W. S. (1979). *Robust Locally Weighted Regression and Smoothing Scatterplots.* J. Amer. Statist. Assoc. **74** 829 – 836.

CLEVELAND, W. S. and DELVIN, S. (1988). *Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting.* J. Amer. Statist. Assoc. **83** , 596–610.

COCHRAN, W. G. (1977). *Sampling Technique*, 3rd ed. Wiley, New Yolk.

DEVILLE, J. C. and SÄRNDAL, C. E. (1992). *Calibration Estimators in Survey Sampling*. J. Amer. Statist. Assoc. **87** 376 – 382.

DORFMAN, A. H. and HALL, P. (1993). *Estimators of the Finite Population Distribution Using Non-parametric Regression*. Ann. Statist. **21** 1452 – 1475.

DORFMAN, A.H.(1992). Nonparametric regression for estimating totals in Finite population. Proceedings of section on survey research methods. Journal of the American Statistical Association, **74** 622-625.

FAN, J. (1992). *Design-adaptive Non-parametric Regression*, J. Amer. Statist. Assoc. **87** 998 – 1004.

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and its Applications*. Chapman and Hall, London.

GASSER, T. and ENGEL (1990). *The Choice of Weights in Kernel Regression Estimation*. Biometrika **97** 377 – 381.

GASSER, T. and MÜLLER, H. G. (1979). *Kernel Estimation of Regression Functions. Smoothing Technique for Gove Estimation, eels. T. Gasser and M. Rosenblatt.* New Yolk; Springer Verlag, 23 – 68.

GODAMBE, V. P. and JOSHI, V. M. (1965). *Admissibility and Bayes Estimation in Sampling Finite Populations I.* Ann. Math. Statist. **36** 1707 – 1722.

GODAMBE, V.P. AND THOMPSON, M.E. (1977). *Robust near optimal estimation in Survey practice.* Bulletin of International Statist. Inst. **47 (3)** 129-146.

HALL et al (1991). *On Optimal data-based Brandwidth Selection in Kernel Density Estimation.* Biometrika, **78** 263-270.

HALL, P. and TURLACH, B. A. (1997). *Interpolation Methods for Adapting to Sparse Design in Non-parametric Regression.* J. Amer. Statist. Assoc. **92** 466 – 472.

HANSEN, M.H. et al (1983). *An evaluation of model-dependent and probability Sampling Inferences in Sample Surveys.* J. Amer. Statist. Assoc. **78** 776-793.

HARDLE, W. (1991). *Smoothing Techniques;* Springer-Verlag.

HÄRDLE, W. (1980). *Applied Non-parametric Regression.*

HORN, S. D., HORN, R. A. and DUNCAN, D. B. (1975). *Estimating heteroscendistic Variances in Linear Models.* J. Amer. Statist. Assoc. **78** 776 – 807.

HORVITZ, D.G and THOMPSON, D. J. (1952). *A Generalisaton of Sampling without replacement from sampling from a finite Universe.* J. Amer. Statist. Assoc. **47** 663-685.

ISAKI, C. T, and FULLER, W. A. (1982). *Survey Design Under the Regression Superpopulation Model.* J. Amer. Statist. Assoc. 77 89 – 96. *J.* Amer. Statist. Assoc.**83** 242-248.

JONES, M.C. AND BRADBURY, I.S. (1993). *Kernel Smoothing for finite Populations. Statistics and Computing* **3** *45-50.*

KUK, A.Y.C. (1993). *A Kernel Method for estimating finite population distribution functions using auxiliary information.* Biometrika **80**, 385-392

LEJENE, M. (1985). *Estimation Nonparametrique par Noyaux: Regression polynomiale Mobile, Revue de statistiques Applique'es,* **33** *43-68.*

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. The Theory of Probability Application.* **9** 141 – 142.

ODHIAMBO, R.O. AND MWALILI, T.M. (2000). *Non parametric regression Method for estimating the error Variance in unistage Sampling.* East African Journal of Science **2 (2)** 107-112.

OPSOMER, J. D. and RUPPERT, D. (1997). *Fitting a Bivariate Adaptive Model by Local Polynomial Regression.* Ann. Statist. **25** 186 – 211.

PRIESTLY, M. B. and CHAO, M. T. (1972). *Nonparametric Function Fitting.* J. Roy. Statist. Soc, B **34** 385 – 392.

ROBINSON, P.M. and SÄRNDAL, C.E. (1983). *Asymptotic Properties of the generalised regression estimation in Probability Sampling.* Sankhyā . Ser. B **45** 240-248.

ROYALL, R. M. (1970). *On Finite population Sampling Under Certain Linear Regression Models.* Biometrika **57** 377 – 387.

ROYALL, R. M. and CUMBERLAND, W. G. (1985). *Conditional Convergence Properties of Finite Population Confidence Intervals.* J. Amer. Statist. Assoc., **80** 355-359.

ROYALL, R. M. and HERSON, J. (1973). *Robust Estimation in Finite Population Sampling, I.* J. Amer. Statistic. Assoc. **68** 880-889.

ROYALL, R.M. AND CUMBERLAND , W.G.(1981). *An empirical Study of the ratio estimator and estimators of its Variance.* Journal of the Amer. Statist. Assoc. **76** 66-77.

RUPPERT, D. and WAND, M. P. (1994). *Multivariate Locally Weighted Least Squares Regression.* Ann. Statist. **22** 1346 – 1370.

SÄRNDAL, C. E. (1980). *On π-inverse weighting Verses Best Linear Unbiased Weighting in Probability.* Biometrika **67** 639 – 650.

SÄRNDAL, C. E., SWENSON, B. and WRETMAN, J. (1989). *The Weighted Residual technique for Estimating the Variance of the General Regression Estimator of the Population Total.* Biometrika **76** 527 – 537.

SÄRNDAL, C. E., SWENSON, B. and WRETMAN, J. (1992). *Model Assisted Survey Sampling.* Springer, New Yolk.

SEBER, G.A.F. (1982). *The estimation of Animal Abundance and Related parameters,* 2nd ed. London: Griffin.

SEN, P.K. (1988). *Asymptotics in finite Population Sampling.* In Handbook of Statistics (P.R.Krishnaiah and C.R. Rao, eds.) **6** 291-331. North-Holland, Amsterdam.

SILVERMAN, B. (1986). *Density Estimation.* Chapman and Hall, London.

SINHA, B.K. (1991). *Wiley Series in Probability and Mathematical Statistic.* Wiley, New York.

STONE, C.J. (1977). *Consistent non parametric Regression.* The Annals of Statistics **5** 595-620.

SUNDBERG, R.(1994). *Precision estimation in Sample Survey Inference: A criterion for choice between variance estimators.* Biometrika **81** 157-172.

TAM, S.E. (1988). *Some results on Robust estimation in finite population Sampling*

THOMPSON, M.E. (1997). *Theory of Sample Surveys.* Chapman and Hall, London.

TSYBAKOR, A.B. (1986). *Robust reconstruction of functions by the local approximation method.* Problems of information transmission. **22** 133-146.

WAHBA, G. (1975). *Smoothing Noisy Data in Spline Functions.* Numerical Mathematics **24** 386 – 393.

WAND, M. P. and JONES, M. C. (1995). *Kernel Smoothing.* Chapman and Hall, London.

WATSON, G.S. (1964). *Smoothing regression Analysis,* Sankhy $\bar{a}$, A; 359-372.

WRIGHT, R. L. (1983). *Finite Population Sampling with Multivariate Auxiliary Information.* J. Amer. Statist. Assoc. **78** 879 – 884.