COMPARISON OF THE RATIO ESTIMATE TO THE LOCAL LINEAR POLYNOMIAL ESTIMATE OF FINITE POPULATION TOTALS

Muga Zablon Maua*

George Otieno Orwa**

Romanus Odhiambo Otieno***

Leo Odiwuor Odongo

Abstract

In this paper, attempt to study effects of extreme observations on two estimators of finite population total theoretically and by simulation is made. We compare the ratio estimate with the local linear polynomial estimate of finite population total given different finite populations. Both classical and the non parametric estimator based on the local linear polynomial produce good results when the auxiliary and the study variables are highly correlated. It is however noted that in the presence of outlying observations the local linear polynomial performs better with respect to design mean square error (MSE) in all the artificial populations generated.

* Bondo University College, School of Mathematics and Actuarial Science, Bondo

** Jomo Kenyatta University, Department of Statistics and Actuarial Science, Nairobi

*** Kenyatta University, Department of Mathematics, Nairobi

1. Introduction. While investigating the characteristics of an entire population using a sample, statisticians employ either parametric or non-parametric methods. Parametric methods require that prior assumptions are made with regard to the distribution of the random variable while non-parametric methods employ analysis without such prior assumptions about the distribution of the random variable. Non parametric methods deal with approaches, some which are model based and others which are model assisted.

In an attempt to make inferences about population characteristics from a sample, statisticians have to bare in mind the possible contribution of outlying observations to the results

Robustness of an estimator will signify insensitivity to small deviations due to the presence of unusual observations.

1.1 General effects of outliers. Barnett et al., (1994) suggested accommodation and transformation as methods of dealing with outliers in a set of data. They explored the use of non-parametric methods in accommodation of outlying observations and further suggested transformations such as the use of square roots or natural logarithms when data points are non-negative to pull outliers into proximity with the rest of the data. Finally they suggested that deletion of outliers may be necessary if they are found to be errors that cannot be corrected.

Cannon et. el., (1999) investigated the effects of outliers on a regression line. In their work, a high leverage point that does not conform to the linear relationship between the variables in question is influential and would considerably change p-values from significance tests. Robin (2000) discussed effects of both deterministic and random outliers. His work considered the effects of outliers on sample means and variances. He suggested the use of visual aids, dot plots, scatter plots and box plots for identification of outliers before one proceeds with the analysis of a given set of data. He further explored a non-parametric or distribution free approach to detect outliers based on computing medians. Madalena (2005) investigated effects of outliers on Mean Square Error curves and variance. Webster (2006) outlines the effects of outlying observations on regression analysis.

1.2 Outlier Robust Estimation. Cassel et. al., (1976) and Sarndal (1980) considered the generalized regression estimators which feature great robustness to model misspecification. Aspects of the ratio and local polynomial regression estimators of finite population total considered in this project. have been previously discussed by various statisticians. Cochran

(1963) constructed a modified ratio estimator corrected for bias. Barnett (1974) showed that the ratio estimator makes use of parametrically specified models and that it is applicable as an estimator in a bivariate set of data where the two population characteristics are highly correlated. Breidt et. al.,(2000) considered estimation of finite population totals in the presence of auxiliary information based on the local polynomial regression. Design-based approaches to dealing with outliers in survey estimation have been described by Kish (1965),Searls (1966) and Hidiroglou et. al.,(1981). Chambers (1982,1986) developed model-based outlier robust techniques for sample surveys. Recent work on thgis area is described in Chambers et. al.,(1993), Lee (1991,1995), Hullinger (1995), Welsh et. al.,(1998), Duchesne (1999).

2. Proposed Ratio Estimator of Finite Population Total. In the ratio estimation, an auxiliary variate x_i, correlated with y_i is obtained for each unit in the sample. The population total X of the x_i must be known. In practice, x_i is often the value of y_i at some previous time when a complete census was taken. The aim of this method is to obtain increased precision by taking advantage of the correlation between y_i and x_i. In our study we have assumed simple random sampling.

Suppose $U = (u_1, u_2, ..., u_N)$ is a finite population of size N. Let Y be the study or survey variable and Y_i be the study variable for the i^{th} individual in the population. Let X be an auxiliary variable such that X is highly correlated with Y and X_i be the auxiliary variable for the i^{th} individual in the population. Information about X can be available before hand or can be generated in the course of study. If information about X is available before hand then the choice of the sampling strategy can be influenced both at the design and estimation stages. If however the existence of the auxiliary information is discovered in the course of study, then it can be made use of by proposing an estimator. Let $\bar{Y}$ and $\bar{X}$ be the population means of Y_i and X_i respectively, then R is the ratio of Y to X in the population such that:

$$R = \frac{Y}{X} = \frac{\sum_{i=1}^{N} Y_i}{\sum_{i=1}^{N} X_i} = \frac{N\bar{Y}}{N\bar{X}} = \frac{\bar{Y}}{\bar{X}}$$

3.1.1.1

For the sample we have;

$$\hat{R} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} = \frac{y}{x} = \frac{n\overline{y}}{n\overline{x}} = \frac{\overline{y}}{\overline{x}} \qquad 3.1.1.2$$

where, $\hat{R}$ is an estimator of R, y_i and x_i are the study variable for the i[th] unit in the sample and auxiliary variable for the i[th] unit in the sample respectively.

The ratio estimator of the population mean is;

$$\hat{\overline{Y}} = \frac{\sum_{i=1}^{n} yi}{\sum_{i=1}^{n} xi}\overline{X} = \frac{\overline{y}}{\overline{x}}\overline{X} = \hat{R}\overline{X}$$

3.1.1.3

For population totals the ratio estimators are;

$$\hat{Y}_R = N\hat{\overline{Y}}_R = \hat{R}X$$

3.1.1.4

The ratio estimator is normally a biased estimator and its bias is determined by;

$$Bias\left(\hat{\overline{Y}}\right) = E\left(\hat{\overline{Y}}\right) - \overline{Y} = E\left(\hat{R}\overline{X} - R\overline{X}\right) = \overline{X}E\left(\hat{R} - R\right) = \overline{X}Bias\hat{R}$$

To obtain bias $\hat{R}$, one may proceed as follows.

Let $$\hat{R} - R = \frac{\overline{y}}{\overline{x}} - R = \frac{\overline{y} - R\overline{x}}{\overline{x}} = \frac{1}{\overline{x}}\left(\overline{y} - R\overline{x}\right) \qquad 3.1.1.5$$

Then;

$$E\left(\hat{R} - R\right) = \frac{1-f}{n\overline{X}^2}\left(RS_x^2 - S_{xy}\right) = \frac{1-f}{n\overline{X}^2}\left(RS_x^2 - \rho S_x S_y\right) = Bias\left(\hat{R}\right) \qquad 3.1.1.6$$

Where the correlation coefficient ρ between y_i and x_i in the finite population is defined by the equation

$$\rho = \frac{E\left(y_i - \bar{Y}\right)\left(x_i - \bar{X}\right)}{\sqrt{E\left(y_i - \bar{Y}\right)\left(x_i - \bar{X}\right)^2}} = \frac{\sum_{1=i}^{N}\left(y_i - \bar{Y}\right)\left(x_i - \bar{X}\right)}{\left(N-1\right)S_y S_x}$$

and $S_{yx} = \rho S_y S_x$ is the covariance between y_i and x_i.

Hence

$$Bias\left(\hat{\bar{Y}}_R\right) = \bar{X}\ Bias\left(\hat{R}\right) = \frac{1-f}{n\bar{X}}\left(RS_x^2 - \rho S_x S_y\right) \qquad 3.1.1.7$$

2.1 Variance of the Proposed Ratio Estimator. Suppose n is large and $MSE\left(\hat{R}\right) = Var\left(\hat{R}\right)$. We assume that $\bar{x}$ and $\bar{X}$ are quite close such that $\hat{R} - R = \frac{\bar{y} - R\bar{x}}{\bar{x}} = \frac{\bar{y} - R\bar{x}}{\bar{X}}$, so that the bias of $\hat{R}$ becomes quite small, Konijn (1973).

Let $d_i = y_i - Rx_i$. Then for a sample of size n, $\bar{d} = \bar{y} - R\bar{x}$ and from the theory of simple random sampling

$$\text{var}\left(\bar{d}\right) = \frac{N-n}{nN} S_d^2 = \frac{N-n}{nN}\sum_{i=1}^{N}\frac{\left(d_i - \bar{D}\right)^2}{N-1}$$

but

$$\bar{D} = E\left(\bar{d}\right) = E\left(\bar{y} - R\bar{x}\right) = \bar{Y} - R\bar{X} = 0$$

Hence,

$$\text{var}\left(\hat{R}\right) = E\left(\hat{R} - R\right)^2 = \frac{1}{\bar{X}^2}\text{var}\left(\bar{d}\right) = \frac{1-f}{n\bar{X}^2}\sum_{i=1}^{N}\frac{\left(y_i - Rx_i\right)^2}{N-1} \qquad 3.1.2.1$$

Equation (3.1.2.1) can be written as

$$\text{var}\left(\hat{R}\right) = \frac{1-f}{n\bar{X}^2}\sum_{i=1}^{N}\frac{\left(y_i - \bar{Y} - R\left(x_i - \bar{X}\right)\right)^2}{N-1}$$

$$= \frac{1-f}{n\bar{X}^2}\left(\frac{\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)^2}{N-1} + \frac{R^2\sum_{i=1}^{N}\left(x_i - \bar{X}\right)^2}{N-1} - \frac{2R\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)\left(x_i - \bar{X}\right)}{N-1} \right)$$

$$= \frac{1-f}{n\bar{X}^2}\left(S_y^2 + R^2S_x^2 - 2R\rho S_x S_Y\right)$$

Hence

$$\operatorname{var}\left(\hat{\bar{Y}}_R\right) = \frac{1-f}{n\bar{X}}\left(S_y^2 + R^2S_x^2 - 2R\rho S_x S_Y\right) \qquad 3.1.2.2$$

and variance of the estimator of the population total is

$$\operatorname{var}\left(\hat{Y}_R\right) = \operatorname{var}\left(N\hat{\bar{Y}}_R\right) = N^2 \operatorname{var}\left(\hat{\bar{Y}}_R\right) = \frac{N^2 1-f}{n\bar{X}}\left(S_y^2 + R^2S_x^2 - 2R\rho S_x S_Y\right) \qquad 3.1.2.3$$

2.2 Estimation of the variance

$\sum_{i=1}^{N}\frac{\left(y_i - Rx_i\right)^2}{N-1}$ can be estimated by; $\sum_{i=1}^{N}\frac{\left(y_i - \hat{R}x_i\right)^2}{n-1}$

Hence

$$\hat{v}ar\left(\hat{R}\right) = \frac{1-f}{n\bar{X}}\left(s_y^2 + \hat{R}s_x^2 - 2\hat{R}s_x s_y\right) \qquad 3.1.3.1$$

The estimates of $\hat{v}ar(\hat{R})$ are:

$$\hat{v}ar_1\left(\hat{R}\right) = \frac{1-f}{n\bar{X}^2}\left(s_y^2 + \hat{R}s_x^2 - 2\hat{R}s_x s_y\right), \qquad 3.1.3.2$$

when $\bar{X}$ is known and

$$\hat{v}ar_2\left(\hat{R}\right) = \frac{1-f}{n\bar{x}^2}\left(s_y^2 + \hat{R}s_x^2 - 2\hat{R}s_x s_y\right), \qquad 3.1.3.3$$

when $\bar{X}$ is unknown

3. Local Polynomial Regression Estimator. Consider a finite population $U_N = \{1,...,i,...,N\}$. Let X_i be an auxiliary variable observed for each $i \in U_N$ .Let for the current discussion such X_i 's be scalars.

Let $t_x = \sum_{i \in U_N} x_i$. A probability sample s is drawn from U_N according to a fixed-size sampling design $p_N(.)$, where $p_N(s)$ is the probability of drawing the sample s. Let n_N be the size of s. Assume $\pi_{iN} = \Pr(i \in s) = \sum_{s:i \in s} p_N(s)$ and $\pi_{ijN} = \Pr(i, j \in s) = \sum_{s:i,j \in s} p_N(s) > 0$,for all $i, j \in U_N$.

The study variable y_i is observed for each $i \in s$. We aim at estimating $t_y = \sum_{i \in U_N} y_i$. Let $\mathrm{I}_i = \begin{cases} 1, i \in s \\ 0, i \notin s \end{cases}$ and let $E_p(I_i) = \pi_i$ where $E_p(.)$ is the averaging of all possible samples from the finite population. Then an estimator $\hat{t}_y$ of t_y is said to be design-unbiased if $E_p(\hat{t}_y) = t_y$.

The Horvitz-Thompson estimator will be taken as the design –unbiased estimator of $\mathrm{t_y}$.

$$\hat{t}_y = \sum_{i \in s} \frac{y_i}{\pi_i} = \sum_{i \in U_N} \frac{y_i I_i}{\pi_i} \qquad 3.2.2.1$$

(Horvitz-Thompson, 1952).

The variance of the Horvitz-Thompson estimator under the sampling design is

$$\mathrm{var}_p(\hat{t}_y) = \sum_{i,j \in U_N} (\pi_{ij} - \pi_i \pi_j) \frac{y_i}{\pi_i} \frac{y_j}{\pi_j} \qquad 3.2.2.2$$

Note that $\hat{t}_y$ does note depend on the auxiliary information x_i

Suppose the finite population of y_i 's is modeled conditional on the auxiliary variable x_i as a realization from an infinite super population ξ, in which

$$y_i = m(x_i) + \varepsilon_i \qquad 3.2.2.3$$

where $m(x_i)$ is a smooth function of x_i 's and $v(x_i)$ is smooth and strictly positive, ε_i are independent and identically distributed random variables with mean zero and variance $v(x_i)$. Given x_i, $m(x_i) = E_\xi(y_i)$ is called the regression function while $v(x_i) = Var_\xi(y_i)$ is called the variance function. Let Kernel function K (.) be such that: $\int_{-1}^{1} K(u)\,du = 1$.Let h_n denote the band width and let $\mathbf{y}_u = (y_i)\, i \in U_N$ be the N-vector of y_i 's in the finite population. Define the $N \times (p+1)$ matrix

$$\mathbf{X}u_i = \begin{pmatrix} 1 & x_1 - x_i \ldots & (x_1 - x_i)^p \\ . & . & . \\ . & . & . \\ . & . & . \\ 1 & x_N - x_i \ldots & (x_N - x_i)^p \end{pmatrix} = \left[1 \;\; x_j - x_i \ldots \; (x_j - x_i)^p\right]_{j \in U_N},$$

and define the $N \times N$ matrix,

$$\mathbf{W}u_i = diag\left\{\frac{1}{h_N} K\left(\frac{x_j - x_i}{h_N}\right)\right\}_{j \in U_N}$$

Let $\mathbf{e}$ represent a vector with a 1 in the rth position and 0 elsewhere i.e. $\mathbf{e} = (1, 0, 0, \ldots, 0)'$ .Then the local polynomial kernel estimator of the regression function at x_i is given by;

$$\mathrm{m_i} = \mathbf{e}_1' \left(\mathbf{X}'_{ui} \mathbf{W}_{ui} \mathbf{X}_{ui}\right)^{-1} \mathbf{X}'_{ui} \mathbf{W}_{ui} \mathbf{y}_u = \mathbf{w}'_{ui} \mathbf{y}_u \qquad 3.2.2.4$$

where , $\mathbf{w}_{ui} = \mathbf{e}' \left(\mathbf{X}'_{ui} \mathbf{W}_{ui} \mathbf{X}_{ui}\right)$

This is well defined as long as $\mathbf{X}'_{\mathrm{ui}} \mathbf{W}_{\mathrm{ui}} \mathbf{X}_{\mathrm{ui}}$ is invertible.

If m_i 's are well known a design-unbiased estimator of t_y is the generalized difference estimator

$$\hat{t}_y = \sum_{i \in s} \frac{y_i - m_i}{\pi_i} + \sum_{i \in N} m_i \qquad 3.2.2.5$$

(Sarndal, Swenson, and Wretman, 1992)

The variance of this estimator is

$$\mathrm{var}\left(\hat{t}_y\right) = \sum_{i, j \in U_N} (\pi_{ij} - \pi_i \pi_j) \frac{y_i - m_i}{\pi_i} \frac{y_j - m_j}{\pi_j} \qquad 3.2.2.6$$

This variance is much smaller than the variance of the Horvitz-Thompson estimator (3.2.2.2). The deviations $(y_i - m_i) = \{(m(x_i) - m_i) + \varepsilon_i\}$ will typically have smaller variation than $\{y_i\}$ for any reasonable smoothing procedure under the model ξ. m_i the estimator of m (.) cannot be found since only y_i in a sample $s \subset U_N$ are known. m_i is replaced by a sample-based consistent estimator. Let $\mathbf{y}_s = (y_i) i \in s$ be the n_N –vector of y_i 's obtained in the sample. Define the $n_N \times (p+1)$ matrix

$$\mathbf{X}_{si} = \left[1\; x_j - x_i \;....(x_j - x_i)^p\right]$$

and the $n_N \times n_N$ matrix;

$$\mathbf{W}_{si} = diag\left\{\frac{1}{\pi_j h_N} K\left(\frac{x_j - x_i}{h_N}\right)\right\}$$

A sample design –based estimator of m_i is given by;

$$\hat{m}_i^o = \mathbf{e}_1^{'}\left(\mathbf{X}_{si}^{'}\mathbf{W}_{si}\mathbf{X}_{si}\right)^{-1}\mathbf{X}_{si}^{'}\mathbf{W}_{si}\mathbf{y}_s = \mathbf{w}_{si}^{o'}\mathbf{y}_u \qquad 3.2.2.7$$

as long as $\mathbf{X'}_{si}\mathbf{W}_{si}\mathbf{X}_{si}$ is invertible. Substituting $\hat{m}_i^o$ into (3.2.2.5) we get the local polynomial regression estimator for the population total.

$$\tilde{t}_y^o = \sum_{i\in s}\frac{y_i - \hat{m}_i^o}{\pi_i} + \sum_{i\in U_N}\hat{m}_i^o \qquad 3.2.2.8$$

Presence of the inclusion probabilities in the smoothing weights $w_{si}^{o'}$ makes our sample-based estimator $\hat{m}_i$ a design-consistent estimator of the finite population smooth m_i, which is based on some bandwidth h_n considered here fixed for any N. In principle, the estimator 3.2.2.7 can be undefined for certain $i \in U_N$, even if the population estimator 3.2.2.4 is defined everywhere. In practice, a large bandwidth is always chosen to ensure that $\mathbf{X'}_{si}\mathbf{W}_{si}\mathbf{X}_{si}$ is invertible. In theory where a fixed band width is used, we consider an adjusted sample estimator which will exist for any sample $s \in U_N$. Let the adjusted sample estimator be given by

$$\hat{m}_i = e_1^{'}\left(\mathbf{X}_{si}^{'}\mathbf{W}_{si}\mathbf{X}_{si} + \text{diag}\left\{\frac{\delta}{N^2}\right\}_{j=1}^{p+1}\right)^{-1}\mathbf{X}_{si}^{'}\mathbf{W}_{si}\mathbf{y}_s = \mathbf{w}_{si}^{'}\mathbf{y}_s \qquad 3.2.2.9$$

for some small $\delta > 0$.

Let $\tilde{t}_y = \sum_{i \in s} \frac{y_i - \hat{m}_i}{\pi_i} + \sum_{i \in U_N} \hat{m}_i$, 3.2.2.10

denote the local polynomial regression estimator that uses adjusted sample smoothers. We consider the estimator when p=0 and when p = 1 using Taylor linearization for the sample smoother $\hat{m}_i$ .Write $m_i = f\left(N^{-1}t_i, 0\right)$ and $\hat{m}_i = f\left(N^{-1}\hat{t}_i, \delta\right)$ for some function f where δ comes from equation 3.2.2.9 and vanishes in the population fit in equation 3.2.2.4.

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

and

$$\hat{t}_i = \left[\sum_{k \in U_N} \frac{1}{h_N} K\left(\frac{x_k - x_i}{h_N} \right) z_{igk}^{\dagger} \frac{I_k}{\pi_k} \right]_{g=1}^{G} = \left[\sum_{k \in U_N} z_{igk}^{*} \frac{I_k}{\pi_k} \right]$$

for some suitable $z_{igk}^{\dagger}$

For the local polynomial regression of degree p, $G = 3p + 2$. If we let $G_1 = 2p + 1$

We can write $z_{igk}^{\dagger} = (x_k - x_i)^{g-1}$ $if \ g < G_1$ and $z_{igk}^{\dagger} = (x_k - x_i)^{g-G-1} y_k$ $if \ g > G_1$

For p =0 we have the kernel regression and (3.2.2.4) is the Nadaraya-Watson estimator based on the entire population.

3.1 Assumptions for proof of theoretical results. We outline the assumptions and prove the theoretical properties

3.1.1 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.

3.1.2 For each N, the x_i are considered fixed with respect to the super population model ξ. The x_i are independent and identically distributed $F(x) = \int_{-\infty}^{x} f(t)\,dt$, where $f(.)$ is a density with compact support $[a_x, b_x]$ and $f(x) > 0$ for all $x \in [a_x, b_x]$.

3.2. Asymptotic Normality. The local polynomial regression estimator inherits the limiting distributional properties of the generalized difference estimator, as demonstrated in the following theorem.

Theorem 4.

Assume that 1-7 hold and let $\hat{t}_y$ and $Var_p(\hat{t}_y)$ be as defined in (3.2.2.5) and (3.2.2.6), respectively. Then

$$\frac{N^{-1}\left(\hat{t}_y - t_y\right)}{Var_p^{1/2}\left(N^{-1}\hat{t}_y\right)} \xrightarrow{L} N(0,1)$$

as $N \to \infty$ implies

$$\frac{N^{-1}\left(\tilde{t}_y - t_y\right)}{\hat{V}^{1/2}\left(N^{-1}\tilde{t}_y\right)} \xrightarrow{L} N(0,1)$$

as $N \to \infty$, where

$$\hat{V}\left(N^{-1}\hat{t}_y\right) = \frac{1}{N^2}\sum_{i,j\in U_N}\left(y_i - \hat{m}_i\right)\left(y_j - \hat{m}_j\right)\frac{\pi_{ij} - \pi_i\pi_j}{\pi_i\pi_j}\frac{I_i I_j}{\pi_{ij}}$$

Proof of Theorem 4:

$$N^{-1}\left(\tilde{t}_y - t_y\right) = \sum_{i\in U_N}\frac{y_i - m_i}{N}\left(\frac{I_i}{\pi_i} - 1\right) + 0_p\left(n_N^{-1/2}\right) = N^{-1}\left(\hat{t}_y - t_y\right) + 0_p\left(n_N^{-1/2}\right)$$

Further, $\hat{V}\left(N^{-1}\tilde{t}_y\right) / AMSE\left(N^{-1}\tilde{t}_y\right) \xrightarrow{p} 1$ by theorem 3, 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 is in turn essentially the same problem as problem as establishing a CLT for the Horvitz- Thompson estimator. The following corollary establishes 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 1-7 hold. Then

$$\frac{N^{-1}\left(\tilde{t}_y - t_y\right)}{\hat{V}^{1/2}\left(N^{-1}\tilde{t}_y\right)} \xrightarrow{L} N(0,1)$$

as $N \to \infty$, where $\hat{V}\left(N^{-1}\tilde{t}_y\right)$ can be written as

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

Proof of Corollary. From the assumptions and Lemma 2(iv),

$$\limsup_{N\to\infty} N^{-1}\sum_{i\in U_N}\left(y_i - \hat{m}_i\right)^4 < \infty,$$

from which the Lyapunov condition(3.25) of Thompson (1997) can be deduced.

Note that

$$Var_p\left(N^{-1}\hat{t}_y\right) = \left(1 - \frac{n_N}{N}\right)\frac{\sum_{i\in U_N}\left(y_i - m_i\right)^2 - N^{-1}\left[\sum_{i\in U_N}\left(y_i - m_i\right)\right]^2}{n_N\left(N - 1\right)}$$

From Theorem 3.2 of Thompson (1997),

$$\frac{N^{-1}\left(\hat{t}_y - t_y\right)}{\left(Var_p\left(N^{-1}\hat{t}_y\right)\right)^{1/2}} \xrightarrow{L} N(0,1),$$

so that the result follows from Theorem 4.

Theorem 5: Under the assumptions 1-7 $\tilde{t}_y$ asymptotically attains the Godambe-Joshi lower bound in the sense that

$$n_N E\left(\frac{\tilde{t}_y - t_y}{N}\right)^2 = \frac{n_N}{N^2}\sum_{i\in U_N} v\left(x_i\right)\frac{1-\pi_i}{\pi_i} + 0(1)$$

Proof of Theorem 5:

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

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

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

Then $n_N E\left(\frac{\tilde{t}_y - t_y}{N}\right)^2 = E\left[b_N^2\right] + E\left[c_N^2\right] + E\left[d_N^2\right] + 2E\left[b_N c_N\right] + 2E\left[b_N d_N\right] + 2E\left[c_N d_N\right]$

By Lemma 8, $E\left[b_N^2\right] \to 0$ as $N \to \infty$. Next,

$$E\left[d_N^2\right] = \frac{n_N}{N^2} \sum_{i,j \in U_N} E\left[\left(m_i - m(x_i)\right)\left(m_i - m(x_j)\right)\right] \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j}$$

$$\leq \left(\frac{n_N \max_{i,j \in U_N : i \neq j} \left|\pi_{ij} - \pi_i \pi_j\right|}{\lambda^2} + \frac{1}{\lambda}\right) \sum_{i \in U_N} \frac{E\left(m_i - m(x_i)\right)^2}{N}$$

$$\to 0$$

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

$$E\left[c_N^2\right] = \frac{n_N}{N^2} \sum_{i \in U_N} v(x_i) \frac{1 - \pi_i}{\pi_i}.$$

so that

$$\limsup_{N \to \infty} E\left[c_N^2\right] \leq \lim \sup_{N \to \infty} \frac{1}{N\lambda} \sum_{i \in U_N} v(x_i) < \infty \text{ by assumption 3.}$$

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

3.4. Outlier Robust Estimation. The pth order local polynomial regression is based on minimizing

$$\sum_{i=1}^{n} \left[y_i - \beta_0 - \ldots - \beta_p (x_i - x)^p\right]^2 K\left(\frac{x_i - x}{h}\right) \quad 3.2.9.1$$

Some authors have suggested the related approach of using a local version of M-estimation which attempts to achieve robustness by replacing 3.2.9.1 by

$$\sum_{i=1}^{n} \rho\left[y_i - \beta_0 - \ldots - \beta_p (x_i - x)^p\right] K\left(\frac{x_i - x}{h}\right) \quad 3.2.9.2$$

$\rho(.)$ is chosen to down weight outliers. This is accomplished by choosing $\rho(.)$ to be symmetric with a unique minimum at zero, so that its derivative $\varphi(.)$ is bounded. Minimization of 3.2.9.2 requires iterative procedure which is stopped after one or two steps, (Fan and Jiang, 1999). However since the iterations start at the least squares local polynomial estimator, the estimator is still potentially sensitive to outliers. True robustness requires an estimator that is not based on the least squares estimator.

Wang et al (1994) investigated the Least Absolute Values (LAV) version of 3.2.9.2 estimating $m(.)$ by minimizing

$$\sum_{i=1}^{n}\left|y_i - \beta_0 - \ldots - \beta_p\left(x_i - x\right)^p\right| K\left(\frac{x_i - x}{h}\right) \qquad 3.2.9.3$$

We investigate the conditional breakdown of 3.2.9.3 and its robustness

3.5. Determining the Conditional Breakdown . From definitions l and m : **Breakdown of an estimator** is the smallest fraction of outliers that can force the estimators to arbitrary values, and is thus the measure of the resistance of the estimator to unusual values. **Breakdown point of an estimator** τ is defined to be $\alpha^* = \min\left[\frac{m}{n}; bias\left(m;\tau,y,X\right) is \ \inf inite\right]$

where $bias\left(m;\tau,y,X\right)$ is the maximum bias that can be caused by replacing any m of the original data points by arbitrary values, (Donoho and Huber, 1983). Any estimator that is not at all resistant to outliers, such as one based on least squares has breakdown $\frac{1}{n}$. Since the local polynomial regression estimate $\hat{m}(.)$ is implemented by solving many local regression problems, each centered at an evaluation point x, its breakdown properties are defined on a local level as well. We restrict ourselves to kernel functions $K(.)$ that are positive on a bounded interval [-1, 1]. Conditional breakdown implies that unlike for parametric models, the breakdown point changes depending on the evaluation point of x.

4. Results

Table 1: Relative MSE of the ratio estimator to the local linear polynomial estimator (artificial population)

	Relative Mean Square Error		
	h	**n=30**	**n=50**
Linear $\sigma_1 = .1$	0.1	0.000027	0.0000189
	0.25	0.000091	0.0000287
	0.5	0.0000964	0.001303
Linear $\sigma_2 = .4$	0.1	0.00211	0.002503
	0.25	0.004523	0.008477
	0.5	0.006464	0.0104
Quadratic $\sigma_1 = .1$	0.1	20.518	98.867
	0.25	7.561	21.967
	0.5	5.1481	16.303
Quadratic $\sigma_2 = .4$	0.1	15.48	18.4534
	0.25	14.29	13.5318
	0.5	9.2867	12.3895
Exponential $\sigma_1 = .1$	0.1	1.628	3.4007
	0.25	1.490	1.8587
	0.5	1.0985	1.3533
Exponential $\sigma_2 = .4$	0.1	2.8626	4.099

	0.25	1.235	1.2419
	0.5	1.006	1.105

4.1 Discussion of results. From table of relative mean square error, it is clear that the ratio estimator performs better than the local linear polynomial estimator when the population is linear irrespective of the variance used. The local linear polynomial regression estimator turns out to be a better estimator when the population involved is either quadratic or exponential. The relative mean square errors increase as the bandwidths increase from 0.1 to 0.5 which implies robustness of the local linear polynomial regression estimator when the Quadratic Kernel is combined with a smaller bandwidth. This trend is also true as the sample size increases from 20 to 30 which indicates that as we increase the likelihood of extreme values in the sample, the non parametric estimator performs better than the parametric estimator

4.2. Conclusion. From the results obtained we conclude that the choice of an appropriate estimator of finite population totals is important. The ratio estimator will be very useful when the variables are highly correlated such that their graph is a line through the origin. The local linear polynomial regression estimator will be handy when data involved does not depict a high linear relationship.

References

Augustyns, I. (1997). Local polynomial smoothing of sparse multinomial data. Unpublished Ph.D thesis, Limburgs University.

Brewer, K.R.W.(1963). Ratio estimation in finite populations: some results deductible from the assumption of an underlying stochastic process. Australian Journal of statistics 5,93-105.

Cassel, C-M., Sarndal, C-E., and Wretman, J.H. (1977). Foundations of Inference in Survey Sampling. Wiley, New York.

Chambers, R.L. (1986), Outlier robust finite population estimation, JASA, Vol.81, No. 396, pp. 1063-1069.

Horvitz, D.G. and D.J. Thompson. (1952). A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association 47, 663-685.

Chen, J. and Qin, J.(1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. Biometrika 80, 107-116.

Cleveland, W.S. (1979). Locally weighted regression and smoothing scatter plots. Journal of the American Statistical association 74, 829-836.

Cleveland, W.S. and Devlin,S.J.(1988). Locally weighted regression: an approach to regression analysis by local fitting. Journal of the American Statistical Association, 83,596-610.

Cochran, W.G. (1977). Sampling Techniques, 3rd ed. Wiley, New York.

Cook, R.D. Holschuch, N. and Weisberg, S. (1982), A note on an alternative outlier model, J.R.S.S., B, 44, No.3, pp. 370-376.

Donoho, D.L. and Huber, P.J. (1983). The notion of breakdown point. In: Bickel, P., Doksum, K. and Hodges, J.L. pp. 157-184.

Dorfman, A.H. and Hall, P. (1993). Estimators of the finite population distribution function using nonparametric regression. Annals of statistics 21, 1452-1475.

Draper, N.R and J.A.John (1981). Influential observations and outliers in regression.Technometrics, 23, 21-26.

Fan, J., Hu,T.C and Truong, Y.K.(1994), Robust non –parametric function estimation. Scandinavian Journal of Statistics, 21,433-446.

Fan, J. and Gijbels, I.(1996). Local Polynomial Modeling and it's applications, Chapman and Hall, London.

Giloni, A. and Padberg, M. (2004). The finite sample breakdown point for local linear regression polynomial. Journal of optimization, 14, 1028-1042.

Godambe, V.P. and Joshi,V.M. (1965). Admissibility and Bayes estimation in sampling finite populations, 1.Annals of Mathematical Statistics 36, 1707-1722.

Hidiroglou, M.A. and Srinath, K.P. (1981), Some estimators of a population total from simple random samples containing large units, JASA, Vol. 78, No. 375, pp. 690-695.

Horvitz, D.G. and D.J.Thompson. (1952). A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association 47, 663-685.

John Wiley and Sons (1996). Outliers in statistical data. New York. 10058-0012.USA.

Opsomer, J.-., and Ruppert,D. (1997). Fitting a bivariate additive model by local polynomial regression. Annals of Statistics 25, 186-211.

Pregibon, D. (1981). Logistic regression diagnostics. The Annals of Statistics, Vol.9, No.4, pp. 705-724.

Royall, R. M. (1970). On finite population sampling under certain linear regression models. Biometrika 57, 377-387.

Ruppert, D., and Wand, M.P. (1994). Multivariate locally weighted least squares regression. Annals of Statistics 22, 1346-1370.

Wand, M.P. and Jones, M.C. (1995). Kernel Smoothing, Chapman and Hall, London.

Wang, F.T. and Scott, D.W. (1994), The L_1 method for robust nonparametric regression.
Journal of the A American Statistical Association.