Results on Robust Model Based Estimation in Finite Populations

Conlet Biketi Kikechi*, Richard Onyino Simwa and Ganesh Prasad Pokhariyal

School of Mathematics, College of Biological and Physical Sciences, University of Nairobi, Nairobi, Kenya

Email address:
kikechiconlet@gmail.com (C.B. Kikechi), rsimwa@gmail.com (R.O. Simwa), pokhariyal@uonbi.ac.ke (G.P. Pokhariyal)

*Corresponding author

Abstract

In this article, we present results on nonparametric regression for estimating unknown finite population totals in a model based framework. Consistent robust estimators of finite population totals are derived using the procedure of local polynomial regression and their robustness properties studied (see Kikechi et al (2017), Kikechi et al (2018) and Kikechi and Simwa (2018)). Results of the bias show that the Local Polynomial estimators dominate the Horvitz-Thompson estimator for the linear, quadratic, bump and jump populations. Further, the biases under the model based Local Polynomial approach are much lower than those under the design based Horvitz-Thompson approach in different populations. The MSE results show that the Local Linear Regression estimators are performing better than the Horvitz-Thompson and Dorfman estimators, irrespective of the model specification or misspecification. Results further indicate that the confidence intervals generated by the model based Local Polynomial procedure are much tighter than those generated by the design based Horvitz-Thompson method, regardless of whether the model is specified or misspecified. It has been observed that the model based approach outperforms the design based approach at 95% coverage rate. In terms of their efficiency, and in comparison with other estimators that exist in the literature, it is observed that the Local Polynomial Regression estimators are robust and are the most efficient estimators. Generally, the Local Polynomial Regression estimators are not only superior to the popular Kernel Regression estimators, but they are also the best among all linear smoothers including those produced by orthogonal series and spline methods. The estimators adapt well to bias problems at boundaries and in regions of high curvature and they do not require smoothness and regularity conditions required by other methods such as the boundary Kernels.

Keywords: Finite population, Local polynomial regression, Model based framework, Nonparametric regression, Robust estimators, Survey sampling.

1. Introduction

The idea of nonparametric regression is introduced by Nadaraya (1964) and Watson (1964). Several types of nonparametric regression methods such as the kernels, penalized splines and orthogonal series are in existence (see Dorfman (1992), Hardle (1989) and Zeng & Little (2003)). In many estimation problems, the sample is used to describe and analyze the target population from which it was selected by estimating population parameters and other descriptive and analytic inferences such as correlations. Some common parameters of interest for the finite population $Y = (x_1, x_2, \dots, x_N)^T$ are the finite population total, the finite population mean, the finite population variance and the finite population proportion respectively given by,

$$T = \sum_{i=1}^{N} y_i \tag{1}$$

$$\bar{T} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{2}$$

$$V(x) = \frac{1}{N}\sum_{i=1}^{N} (x - \bar{x})^2 \tag{3}$$

$$P = \frac{A}{N} \tag{4}$$

Inferences may explore properties of the process that generate the population values (see Bolfarine and Zacks (1991)). We assume that the finite population has been generated by a super population model $\xi = f(x, y, \varphi)$ and we are interested in estimating the population parameters φ, where $\varphi = \alpha + \beta x_i$. The super population model can be applied to predict the unobserved values $y_i's$ after obtaining estimates of α and β using the known auxiliary information x_i, $i = 1,2 \dots, N$(see Montanari & Ranalli (2005) and Sanchez Borrego (2009)). Using the model ξ, the nonparametric estimator of totals, T has been derived by Dorfman (1992) who has been able to prove the asymptotic unbiasedness and MSE consistency of this estimator. The estimator, however suffers from sparse sample problem, and more work needs to be done to come up with another technique that can overcome this problem. This is where the local polynomial procedure comes in (see Kikechi et al (2017), Kikechi et al (2018) and Kikechi and Simwa (2018)).

This study therefore considers a model based approach to robust finite population total estimation using the procedure local polynomial regression. It is typically of interest to estimate $m(x)$, using Taylor's expansion of the form:

$$m(x) \approx m(x_0) + (x - x_0)m'(x_0) + \frac{(x - x_0)^2}{2!} m''(x_0) + \cdots + \frac{(x - x_0)^p}{p!} m^p(x_0) \quad (5)$$

Then the estimate of $m(x)$ at any value of x is obtained by the minimization problem,

$$\min_{\beta} \sum_{i=1}^{n} \left\{Y_i - \beta_0 - \beta_1(X_i - x) - \beta_2(X_i - x)^2 - \cdots - \beta_p(X_i - x)^p\right\}^2 K_b(x - X_i) \quad (6)$$

with respect to $\beta_0, \beta_1, \ldots, \beta_p$, where β denotes the vector of coefficients $\beta = (\beta_0, \beta_1, \ldots, \beta_p)^T$. The result is therefore a weighted least squares estimator with weights $K_b(x - X_i)$.

Using the notations,

$$X = \begin{bmatrix} 1 & x - x_1 & \ldots & (x - x_1)^p \\ 1 & x - x_2 & \cdots & (x - x_2)^p \\ & \vdots & \ddots & \vdots \\ 1 & x - x_n & \cdots & (x - x_n)^p \end{bmatrix}, \qquad Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

and

$$W = \begin{bmatrix} K_b(x - X_1) & 0 & \ldots & 0 \\ 0 & K_b(x - X_2) & \cdots & 0 \\ & \vdots & \ddots & \vdots \\ 0 & 0 & 0 \quad \cdots & K_b(x - X_n) \end{bmatrix},$$

we can compute $\bar{\beta}$ which minimizes (6) by usual formula for a weighted least squares estimator,

$$\bar{\beta}(x) = (X^T W X)^{-1} X^T W Y \quad (7)$$

Then, the local polynomial estimator of the regression function $m(x)$ is,

$$\bar{m}(x) = \bar{\beta}_0(x) = e_1^T (X^T W X)^{-1} X^T W Y \quad (8)$$

where e_1 is the $n \times 1$ vector having 1 in the first entry and 0 elsewhere.

The weighted least squares principle to be explored in the local polynomial approximation procedure, opens a wealth of statistical knowledge and thus providing easy computations and generalizations (see Fan and Gijbels (1996)). The local polynomial regression is one of the most successfully applied design adaptive non parametric regression. This estimation procedure is an attractive choice due to its flexibility and asymptotic performance. Because of its simplicity, it can be tailored to work for many different distributional assumptions. It does not require smoothness and regularity conditions required by other methods such as boundary kernels. The procedure has also the advantage of adapting well to bias problems at boundaries and in regions of high curvature. Furthermore, it is easy to understand and interpret. The estimate is also linear in the response, provided the fitting criterion is least squares and model selection does not depend on the response. See Stone (1977), Fan (1992), Fan (1993) and Ruppert and Wand (1994) among others.

In this article, we combine and present results from simulation experiments carried out by Kikechi et al (2017), Kikechi et al (2018) and Kikechi and Simwa (2018).

2. The Proposed Robust Estimators

Using the procedure of local polynomial regression for $P = 0$ and $P = 1$, the superpopulation model considered for estimating the finite population total estimators is given by,

$$Y_i = m(X_i) + \sigma^2(X_i)\varepsilon_i \tag{9}$$

Specifically, the following assumptions hold for the model considered in the nonparametric regression estimation of $m(x_i)$:

$$E(Y_i/X_i = x_i) = m(x_i)$$

$$Cov(Y_i, Y_j/X_i = x_i, X_j = x_j) = \begin{cases} \sigma^2(x_i), & i = j \\ 0 \quad , & i \neq j \end{cases} \quad i = 1, 2, 3, \dots., N \quad j = 1,2,3, \dots, N\,. \tag{10}$$

The properties of the error are given by,

$$E(\varepsilon_i/X_i = x_i) = m(x_i)$$

$$Cov(\varepsilon_i, \varepsilon_j/X_i = x_i, X_j = x_j) = \begin{cases} \sigma^2(x_i), & i = j \\ 0 \quad , & i \neq j \end{cases} \quad i = 1, 2, 3, \dots., N \quad j = 1,2,3, \dots, N\,. \tag{11}$$

The functions $m(x_i)$ and $\sigma^2(x_i)$ are assumed to be smooth and strictly positive.

Consider the Taylor series expansion of $m(x_i)$ expressed as,

$$m(x_i) = m(x_j + ht) = m(x_j) + htm^{'}(x_j) + \frac{h^2t^2}{2}m^{''}(x_j) + \frac{h^3t^3}{3}m^{'''}(x_j) + \cdots$$

$$= m(x_j) + (x_i - x_j)m^{'}(x_j) + \frac{(x_i - x_j)^2}{2!}m^{''}(x_j) + \frac{(x_i - x_j)^3}{3!}m^{'''}(x_j) \dots . \qquad (12)$$

The Taylor series expansion is written in a general form expressed as,

$$y_i = \alpha + (x_i - x_j)\beta + \varepsilon_i \qquad (13)$$

where x_i lies in the interval $[x_j - h,\ x_j + h]$ and

$$\varepsilon_i = \frac{(x_i - x_j)^2}{2!}m^{''}(x_j) + \frac{(x_i - x_j)^3}{3!}m^{'''}(x_j) + \cdots$$

The constants α and β are solved using the least squares procedure by making ε_i the subject of the formulae, squaring both sides, summing over all possible sample values and applying the weights to obtain a solution to the weighted least squares problem of the form

$$\sum_{i \in S} \varepsilon_i^2 = \sum_{i \in S} \left(y_i - \alpha - \beta(x_i - x_j)\right)^2 K\left(\frac{x_i - x_j}{h}\right) \qquad (14)$$

The robust estimators for the mean regression functions and for the finite population totals as derived by Kikechi et al (2017), Kikechi et al (2018) and Kikechi and Simwa (2018) are defined as;

$$\bar{m}_0(x_j) = \sum_{i \in S} \left\{ \frac{\left(S_2(x_j; h) - S_1(x_j; h)(x_i - x_j)\right)}{S_0(x_j; h)S_2(x_j; h) - S_1(x_j; h)^2} K\left(\frac{x_i - x_j}{h}\right) y_i \right\}$$

$$= \sum_{i \in S} w_i(x_j) y_i \qquad (15)$$

where

$$w_i(x_j) = \frac{\left(S_2(x_j; h) - S_1(x_j; h)(x_i - x_j)\right)}{S_0(x_j; h)S_2(x_j; h) - S_1(x_j; h)^2} K\left(\frac{x_i - x_j}{h}\right) y_i$$

Implying that the finite population total estimator for $P = 0$ can be estimated using,

$$\bar{T}_0 = \sum_{i\in S} y_i + \sum_{j\in R} \bar{m}_0(x_j)$$

$$= \sum_{i\in S} y_i + \sum_{j\in R} \left\{ \sum_{i\in S} \left\{ \frac{\left(S_2(x_j;h) - S_1(x_j;h)(x_i - x_j)\right)}{S_0(x_j;h)S_2(x_j;h) - S_1(x_j;h)^2} K\left(\frac{x_i - x_j}{h}\right) y_i \right\} \right\} \tag{16}$$

$$\bar{M}_1(x_j) = \sum_{i\in S} \left\{ \frac{\left(S_2(x_j;h) - S_1(x_j;h)(x_i - x_j)\right)}{S_0(x_j;h)S_2(x_j;h) - S_1(x_j;h)^2} K\left(\frac{x_i - x_j}{h}\right) y_j \right\}$$

$$+(x_i - x_j) \sum_{i\in S} \left\{ \frac{\left(S_0(x_j;h) - S_1(x_j;h)(x_i - x_j)\right)}{S_0(x_j;h)S_2(x_j;h) - S_1(x_j;h)^2} K\left(\frac{x_i - x_j}{h}\right) y_j \right\}$$

$$= \sum_{i\in S} w_i(x_j)\, y_j + (x_i - x_j) \sum_{i\in s} w_i'(x_j)\, y_j \tag{17}$$

where,

$$w_i(x_j) = \frac{\left(S_2(x_j;h) - S_1(x_j;h)(x_i - x_j)\right)}{S_0(x_j;h)S_2(x_j;h) - S_1(x_j;h)^2} K\left(\frac{x_i - x_j}{h}\right) \tag{18}$$

and,

$$w_i'(x_j) = \frac{\left(S_0(x_j;h) - S_1(x_j;h)(x_i - x_j)\right)}{S_0(x_j;h)S_2(x_j;h) - S_1(x_j;h)^2} K\left(\frac{x_i - x_j}{h}\right) \tag{19}$$

Implying that the finite population total estimator for $P = 1$ can be estimated using,

$$\bar{T}_{LL} = \sum_{i\in S} y_i + \sum_{j\in R} \bar{m}_{LL}(x_j)$$

$$= \sum_{i\in S} y_i + \sum_{j\in R} \left\{ \sum_{i\in S} \left\{ \frac{\left(S_2(x_j;h) - S_1(x_j;h)(x_i - x_j)\right)}{S_0(x_j;h)S_2(x_j;h) - S_1(x_j;h)^2} K\left(\frac{x_i - x_j}{h}\right) y_i \right\} \right\}$$

$$+ \sum_{j\in R} \left\{ \left(\frac{x_i - x_j}{S_0(x_j;h)S_2(x_j;h) - S_1(x_j;h)^2} \right) \sum_{i\in S} \left\{ \left(S_0(x_j;h)(x_i - x_j) - S_1(x_j;h) \right) k\left(\frac{x_i - x_j}{h}\right) y_i \right\} \right\} \tag{20}$$

3. Properties of Model Based Robust Estimators

Considering the fixed equally spaced design model, the following assumptions made in Ruppert and Wand (1994) are used to derive the properties of the model based robust estimators of finite population totals:

(i) The x_j variables lie in the interval $(0, 1)$.

(ii) The function $m''(.)$ is bounded and continuous on $(0, 1)$.

(iii) The kernel $K(t)$ is symmetric and supported on $(-1, 1)$. Also $K(t)$ is bounded and continuous satisfying the following: $\int_{-\infty}^{\infty} K(x)\,dx = 1$, $\int_{-\infty}^{\infty} xK(x)\,dx = 0$, $\int_{-\infty}^{\infty} x^2K(x)\,dx > 0$, $\int_{-\infty}^{\infty} K^2x\,dx < \infty$, $d_k = \int_{-\infty}^{\infty} K^2(t)\,dt$

(iv) The bandwidth h is a sequence of values which depend on the sample size n and satisfying $h \to 0$ and $nh \to \infty$, as $n \to \infty$.

(v) The point x_j at which the estimation is taking place satisfies $h < x_j < 1 - h$.

The expectation, the bias, the variance, the MSE, the unbiasedness and efficiency and the asymptotic relative efficiency of the model based robust estimators have been derived by Kikechi et al (2017), Kikechi et al (2018) and Kikechi and Simwa (2018).

4. Simulation Study

4.1 Description of the data sets

Simulation experiments are carried out to evaluate the performances of the estimators. The data are generated from the superpopulation model of the form,

$$Y_i = m(X_i) + \sigma^2(X_i)\varepsilon_i \qquad i = 1,2,\dots,n \tag{21}$$

The data sets are obtained by simulation using specific models having relations $y_i = 1 + 2(x - 0.5) + \varepsilon_i$, $y_i = 1 + 2(x - 0.5)^2 + \varepsilon_i$ and $y_i = 1 + 2(x - 0.5) + \exp(-200(x - 0.5)^2 + \varepsilon_i$ and $y_i = 1 + 2(x - 0.5)I_{(x \le 0.65)} + 0.65I_{(x > 0.65)}$ for the linear, quadratic, bump and jump populations respectively. The x_i's are generated as independent and identically distributed (iid) uniform (0, 1) random variables. The errors are assumed to be independent and identically distributed (iid) random variables with mean 0 and constant variance. The comparisons of the estimators of T according to their performances is based on Horvitz Thompson, Cochran, Dorfman and the local polynomial regression estimators $\bar{T}_0$ and $\bar{T}_1$ among others.

The Epanechnicov kernel given by $\frac{3}{4\sqrt{5}}\left(1-\frac{1}{5}t^2\right)|t|<\sqrt{5}$ is used for kernel smoothing on each of the populations due to its simplicity and easy computations using well designed computer programs. In Silverman (1986), the search for optimal bandwidth is done within the interval, $\frac{\sigma}{4n^{1/5}} \le \frac{3\sigma}{2n^{1/5}}$ where σ is the standard deviation of the $x_i's$. The bandwidths are data driven and are determined by the least squares cross validation method. For each of the four artificial populations of size 200, samples are generated by simple random sampling without replacement using sample size $n = 60$. For each combination of mean function, standard deviation and bandwidth, 500 replicate samples are selected and the estimators calculated.

Table 1: Various Estimators used for comparisons in the simulation experiments

$\bar{T}_{HT}$	$Horvitz-Thompson$	$Horvitz\ and\ Thompson$ (1952)
$\bar{T}_{REG}$	$Linear\ Regression$	$Cochran$ (1977)
$\bar{T}_{DORF}$	$Dorfman$	$Dorfman$ (1992)
$\bar{T}_{LP}$	$Local\ Polynomial$	$Robust\ estimators$

Table 2. Computational formulae of population totals for different estimators

Estimator	Formulae
$Horvitz-Thompson, \bar{T}_{HT}$	$\bar{T}_{HT} = \sum_{i=1}^{n} \frac{y_i}{\pi_i}$
$Linear\ Regression, \bar{T}_{REG}$	$\bar{T}_{REG} = \sum_{i\in S} y_i + \sum_{i\in R} (\bar{\alpha} + \bar{\beta} x_i)$
$Dorfman, \bar{T}_{DORF}$	$\bar{T}_{DORF} = \sum_{S} Y_i + \sum_{P-s} \bar{m}\,(x_j)$
$LPRE, \bar{T}_0$	$\bar{T}_0 = \sum_{i\in S} Y_i + \sum_{j\in R} \bar{m}_0\,(x_j)$
$LPRE, \bar{T}_1$	$\bar{T}_1 = \sum_{i\in S} Y_i + \sum_{j\in R} \bar{m}_1\,(x_j)$

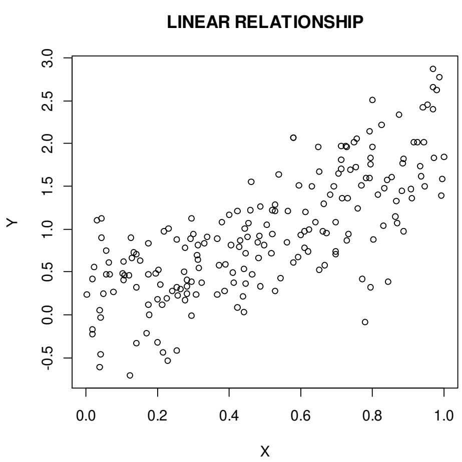

Figure 1. Scatter Diagram for the Linear Data

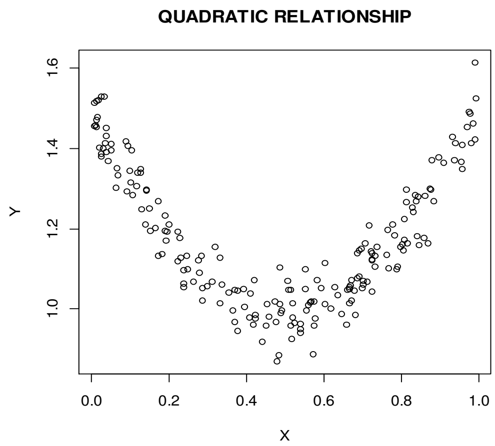

Figure 2. Scatter Diagram for the Quadratic Data

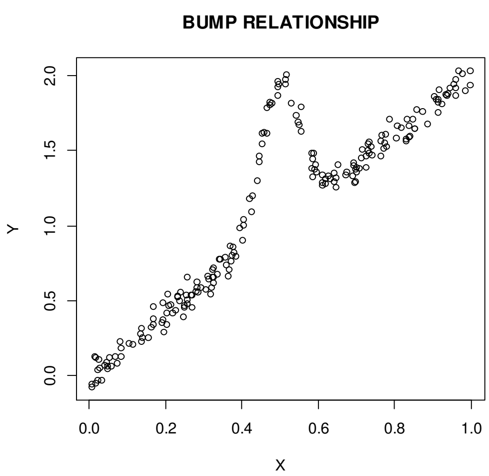

Figure 3. Scatter Diagram for the Bump Data

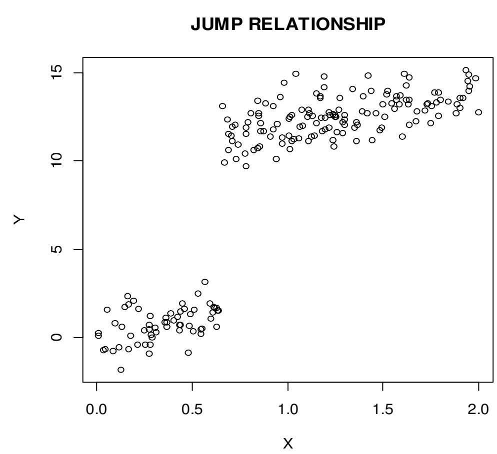

Figure 4. Scatter Diagram for the Jump Data

The population totals, prediction errors, the biases, absolute biases, efficiencies, MSEs and AREs for the estimators of finite population totals have been computed. The relative efficiencies (RE) which examine the robustness of various estimators, i.e. the Horvitz-Thompson estimator, the REG estimator and the Dorfman estimator versus the proposed robust estimators have also been computed. Further, the 95 % confidence intervals (CI) and the average lengths (AL) of the confidence intervals of various estimators have been constructed (see Kikechi et al (2017), Kikechi et al (2018) and Kikechi and Simwa (2018)).

4.2 Results

The results for the absolute biases, mean squared errors, relative efficiencies, confidence intervals and average length of confidence intervals for the various estimators are provided in tables 3, 4, 5, 6, 7 and 8 respectively.

Table 3: The Absolute Bias of various estimators in four populations

ABSOLUTE BIAS				
	HORVITZ-THOMPSON (HT)	**LINEAR REGRESSION (REG)**	**DORFMAN (DORF)**	**LOCAL LINEAR (LL)**
Linear	139.1395	3.650095	3.628214	3.626798
Quadratic	163.4725	1.226636	0.403125	0.4323062
Bump	157.7427	2.018801	0.4777851	0.4087753
Jump	1219.668	21.785	9.760465	9.485367

Table 4: The Mean Squared Error (MSE) of various estimators in the four populations

THE MEAN SQUARE ERROR (MSE)				
	HORVITZ-THOMPSON (HT)	**LINEAR REGRESSION (REG)**	**DORFMAN (DORF)**	**LOCAL LINEAR (LL)**
Linear	514.9775	15.36639	15.74559	15.47903
Quadratic	453.5207	1.521063	0.1713249	0.160443
Bump	548.131	4.551133	0.2942485	0.1894413
Jump	35691.94	512.8734	110.7915	97.02299

Table 5: Relative Efficiency of various estimators versus the proposed estimators

RELATIVE EFFICIENCY			
	HORVITZ-THOMPSON (HT)	**LINEAR REGRESSION (REG)**	**DORFMAN (DORF)**
	Relative Efficiency	**Relative Efficiency**	**Relative Efficiency**
Linear	0.09467563	0.8093	0.95664
Quadratic	0.000464731	0.9954403	0.962707
Bump	0.0002038478	0.02743355	0.9433107
Jump	0.003577862	0.1901854	0.9706123

Table 6: Confidence Intervals of various Estimators with respect to the four populations

95% CONFIDENCE INTERVALS								
	HORVITZ-THOMPSON (HT)		**LINEAR REGRESSION (REG)**		**DORFMAN (DORF)**		**LOCAL LINEAR (LL)**	
	Lower Limit	Upper Limit	Lower Limit	Upper Limit	Lower Limit	Upper Limit	Lower Limit	Upper Limit
Linear	65.43579	78.35652	62.92036	63.24861	62.75978	63.01298	62.62953	63.06378
Quadratic	61.74714	62.41275	60.29736	60.30645	60.25827	60.27853	60.44418	60.47615
Bump	88.43077	92.85335	93.01087	93.14516	92.06424	93.34889	91.91642	93.18671
Jump	503.6836	565.5807	479.9458	495.7306	460.7667	479.1529	465.1171	483.1778

Table 7: Average Length of Confidence Intervals of various Estimators

AVERAGE LENGTH OF CONFIDENCE INTERVALS				
	HORVITZ-THOMPSON (HT)	**LINEAR REGRESSION (REG)**	**DORFMAN (DORF)**	**LOCAL LINEAR (LL)**
Linear	12.92073	0.3282467	0.2532001	0.4342478
Quadratic	0.6656047	0.009090092	0.02025908	0.03197243
Bump	4.422574	0.1342954	1.284649	1.270295
Jump	61.8971	15.78477	18.38621	18.06073

Table 8: The Bias and MSE for $\bar{T}_0$ and $\bar{T}_1$ in the three artificial populations

	Linear		**Quadratic**		**Bump**	
	$\bar{T}_0$	$\bar{T}_1$	$\bar{T}_0$	$\bar{T}_1$	$\bar{T}_0$	$\bar{T}_1$
BIAS	5.507608	3.777348	4.7372	0.45116	5.293896	0.4187236
MSE	100.8874	15.40735	18.40769	0.1601695	43.9272	0.1896261

5. Discussion

In all the populations considered according to table 3, the Horvitz-Thompson estimator was the poorest resulting in large biases as compared to the other three finite population total estimators. For all the biases computed, the Local Linear Regression estimator is superior and dominates the Horvitz-Thompson estimator and the Linear Regression estimator in all the populations in consideration. The Local Linear regression estimator also dominates the Dorfman estimator for all the populations except when the population is quadratic.

The MSE results in table 4 indicate that the Local Linear estimators outperform the Linear Regression estimator in all the populations except when the population is linear. The Local Linear Regression estimators are not only superior to the popular Kernel Regression estimators, but they are also the best among all linear smoothers including those produced by orthogonal series and spline methods. In general, Local Linear regression estimation removes a bias term from the kernel estimator, that makes it have better behavior near the boundary of the $x's$ and smaller MSE everywhere.

Further, results in table 5 show that the relative efficiency of the proposed Local Linear estimators to the Horvitz-Thompson estimator, the REG estimator and the Dorfman estimator is less than1. This implies that the proposed Local Linear estimators have a smaller variance than the three estimators and thus the three estimators are less efficient. Generally, the Local Linear regression estimators outperform the HT estimator, the REG estimator and the DORF estimator in all the populations implying that they are robust and are the most efficient estimators.

In table 6, the confidence intervals indicate that the Local Linear regression method dominates the REG and Dorfman methods when the model is incorrectly specified. Generally, the model based estimators are much far better than the traditional design based estimators. The biases under the model based approach are also much lower than those for the design based approach in different populations under consideration.

Finally, we observe in table 8 that the biases and MSEs computed for the local polynomial regression estimator $\bar{T}_1$ are small in all the three populations. The results therefore indicate that the local polynomial regression estimator $\bar{T}_1$ is superior and dominates the local polynomial regression estimator $\bar{T}_0$ for the linear, quadratic and bump populations and thus $\bar{T}_1$ is the most efficient estimator.

6. Conclusion

In this article, we have reviewed and presented model based robust estimators of finite population totals using the procedure of local linear regression as studied by Kikechi et al (2017), Kikechi et al (2018) and Kikechi & Simwa (2018). Results of the bias, mean squared error, relative efficiency, confidence intervals and average length of confidence intervals for the various estimators have been presented. The bias results show that the local linear regression estimators dominate the Horvitz-Thompson estimator for the linear, quadratic, bump and jump populations. The MSE results show that the local linear estimators are performing better than the Horvitz-Thompson estimator and Dorfman estimator, irrespective of the model specification or misspecification. Results also show that the local linear regression estimators are robust and are the most efficient ones.

Results further indicate that the confidence intervals generated by the model based local linear procedure are much tighter than those generated by the design based Horvitz-Thompson method, regardless of whether the model is specified or misspecified. It has been observed that the model based approach outperforms the design based approach at 95% coverage rate.

Generally, the local linear regression estimators are not only superior to the popular kernel regression estimators, but are also the best among all linear smoothers including those produced by orthogonal series and spline methods. The estimators adapt well to bias problems at boundaries and in regions of high curvature and do not require smoothness and regularity conditions required by other methods such as boundary kernels. Simulation experiments carried out on the proposed Local Linear regression estimators in comparison with some estimators that exist in the literature indicate that the proposed estimators are robust and are the most efficient estimators.

Further, the local polynomial regression estimators $\bar{T}_0$ and $\bar{T}_1$ of finite population totals have been studied and comparisons made. Analytically, variance comparisons are explored using the local polynomial regression estimator $\bar{T}_0$ for $\mathrm{P} = 0$ and the local polynomial regression estimator $\bar{T}_1$ for $\mathrm{P} = 1$ in which results indicate that the estimators are asymptotically equivalently efficient. Simulation experiments carried out in terms of the biases and MSEs show that the local polynomial regression estimator $\bar{T}_1$ outperforms the local polynomial regression estimator $\bar{T}_0$ in all the three artificial populations and therefore, $\bar{T}_1$ is the most efficient estimator.

References

Bolfarine, H. and Zacks, S. (1991). "Bayes and Minimax Prediction in Finite Populations." *Journal of Statistical Planning and Inference*, 28 (2), 139-151.

Cochran, W.G. (1977). "Sampling techniques." Newyork: *John wiley and sons Ltd.*

Dorfman, A. (1992). Nonparametric Regression for Estimating Totals in Finite Populations, Proceedings of the Section on Survey Research Methods. *American Statistical Association*, 622-625.

Fan, J. (1992). Design Adaptive Nonparametric Regression. *Journal of American Statistical Association, 87*, 998-1004.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Annals of Statistics, 21*, 196–216. https://doi.org/10.1214/aos/1176349022

Fan J. and Gijbels I. (1996). *Local Polynomial Modeling and its Applications.* London: Chapman and Hall.

Hardle, W. (1989). "Applied Nonparametric Regression Analysis." *Cambridge: Cambridge University Press.*

Horvitz, D.G. and Thompson, D.J. (1952) A Generalization of Sampling without Replacement from a Finite Universe. *Journal of American Statistical Association*, 47, 663-685. https://doi.org/10.1080/01621459.1952.10483446

Kikechi C.B., Simwa R.O. and Pokhariyal G.P. (2017). On Local Linear Regression Estimation in Sampling Surveys. *Far East Journal of Theoretical Statistics,, 53*(5), 291-311. http://dx.doi.org/10.17654/TS053050291

Kikechi C.B., Simwa R.O. and Pokhariyal G.P. (2018). On Local Linear Regression Estimation of Finite Population Totals in Model Based Surveys. *American Journal of Theoretical and Applied Statistics, 7*(3), 92-101. https://doi: 10.11648/j.ajtas.20180703.11

Kikechi, C.B. and Simwa, R.O. (2018). On Comparison of Local Polynomial Regression Estimators for $P = 0$ and $P = 1$ in a Model Based Framework. *International Journal of Statistics and Probability, 7*(4), 104-114. https://doi.org/10.5539/ijsp.v7n4p104

Montanari, G.E. and Ranalli, M.G. (2005) Nonparametric Model Calibration Estimation in
Survey Sampling. *Journal of the American Statistical Association*, 100, 1429-1442.
https://doi.org/10.1198/016214505000000141

Nadaraya, E.A. (1964). On Estimating Regression. *Theory of Probability Applications, 10*, 186-190.

Ruppert D. and Wand M.P. (1994). Multivariate Locally Weighted Least Squares Regression. *Annals of Statistics, 22*, 1346-1370. https://doi.org/10.1214/aos/1176325632

Sanchez-Borrego, I.R. and Rueda, M. (2009) A Predictive Estimator of Finite Population
Mean Using Nonparametric Regression. Computational Statistics 24,1-14. https://doi.org/10.1007/s00180-008-0140-x

Silverman, B. (1986). "Densty Estimation for Statistics and Data Analysis," Monographs on Statistics and Applied Probability, Chapman and Hall, London.

Stone, C. (1977). Consistent Nonparametric Regression. *Annals of Statistics, 5*, 595-645.

Watson, G. (1964). Smooth Regression Analysis. *Sankhya Series A, 26*, 359-372.

Zheng, H. and Little, R.J. (2003). “Penalized Spline Model-Based Estimation of the Finite Population Total from Probability-Proportional-To-Size Samples.” *Journal of Official Statistics* 19, 99–117.