



Optimal Nonparametric Regression Estimation of Finite Population Total Using Nadaraya Watson Incorporating Jackknifing

Imboga Orang'o Herbert, George Otieno Orwa, Romanus Odhiambo Otieno

Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

Email address:

imbogaaherbert@yahoo.com (I. O. Herbert)

To cite this article:

Imboga Orang'o Herbert, George Otieno Orwa, Romanus Odhiambo Otieno. Optimal Nonparametric Regression Estimation of Finite Population Total Using Nadaraya Watson Incorporating Jackknifing. *International Journal of Theoretical and Applied Mathematics*. Vol. 3, No. 3, 2017, pp. 122-128. doi: 10.11648/j.ijtam.20170303.14

Received: April 17, 2017; **Accepted:** May 27, 2017; **Published:** June 30, 2017

Abstract: In this study a model based approach is adopted and a robust estimator of the jackknifed Nadaraya Watson estimator of the finite population total is proposed by incorporating the jackknifed procedure into the nonparametric regression estimator (the case of Nadaraya Watson). The study sought to estimate the finite population total using the proposed estimator (Jackknifed Nadaraya Watson). The study also looked at the various approaches of estimation of finite population totals and their properties. To measure the performance of each estimator, the study considered the average bias, the efficiency by the use of mean squared error and robustness using the rate of change of efficiency. Numerical study using simulated population was employed to examine the performance of the proposed estimator and compared it with the already existing estimators (Horvitz-Thompson, Nadaraya Watson, Ratio estimator). The simulation experiment showed that the proposed estimator records better results in terms of Bias and mean squared errors (MSE).

Keywords: Homoscedasticity, Jackknifing, Optimal, Robustness, Efficiency

1. Introduction

In many complex surveys, available information about the study population can be used at the design and estimation stages to construct efficient procedures for the finite population quantities i.e. population total or mean so as to increase the precision of the estimators of such population quantities. The information can be collected by national census, official registers, natural resources inventories and remote sensing data. Estimation being the main concern in surveys, emphasis is usually on the use of the auxiliary information. One of the approaches is to assume a working model and more often a linear model describing the relationship between the survey variable and auxiliary variable is selected. Estimators are then derived from this linear model. However for efficient use of any of these estimators prior knowledge of the specific parametric structure of the population needs to be known and this is usually problematic especially if the model is to be used for many variables [1]. Because of these concerns more focus has been given to non-parametric models describing the relationship between the auxiliary variables and the study variables are assumed [2]. The idea of nonparametric traces its

origin in works by [3]. Non-parametric based estimation is often more robust and flexible than inference based on parametric regression models or design probabilities (as in the case of design-based inference) [4]. A variety of approaches exist in the construction of more efficient estimators and they include; Model-based and design based methods. In Model-based approach, the idea is based on super population models which assumes that the population under study is a realization of a random variable having a super population model ζ . The model ζ is used to predict the non-sampled values of the population hence finite population quantities [5]. [6] First considered non parametric models for ζ and obtained a local polynomial regression estimator as a generalization of the ordinary generalized regression estimator. From the simulation, their study showed the proposed estimator performed better than the other parametric estimators. [7] Improved on [6] estimator and developed a model-based local polynomial regression estimator applicable to direct sampling designs i.e. simple random sampling and systematic sampling. Their estimator demonstrated better results than [7]. In this study auxiliary information is used to determine the estimate of finite population total using non-parametric regression in the

case where jackknifing procedure is incorporated into the Nadaraya Watson estimator.

2. Review of the Jackknifing Estimator

Jackknifing is one of the re-sampling techniques that are commonly used to reduce bias found in estimators and also assessing variability. Basically in jackknifing procedure, the idea is generalized into splitting the sample into g groups of size h each. This thus implies that $n = gh$. The procedure involves the following: Let Y_1, Y_2, \dots, Y_n be a sample of independent and identically distributed random variables. Again let $\hat{\theta}$ be an estimator of the parameter θ based on sample of size n . Let $\hat{\theta}_{-i}$ be the corresponding estimator based on the sample of size $(g - 1)h$, where the i^{th} group of size h has been deleted. [8] and [9] defined that

$$\tilde{\theta}_i = g\hat{\theta} - (g - 1)\hat{\theta}_{-i} \text{ for } i = (1, 2, 3, \dots, g) \tag{1}$$

Then the estimator

$$\tilde{\theta} = \frac{1}{g} \sum_{i=1}^g \tilde{\theta}_i = g\hat{\theta} - (g - 1) \frac{1}{g} \sum_{i=1}^g \hat{\theta}_{-i} \tag{2}$$

It’s important to note that the estimator (2) has the property that it eliminates the order $\frac{1}{n^2}$ term from bias of the form $E(\hat{\theta}) = \theta + \frac{\alpha_1}{n} + O(\frac{1}{n^2})$

3. Jackknifing in Model Based Estimation of Finite Population Totals

The algorithm of estimating the total using the Nadaraya Watson with the Jackknife technique is given in this section. The technique of deleting a single case from the Original sample (delete one jackknife) sequentially will be adopted. Suppose that a database consists of n vectors $(Y_i, X_{1i}, X_{2i}, \dots, X_{ni})$, where Y_i is the study variable and $X_{1i}, X_{2i}, \dots, X_{ni}$ are considered auxiliary variables. Let $X_i = X_{1i}, X_{2i}, \dots, X_{ni}$ and $d_k = (y_k, x_k), k = 1, 2, \dots, n$ denote the values associated with the i^{th} observation. In this case, the set observations is the vector (d_1, d_2, \dots, d_n) . Then the Jackknife Procedure based on delete one is as follows:

- i. Draw n sized sample from a population randomly and label the elements of the vector $d_k = (y_k, x_k), k = 1, 2, 3, \dots, n$.
- ii. Omit one observation of the vector $d_k = (y_k, x_k), k = 1, 2, 3, \dots, n$. and label the remaining $n - 1$ sized observations set $Y_1^J = (y_1, y_2, \dots, y_{n-1})$ and $X_1^J = (x_1, x_2, \dots, x_{n-1})$ as delete -one Jackknife sample d_1^J .

$$Var(\hat{e}_{ij}) = Var(Y_{ij}) - E[Y_{ij}]^2 - 2Cov(Y_{ij}, \hat{\mu}(x_{ij})) + 2E[Y_{ij}]E[\hat{\mu}(x_{ij})] + Var(\hat{\mu}(x_{ij})) + E[\hat{\mu}(x_{ij})]^2 \tag{4.4}$$

But Y_{ij} and $\hat{\mu}(x_{ij})$ are independent, hence the above equation (4.1) simplifies to

$$\begin{aligned} Var[\hat{e}_{ij}] &= Var(Y_{ij}) + Var(\mu(x_{ij})) \\ &= \sigma^2(x_{ij}) + Var(\mu(x_{ij})) \end{aligned} \tag{6}$$

- iii. Obtain the Nadaraya Watson estimate \hat{m}^{J1} from d_1^J .
- iv. Omit another observation of the vector $d_i = (y_i, x_i)$, and label the remaining $n - 1$ Sized observations set $Y_2^J = (y_1, y_2, \dots, y_{n-1})$ and $X_2^J = (x_1, x_2, \dots, x_{n-1})$ as delete -one Jackknife sample d_2^J .
- v. Obtain the Nadaraya Watson estimate \hat{m}^{J2} from d_2^J .
- vi. Similarly omit one of the n observations (there are n jackknife samples and each of them and have $n - 1$ observations) and estimate the Nadaraya Watson $\hat{m}^{JK}(x_j)$, where $\hat{m}^{JK}(x_j)$ is the Jackknife Nadaraya Watson estimate after deleting the k^{th} observation from $d_k = (y_k, x_k)$.

4. The Proposed Estimator

Dorfman [3] introduced a non-parametric regression estimator for finite population total based on a sample drawn from the population. Taking into consideration a population consisting of N units, the author sought to estimate the finite population total defined by:

$$T = \sum_{i=1}^N y_i \text{ for } i = (1, 2, 3, \dots, N) \tag{3}$$

We propose the Jackknife estimate of the population total to be defined as

$$\hat{T}_{jnw} = \sum_{i=1}^n y_i + \sum_{j \neq i}^N \hat{m}^J(x_j) \tag{4}$$

$$= \sum_{i=1}^n y_i + \sum_{j \neq i}^N \frac{1}{n} \sum_{k=1}^n w_{ij}(x_{ij}) y_{ij} \tag{5}$$

Where

$$\hat{m}^J(x_j) = \frac{1}{n} \sum_{k=1}^n \hat{m}^{jk}(x_j)$$

Now on deriving the asymptotic variance of the error term, we note that It’s easy and it can be shown that $E[\hat{e}_{ij}] = 0$. Therefore,

$$\begin{aligned} Var[\hat{e}_{ij}] &= E[e_{ij}^2] \\ &= E[Y_{ij} - \hat{\mu}(x_{ij})]^2 \\ &= E[Y_{ij}^2] - 2E[Y_{ij}\hat{\mu}(x_{ij})] + E[\hat{\mu}(x_{ij})]^2 \end{aligned}$$

But

$$Cov(Y_{ij}, \mu(x_{ij})) = E(Y_{ij}, \hat{\mu}(x_{ij})) - E(Y_{ij})E(\mu(x_{ij}))$$

From this we get that

Computing $Var(\mu(x_{ij}))$, we have that

$$\begin{aligned} Var[\hat{\mu}(x_{ij})] &= Var\left[\frac{n}{n-1}\sum_{ies} (n-1)^{-1}b^{-1}K\left(\frac{x_{ij}-x_{ij}}{b}\right)Y_{ij}\right](d_s(\hat{x}_{ij}))^{-1} \\ &= \left(\frac{n}{n-1}\right)^2 \sum_{ies} (n-1)^{-2}b^{-2}K\left(\frac{x_{ij}-x_{ij}}{b}\right)^2 \sigma^2(x_{ij})(d_s(\hat{x}_{ij}))^{-2} \end{aligned}$$

Now next we obtain the asymptotic expansion of $Var[\hat{\mu}(x_{ij})]$ as follows

$$[d_s(x_{ij})]^{-1} = [d_s(x_{ij})]^{-1} \left[1 - \frac{b^2}{d_s(x_{ij})} K_2 d_s''(x_{ij}) + O(b^3 + (n-1)^{-1/2}b^{-1/2})\right]$$

Therefore

$$\begin{aligned} [d_s(x_{ij})]^{-2} &= [d_s(x_{ij})]^{-2} \left[1 - \frac{b^2}{d_s(x_{ij})} K_2 d_s''(x_{ij}) + O(b^3 + (n-1)^{-1/2}b^{-1/2})\right]^2 \\ &= [d_s(x_{ij})]^{-2} \left[1 - \frac{b^2}{d_s(x_{ij})} K_2 d_s''(x_{ij}) + O(b^3 + (n-1)^{-1/2}b^{-1/2})\right] \end{aligned} \quad (7)$$

But we know that

$$K\left(\frac{x_{ij}-x_{ij}}{b}\right)\mu(x_{ij}) = b\mu(x_{ij})d_s'(x_{ij}') + O(b^3 + b^{1/2})$$

Then similarly

$$K\left(\frac{x_{ij}-x_{ij}}{b}\right)\sigma(x_{ij}) = \sigma(x_{ij})d_s(x_{ij}) + O(b^3 + b^{1/2})$$

Now

$$\begin{aligned} \left[b^{-1}(n-1)^{-1}K\left(\frac{x_{ij}-x_{ik}}{b}\right)\sigma(x_{ij})\right]^2 &= b^{-2}(n-1)^{-1}[b\sigma(x_{ij})d_s(x_{ij}) + O(b^3 + b^{1/2})]^2 \\ &= b^{-2}(n-1)^{-2}[b^2\sigma^2(x_{ij})d_s^2(x_{ij}) + 2b\sigma(x_{ij})d_s(x_{ij})O(b^3 + b^{1/2}) + [O(b^3 + b^{1/2})]^2] \\ &= (n-1)^{-2}[\sigma^2(x_{ij})d_s^2(x_{ij}) + 2\sigma^2(x_{ij})d_s(x_{ij})O(b^2 + b^{-1/2}) + (O(b^4 + b^{-1/2}))] \end{aligned}$$

Therefore

$$\sum_{ies} \left[b^{-1}(n-1)^{-1}K\left(\frac{x_{ij}-x_{ik}}{b}\right)\sigma(x_{ij})\right]^2 (n-1)^{-1}[\sigma^2(x_{ij}) + O(b^4 + b^{-1/2}) + O(b^4 + b^{-1})]$$

This then follows that

$$Var(\mu(x_{ij})) = \left(\frac{n}{n-1}\right)^{1/4} \sum_{ies, i \neq k} (n-1)^{-2}b^{-2}K\left(\frac{x_{ij}-x_{ik}}{b}\right)^2 \sigma(x_{ij})(\hat{d}_s(x_{ij}))^2$$

So that

$$\begin{aligned} Var(\hat{\mu}(x_{ij})) &= \left(\frac{n}{n-1}\right)^{1/4} \left[(n-1)^{-1} \{ \sigma^2(x_{ij})d_s(x_{ij}) + O(b^2 + b^{-1/2}) + O(b^2 + b^{-1}) \} \right. \\ &\quad \left. * [d_s(x_{ij})]^{-2} \left[1 - \frac{b^2}{d_s(x_{ij})} K_2 d_s''(x_{ij}) + O(b^3 + (n-1)^{-1/2}b^{-1/2}) \right] \right] \end{aligned}$$

$$\begin{aligned}
 & \left(\frac{n}{n-1}\right)^{-\frac{1}{4}} (n-1)^{-1} \left\{ \sigma^2(x_{ij}) + b^2 \sigma^2(x_{ij}) K_2 d_s(x_{ij}) d''(x_{ij}) + O(b^3 + (n-1)^{-\frac{1}{2}} b^{-\frac{1}{2}}) \right\} \\
 &= \left(\frac{n}{n-1}\right)^{-\frac{1}{4}} (n-1)^{-1} [\sigma^2(x_{ij}) + O((n-1)^{-\frac{1}{2}} b^{-\frac{1}{2}})] \\
 &= \left(\frac{n}{n-1}\right)^{-\frac{1}{4}} (n-1)^{-1} [\sigma^2(x_{ij}) + O((n-1)^{-\frac{1}{2}} b^{-\frac{1}{2}})] \\
 &= \left(\frac{n}{n-1}\right)^{-\frac{1}{4}} \frac{1}{n-1} [\sigma^2(x_{ij}) + O((n-1)^{-\frac{1}{2}} b^{-\frac{1}{2}})] \\
 &= \left(\frac{n}{n-1}\right)^{-\frac{1}{4}} \left[\frac{\sigma^2(x_{ij})}{n-1} + \frac{1}{n-1} \sigma^2(x_{ij}) ((n-1)^{-\frac{1}{2}} b^{-\frac{1}{2}}) \right] \tag{8}
 \end{aligned}$$

But for large number of subgroups $n - 1 \cong n$. therefore

$$Var[\hat{\mu}(x_{ij})] = \frac{\sigma(x_{ij})}{n-1} + O\left((n-1)^{-\frac{3}{2}} b^{-\frac{1}{2}}\right)$$

And as $nb \rightarrow \infty$ and so

$$Var[\hat{\mu}(x_{ij})] = \frac{\sigma(x_{ij})}{n-1}$$

Hence

$$\begin{aligned}
 Var[\hat{\mu}(x_{ij})] &= \sigma^2(x_{ij}) + \frac{\sigma^2(x_{ij})}{n-1} \\
 &= \frac{n}{n-1} \sigma^2(x_{ij})
 \end{aligned}$$

But as n tends to be large, $n - 1 \cong n$ therefore follows that $Var[e_{ij}] = \sigma^2 x_{ij}$

5. Description of the Population

In the study four populations are considered, which are generated from regression model [10] defined as;
 $y_i = m(x_i) + e_i$ With the following mean functions;

$$\begin{aligned}
 \text{linear: } m_1 &= 1 + 0.5 * (X - 0.25) \\
 \text{Quadratic: } m_2 &= 1 + 0.5 * (X - 0.25)^2
 \end{aligned}$$

With the following regression functions;

$$\text{Homogeneous Population: } Y_1 = m_x + e_i$$

$$\text{Heterogeneous Population: } Y_2 = m_x + e_i * X$$

e_i : rnorm (1000, mean=0, sd=.1) #the random error

X : runif (1000, min = 0, max = 1) #the explanatory variable

6. Simulation Results

The results of the study are summarized in the tables 1 to 8 and the plots 1 to 4. On each population the performance of

each estimator is analyzed using the average bias and mean squared error. The average bias is an indication of the measure of how close the estimator is from the true value, while the MSE is used to assess the efficiency of an estimator. For each combination of mean functions, standard deviation (s.d=0.1), and at optimal bandwidth $h=0.1429514$, 1000 replicates from samples of the four (4) populations are selected and the estimators calculated. Each of the table 1-4 show the ratios of the MSE’s for all the estimators. The performance of the Jackknifed Nadaraya Watson Estimator in the four selected populations (I, II, II, and IV) is consistently good. With respect to the MSE the Jackknifed Nadaraya Watson Estimator dominates all the other estimators.

In table 3 the Jackknifed Nadaraya Watson estimator performs better than Nadaraya Watson estimator and the Ratio estimator except the Horvitz-Thompson estimator, this is because the mean function is correctly specified and the other estimators are considered competitive.

The Biases for the Jackknifed Nadaraya Watson are smaller compared to the Horvitz-Thompson estimator and the Ratio Estimator but slightly larger compared to the Nadaraya Watson. Table 1-3 shows biases of the Jackknifed Nadaraya Watson having negative values a manifestation that the Jackknifed Nadaraya Watson estimator tends to underestimate the True population total while the Nadaraya Watson estimator also tends to overestimate the True population total but only in cases where the sample size is large enough.

The confidence lengths generated by the Nadaraya Watson estimator are much tighter than those generated by JNW, HT, and Ratio estimator’s at large bandwidths but at lower bandwidths, the JNW gives much tighter confidence lengths compared to the other estimators. Note that the best performing estimator is one whose coverage rate is close to the true population total and its length is small i.e a smaller confidence length is better because it implies that the true population total is captured within a small range and therefore the results are more precise.

Table 1. Population I (Linear and Homoscedastic).

	NW Estimator	JNW Estimator	HT Estimator	RATIO Estimator
Pop. Estimate	1122.905	1120.369	1121.751	1709.218
Bias	1.239	-1.297	-4.278	587.553
MSE	16.441	1.705	18.445	386049.800
Confidence Length	5.357	5.992	7.915	293.345

True Population Total is 1121.665

Table 2. Population II (Linear and Heteroscedastic).

	NW Estimator	JNW Estimator	HT Estimator	RATIO Estimator
Pop. Estimate	1123.959	1120.219	1122.934	1720.742
Bias	0.982	-2.759	-5.680	597.764
MSE	9.717	7.632	32.367	397490.700
Confidence Length	4.165	4.985	6.957	284.892

True Population Total is 1122.978

Table 3. Population III (Quadratic and Heteroscedastic).

	NW Estimator	JNW Estimator	HT Estimator	RATIO Estimator
Pop. Estimate	1071.964	1070.454	1071.479	1881.422
Bias	0.523	-0.987	-0.169	809.981
MSE	5.613	0.983	0.072	711719.400
Confidence Length	3.207	3.600	4.694	337.541

True Population Total is 1071.441

Table 4. Population IV (Quadratic and Homoscedastic).

	NW Estimator	JNW Estimator	HT Estimator	RATIO Estimator
Pop. Estimate	1070.912	1070.604	1070.296	1869.899
Bias	0.781	0.477	1.231	799.769
MSE	11.899	0.236	1.584	696017.500
Confidence Length	4.503	4.825	5.540	343.391

True Population Total is 1070.129

6.1. Conditional Properties of the Nonparametric Estimators

The figures below show behavior of the conditional bias for each estimator when various mean functions were used. In figure 1 and 2 the Ratio estimator and the HT estimator performed well when a linear function was used. This is because the Ratio Estimator is the Best Linear Unbiased Estimator (BLUE). It can be observed that the biases to the left of the population mean of the auxiliary variable ($\bar{x} = 248$), are small but reduce towards the right. As the population grows larger the Jackknifed Nadaraya Watson estimator performs well than all the other estimators.

In figure 3 and 4 the Quadratic mean functions were used. The proposed estimator (JNW) gives better estimates of the population total compared to the ones realized using Nadaraya Watson estimator that was proposed by [3] and the Ratio estimator. It can be observed that the biases of JNW remains lowest throughout.

6.2. Plots for the Conditional Biases

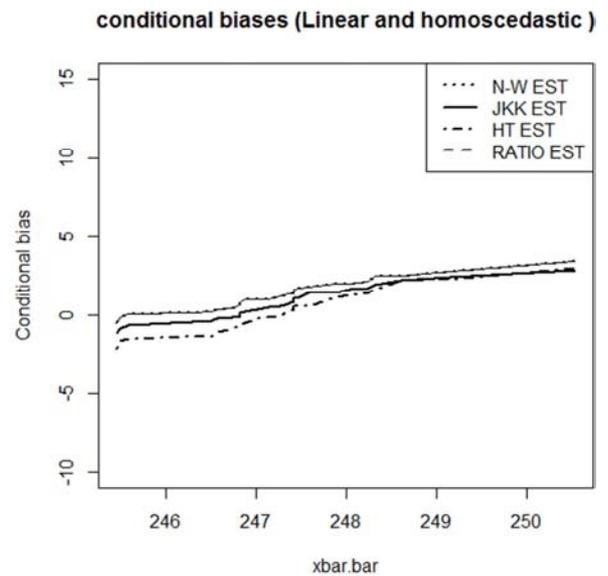


Figure 1. Plot of conditional biases (Linear and Homoscedastic).

conditional biases (Linear and heteroscedastic)

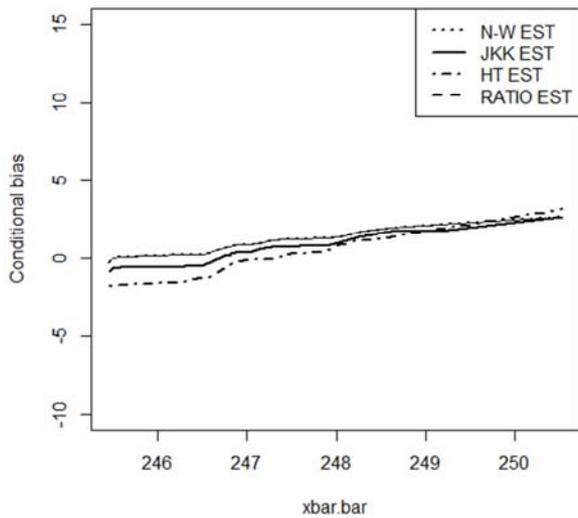


Figure 2. Plot of conditional biases (Linear and Heteroscedastic).

conditional biases (Quadratic and heteroscedastic)

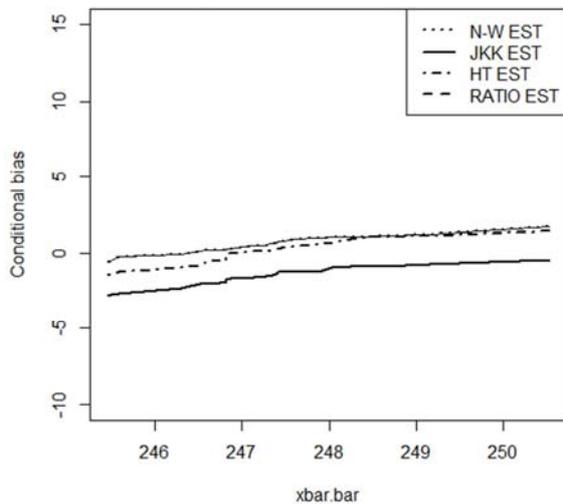


Figure 3. Plot of conditional biases (Quadratic and Heteroscedastic).

conditional biases (Quadratic and homoscedastic)

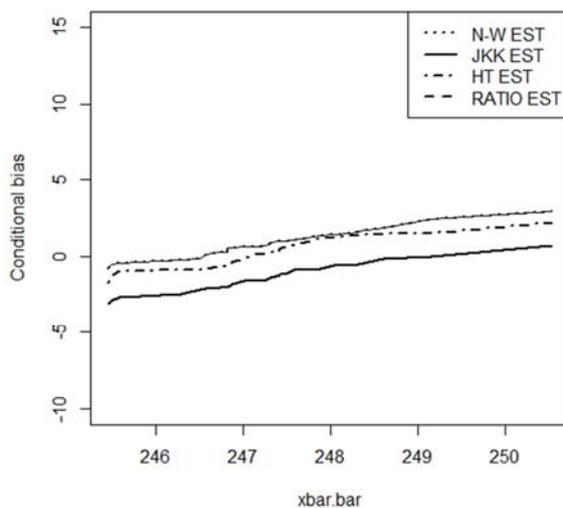


Figure 4. Plot of conditional biases (Quadratic and Homoscedastic).

6.3. Performance of the Estimators at Varying Bandwidths

From table 5 and 6 the two estimators (Nadaraya Watson and Jackknifed Nadaraya Watson) perform well under the linear and homoscedastic condition. However in comparison with JNW the Nadaraya Watson estimator has lower bias and the values of JNW are negative indicating that it tends to underestimate the value of the True population total. The small values of the bias indicate that the estimated values of the population total are closer from the True population value. The values of the MSE from the JNW outperforms the Nadaraya Watson and this implies that the estimator (JNW) has efficiency in linear and homoscedastic population structure.

Table 7 and 8 (Quadratic structure) of the population, it can be noted that the Nadaraya Watson estimator has the least absolute bias compared to the jackknifed Nadaraya Watson estimator. It's important to also note that the Nadaraya Watson is the best estimator for Quadratic and homoscedastic population. The JNW estimator has the lowest MSE especially for small values of bandwidth (h) making it the most efficient estimator compared to the Nadaraya estimator.

Table 5. Population I (Linear and Homoscedastic).

	Bandwidth	NW Estimator	JNW Estimator
Population estimate	H=0.11	1122.924	1121.587
	H=0.12	1122.92	1121.577
	H=0.15	1122.892	1122.535
	H=0.2	1122.812	1121.445
	H=0.5	1122.325	1120.369
	H=1	1121.795	1118.337
	H=1.5	1121.698	1117.751
	H=2	1121.669	1117.494
True Population Total is 1121.665			
Bias	H=0.11	1.258162	-0.078820
	H=0.12	1.254551	-0.088061
	H=0.15	1.226852	-0.130155
	H=0.2	1.146493	-0.220614
	H=0.5	0.659967	-1.296946
	H=1	0.129356	-3.328162
	H=1.5	0.032733	-3.994402
	H=2	0.033173	-4.171915
M.S.E	H=0.11	16.43041	0.0324638
	H=0.12	16.43709	0.0339747
	H=0.15	16.43905	0.0430443
	H=0.2	16.45676	0.0744916
	H=0.5	18.47527	1.704679
	H=1	24.42786	11.0947
	H=1.5	27.65575	15.97182
	H=2	28.92265	17.42106

Table 6. Population II (Linear and Heteroscedastic).

	Bandwidth	NW Estimator	JNW Estimator
Population estimate	H=0.11	1122.967	1121.295
	H=0.12	1123.965	1121.293
	H=0.15	1123.95	1121.276
	H=0.2	1123.898	1121.216
	H=0.5	1123.499	1120.219
	H=1	1123.016	1118.224
	H=1.5	1122.907	1117.563
	H=2	1122.897	1117.391
True Population Total is 1122.978			

	Bandwidth	NW Estimator	JNW Estimator
Bias	H=0.11	0.989576	-1.683056
	H=0.12	0.987078	-1.68474
	H=0.15	0.972403	-1.701852
	H=0.2	0.920972	-1.761982
	H=0.5	0.521613	-2.75904
	H=1	0.038213	-4.753231
	H=1.5	-0.07104	-5.414173
	H=2	-0.08026	-5.586416
M.S.E	H=0.11	9.671998	2.855903
	H=0.12	9.696557	2.861541
	H=0.15	9.773706	2.919371
	H=0.2	9.920756	3.127357
	H=0.5	12.23737	7.631814
	H=1	18.32091	22.60808
	H=1.5	21.58017	29.32666
	H=2	22.85392	31.22104

Table 7. Population III (Quadratic and Heteroscedastic).

	Bandwidth	NW Estimator	JNW Estimator
Population estimate	H=0.11	1072.881	1072.288
	H=0.12	1071.987	1072.299
	H=0.15	1071.928	1072.31
	H=0.2	1071.752	1072.253
	H=0.5	1069.863	1070.454
	H=1	1068.892	1068.844
	H=1.5	1069.679	1069.449
	H=2	1070.296	1070.045
True Population Total is 1071.441			
Bias	H=0.11	0.559088	0.8468487
	H=0.12	0.545464	0.8576714
	H=0.15	0.486597	0.8689436
	H=0.2	0.311249	0.8114198
	H=0.5	-1.57864	-0.987426
	H=1	-2.54926	-2.597221
	H=1.5	-1.76273	-1.992509
	H=2	-1.14563	-1.396723
M.S.E	H=0.11	5.578517	0.7259672
	H=0.12	5.583277	0.7443973
	H=0.15	5.592069	0.7637942
	H=0.2	5.607622	0.6669852
	H=0.5	9.383778	0.9823074
	H=1	15.37644	6.751466
	H=1.5	12.95653	3.97562
	H=2	11.57031	1.956261

Table 8. Population IV (Quadratic and Homoscedastic).

	Bandwidth	NW Estimator	JNW Estimator
Population estimate	H=0.11	1070.953	1072.56
	H=0.12	1070.942	1072.583
	H=0.15	1070.87	1072.57
	H=0.2	1070.666	1072.482
	H=0.5	1068.689	1070.604
	H=1	1067.671	1068.957
	H=1.5	1068.47	1069.556
	H=2	1069.097	1070.147
True Population Total is 1070.129			
Bias	H=0.11	0.824427	2.451085
	H=0.12	0.812937	2.454351
	H=0.15	0.741046	2.44064
	H=0.2	0.536770	2.352787
	H=0.5	-1.44029	0.4746665
	H=1	-2.45813	-1.172151
	H=1.5	-1.65901	-0.572738
	H=2	-1.03217	0.0177783
M.S.E	H=0.11	11.91843	6.019772

	Bandwidth	NW Estimator	JNW Estimator
Bias	H=0.12	11.88930	6.035778
	H=0.15	11.81215	5.968596
	H=0.2	11.66469	5.547329
	H=0.5	14.75919	0.2367553
	H=1	20.69225	1.383034
	H=1.5	18.34619	0.3367494
	H=2	17.05216	0.0089369

7. Conclusions

From the summary tables above, it can be seen that the proposed estimator generally gives a total with a very small bias as compared to the Nadaraya Watson, Horvitz - Thompson, Ratio Estimator. Also, the proposed model can be seen to have a small Mean Square Error (MSE) as compared to Nadaraya Watson estimator. The graphs also shows that the population total estimate and conditional biases are more consistent in the proposed estimator as compared to Nadaraya Watson, Horvitz -Thompson, Ratio Estimator.

References

- [1] C. Wu and R. Little. A model-Calibration approach to using complete auxiliary information from Survey data. *Journal of American Statistical Association* 96, 185-193, 2001.
- [2] R. V. Allan Dorfman and R. Chambers. Estimation of Finite population distribution using non-parametric regression. *Annals of statistics* 21, 1452-21475, 1993.
- [3] R. V. Allan Dorfman and R. M Royall. Non-parametric regression for estimating population totals in finite populations. *American Statistical Association*, 1992.
- [4] Nadaraya, E. A. (1964) On Estimating Regression. *Theory of Probability and Applications*, 9, 141-142.
- [5] Dorfman, A. H. (1992) Nonparametric Regression for Estimating Totals in Finite Population. In *Section on Survey Research Methods*. *Journal of American Statistical Association*, 622-625.
- [6] Sanchez-Borrego, I. R. and Rueda, M. (2009) A Predictive Estimator of Finite Population Mean Using Nonparametric Regression. *Computational Statistics*, 24, 1-14.
- [7] Breidt, F. J. and Opsomer, J. D. (2000) Local Polynomial Regression Estimators in Survey Sampling.
- [8] Quenouille, M. H. (1949). Approximate tests of correlation in time series. *J. R. Statist. Soc. B* 11, 68-64.
- [9] Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika* 43, 54-60.
- [10] Otieno, R and Mwalili, T (2000) Non parametric regression for finite population estimation. *East Africa Journal of Science*, Vol II, part 2, 107-112.