Science Journal of Applied Mathematics and Statistics
2021; 9(1): 20-32
http://www.sciencepublishinggroup.com/j/sjams
doi: 10.11648/j.sjams.20210901.13
ISSN: 2376-9491 (Print); ISSN: 2376-9513 (Online)

Nonparametric Penalized Spline Model Calibrated Estimator in Complex Survey with Known Auxiliary Information at Both Cluster and Element Levels

Nthiwa Janiffer Mwende[1, *], Ali Salim Islam[1], Pius Nderitu Kihara[2]

[1]Department of Mathematics, Egerton University, Nakuru, Kenya

[2]Department of Financial and Actuarial Mathematics, Technical University of Kenya, Nairobi, Kenya

Email address:

janienthiwa@yahoo.com (N. J. Mwende), asislam54@yahoo.com (A. S. Islam), piuskihara@yahoo.com (P. N. Kihara)

[*]Corresponding author

To cite this article:

Nthiwa Janiffer Mwende, Ali Salim Islam, Pius Nderitu Kihara. Nonparametric Penalized Spline Model Calibrated Estimator in Complex Survey with Known Auxiliary Information at Both Cluster and Element Levels. *Science Journal of Applied Mathematics and Statistics.* Vol. 9, No. 1, 2021, pp. 20-32. doi: 10.11648/j.sjams.20210901.13

Received: November 27, 2020; **Accepted**: February 16, 2021; **Published**: March 10, 2021

Abstract: The present study uses penalized splines (p- spline) to estimate the functional relationship between the survey variable and the auxiliary variable in a complex survey design; where a population divided into clusters is in turn subdivided into strata. This study has considered a case of auxiliary information present at two levels; at both cluster and element levels. The study further applied model calibration technique by penalty function to estimate the population total. The calibration problems at both levels have been treated as optimization problems and solved using penalty functions to derive the estimators for this study. The reasoning behind model calibration is that if the calibration constraints are satisfied by the auxiliary variable, the study expects that the variable of interest's fitted values meets such constraints. This study runs a Monte Carlo simulation to assess the finite sample performance of the penalized spline model calibrated estimator under complex survey data. Simulation studies were conducted to compare the efficiency of p-spline model calibrated estimator with Horvitz Thompson estimator (HT) by mean squared error (MSE) criterion. This study shows that the p-spline model-based estimator is generally more efficient than the HT in terms of the mean squared error. The results have also shown that the estimator obtained is unbiased, consistent and very robust because it does not fail if the model is misspecified for the data.

Keywords: Penalized Spline, Nonparametric Model, Auxilliary Information and Optimization Problem

1. Introduction

In recent years, Nonparametric estimation methods have gained considerable attention due to their flexibility. One of these methods is the penalized spline estimation. This method requires knowing the number and locations of knots, the degree of the polynomial, and the degrees of freedom. The degrees of freedom are known as the equivalent number of parameters. There are practical rules in the existing literature to determine the degree of the polynomial and the number and locations of knots. The degrees of freedom are established according to the user's experience.

Splines can be classified as regression splines, cubic splines, B-splines, penalized-splines, natural splines, thin-plate splines, and smoothing splines, according to De Boor (2001), [3]. Nonparametric regression using splines has undergone extensive development in recent years. Smoothing splines (Eubank 1988; Wahba 1990) [6, 13] use a knot at each distinct value except the boundary values of the X-variable and control overfitting by applying a rough-ness penalty. Penalized splines (p-splines), formally introduced by Eilers and Marx (1996), [5], are in general computationally inexpensive and allow flexible knot selection yet yield sound performance. P-splines are also easy to implement: there is a close relationship which implies that they can be fitted using widely-available statistical software such as R software and S-Plus (Pinheiro and Bates, 2000), [9] function lme ().

The Horvitz-Thompson (HT) estimator (Horvitz and Thompson 1952), [7] is a standard design-unbiased estimator of population total and weights cases by the inverse of their inclusion probabilities (Zhen and little 2003), [15]. This estimator was used in this study as a baseline comparison with p-spline model calibrated estimator.

The concept of auxiliary variables in the present scholarship in statistics denotes independent or predictor variables in a regression analysis. As the name suggests, the variables offer additional information and may be used to improve the estimation of population parameters. Breidt and Opsomer (2000), [1] noted that micro-econometric research is frequently performed using data collected by surveys, which may contain auxiliary information for every unit of the population of interest. As can be expected, many of these surveys use complex sampling plans to reduce costs and increase the estimation efficiency for subgroups of the population. Complex sampling designs result in unequal sampling probabilities for the units in the sample and create data with correlations between observations, violating the assumption that the data are independently and identically distributed (iid) (Sayed, 2010), [12]. This means that the number of population units represented by a given sample unit is not uniform across all of the units in the sample. Although the word complex survey has been used mostly by researchers to refer to different combinations of sampling plans, however, in this research, complex survey refers to a mixture of both stratified and cluster sampling methods (Clair, 2016), [2].

In this study, Nthiwa Janiffer Mwende, et al. (2020), [8] work which considered auxiliary information at cluster level only is extended. This study extends this work to consider auxiliary information available at both the element and cluster levels. The extension is by treating the two levels of calibration problems in a cluster- Strata sampling, as constrained nonlinear optimization problems which is then converted into unconstrained optimization problems. The resulting problems were then solved using the penalty function method to obtain the weights at both cluster and cluster- strata element levels assigned to sample observations from some chi-square distance measures.

This study considered the generalized calibration procedure using model calibration as proposed by Wu and sitter (2001), [14]. They considered generalized linear and nonparametric regression models for the super population model ψ given in the equation below:

$$y_i = h(x_i) + \varepsilon_i \tag{1}$$

where $\{\varepsilon\}_{i=1}^{n}$ is a sequence of independent and identically distributed random variables with $E(\varepsilon_i) = 0$ and $E(\varepsilon_i^2) = \sigma^2$ and $h(x_i)$ is a smooth function that can be estimated using nonparametric methods like kernel, neural network, and penalized splines. Given n pair of sample observations $(x_1, y_1), \dots, (x_n, y_n)$ from a population of size N, of interest, is the estimator $\hat{h}(x)$ of $h(x) = E_{\psi}(y/x)$. For model calibration, calibration is performed to the population mean of the fitted values $\hat{h}_i(x_i)$ (Wu and sitter, 2001), [14]. The study considers a model calibration estimator for the population total Y_t given below.

$$\hat{Y} = \sum_{i \in a} w_i y_i \tag{2}$$

Where a is a set of sampled units under a general sampling design while w_i's are design weights such that for a given metric, are as close as possible in an average sense to the $z_i = \frac{1}{\pi_i}$. These weights are obtained by minimizing a given distance measure between the w_i's and z_i's subject to some constraints. The chi-squared distance measure to be minimized is as provided in the equation below

$$\delta_a = \sum_{i \in a} \frac{(w_i - z_i)^2}{q_i z_i} \tag{3}$$

where q_i 's are known positive constants uncorrelated with the z_i 's, (Deville and Sarndal, 1992), [4] subject to two constraints equation given below

$$\begin{cases} \sum_{i \in a} w_i = N \\ \sum_{i \in a} w_i \hat{h}_i = \sum_{i=1}^{N} \hat{h}_i \end{cases} \tag{4}$$

where $\hat{h}_i = \hat{h}_i(x_i)$.

Model calibration method is intended to provide good efficiency if the model is correctly specified but maintains desirable properties like design consistency if the model is misspecified (Sahar, 2012), [11]. The simulations in this study suggest that for estimation of the finite population total, p-spline model-based predictive estimators are, in general, more efficient than the HT estimator. In situations that favor the HT estimator, the nonparametric model-based estimators are only slightly less efficient.

2. Fitting of Missing Values

In this section, the study considered fitting missing values for a population divided into clusters which are then subdivided into strata. This section considered a case where there is auxiliary information known at both cluster and cluster element Levels. The study defined $Q = x_1, x_2, \dots, x_U$ as a population of auxiliary variables of size U with auxiliary variable; x_i being known at i^{th} cluster and auxiliary variable; x_{ijk} being known at strata elements level. This study considered a population of clusters G to be partitioned into C clusters, each of size $M_i, i = 1, 2, \dots, C$.

Further, each cluster contains L_i strata each of size $N_j, j = 1, 2, \dots, L_i$. Let also y_{ijk} be k^{th} observation in the sample from the j^{th} stratum of i^{th} cluster and x_i be the corresponding auxiliary variable at cluster level. At stage one, a probability sample c of size m_i of clusters is drawn from C according to a fixed design $P_1(\bullet)$ (by simple random sampling), where $P_1(c)$ is the probability of drawing the sample c of size m_i from C. The first order cluster inclusion probabilities, $P_1(\bullet)$ is $\pi_i = pr(i\varepsilon c) = \sum_{i\varepsilon c} P_1(c)$ and $\pi_{i,t} = pr(i,t\varepsilon c) = \sum_{i,t\varepsilon c} P_1(c)$. The first and the second order probabilities are the probability of including cluster i in the sample and the probability of including clusters i and t in the sample respectively. At stage two, for every sampled cluster $i \in c$, the study chose a sample b_i of elements of size $n_i, i = 1,2,...,c$, where $n_i = n_{i1} + n_{i2} + ... + n_{iL_i}$. Given that $n_{i1}, n_{i2}, ..., n_{iL_i}$ are sample sizes of the sample chosen from L_i strata by proportional allocation with inclusion probabilities $\pi_{k/j/i} = pr(k, j\varepsilon r_i / i\varepsilon c)$ and $\pi_{k,p/j/i} = pr(k, p\varepsilon r_i / i\varepsilon c)$. In this case, the first and second order probabilities are the probability of including element k in the sample b_i of the i^{th} cluster and the probability that unit k and p are both included in the sample b_i respectively. This study let;

$$t_i = h(x_i) + \varepsilon_i;\ i = 1,2,...,C \tag{5}$$

to be the ith cluster total, where $\hat{h}(x_i)$ is a smooth function of x. Let also $\hat{t}_c = \left[\hat{t}_i\right]_{i\varepsilon c}$ be the m_i dimension vector of $\hat{t}_i$'s which is obtained in the sample of clusters.

This study modelled $\hat{h}(x_i)$ in equation (5) by way of penalized spline and performed model calibration on $\hat{h}(x_i)$. Since some information is available at the element level such that for each element in the jth strata of the ith cluster, a nonparametric variable x_{ijk} is available then the first step was to obtain the nonparametric fit for elements; $h_{x_{ijk}}$ using the auxiliary variable x_{ijk} at element level nonparametrically before an estimate of cluster total through stratification for the ith cluster total was found.

Suppose that not all element values of the variable of interest in a given cluster- strata are available and have to be imputed. This study derived a model calibrated estimator of element within cluster as follows;

$$E(y_{ijk}) = \hat{h}(x_{ijk}) \tag{6}$$

where $\hat{h}(x_{ijk})$ and x_{ijk} are defined for every element k in the jth stratum of ith cluster. For simplicity, this study uses $\hat{h}_{ijk}$ for $\hat{h}(x_{ijk})$.

Firstly, when penalized splines are used to fit missing values, the present study defined the nonparametric sample fit for elements within clusters, $E_{\psi_1}(h_{ijk})$ at x_{ijk} due Breidt and Opsomer, (2000) [1] as

$$\hat{h}_{x_{ijk}} = J^T_{psijk} y_{ijk} \tag{7}$$

where J^T_{psijk} is defined by

$$J^T_{psijk} = X_{rijk}(X_{Ob}{}^T W_b X_{Ob} + A\alpha)^{-1} X_{Ob}{}^T W_b \tag{8}$$

In which a matrix X_O is considered with rows

$$X^T_{Oijk} = \left\{1, x_{ijk}, ..., x^q_{ijk}, \left(x_{ijk} - k_1\right)^q_+, ..., \left(x_{ijk} - k_l\right)^q_+\right\} \tag{9}$$

for $i\varepsilon G$, q is the degree of the spline, and the k_l are the knots, while $(x - k_1)_+ = x - k_1$ if $x > k_1$ and 0 if $x \le k_1$. Further, X_{Ob} in equation 8 is the sub matrix of X_0 which consists of the rows X^T_{Oijk} for which the cluster element $k\varepsilon b_i$, $A_\alpha = diag\{0, ..., 0, \alpha, ..., \alpha\}$ with q + 1 zeros on the diagonal followed by l penalty constants α.

The study considered the diagonal matrix of inverse inclusion probabilities as; $W = diag, i \in G\left\{\frac{1}{\pi_{ijk}}\right\}$ and its sample submatrix in equation 8 given as; $W_b = diag, k \in b_i\left\{\frac{1}{\pi_{ijk}}\right\}$. As a result, the model calibrated estimator for the sampled cluster total was defined by;

$$\hat{t}_i = \sum_{i\varepsilon C} w_{ijk}\hat{y}_{ijk}. \tag{10}$$

The optimal weights; w_{ijk} in equation (10) was obtained by the penalty function method. They were obtained by minimizing the chi square distance below as discussed by Deville and Sarndal (1992), [4].

$$\delta_c = \sum_{k\in b_i} \frac{\left(w_{ijk} - z_{ijk}\right)^2}{q_{ijk} z_{ijk}}. \tag{11}$$

subject to

$$\sum_{k\varepsilon b_i} w_{ijk} = M_i$$

and

$$\sum_{k=1}^{n_i} w_{ijk}\hat{h}_{ijk} = \sum_{i=1}^{M_i} \hat{h}_{ijk}. \tag{12}$$

where $z_{ijk} = \frac{1}{\pi_{ijk}}$ is the inverse of inclusion probability and q_{ijk} are some known positive constants, uncorrelated with z_{ijk}. This study followed same optimization procedure by Rao, 1984, [10] and considered an optimization problem of the form

Minimize

$$\delta_c = \sum_{k\in b_i} \frac{\left(w_{ijk} - z_{ijk}\right)^2}{q_{ijk} z_{ijk}} \text{ subject to}$$

$$\begin{cases} l_1(w_c) = \sum_{k=1}^{n_i} w_{ijk} - M_i = 0 \\ l_2(w_c) = \sum_{k=1}^{n_i} w_{ijk}\hat{h}_{ijk} - \sum_{k=1}^{M_i} \hat{h}_{ijk} = 0 \end{cases} . \quad (13)$$

where $\hat{h}_{ijk}$ is a nonparametric fit of the missing value y_{ijk}. The study then constructed an unconstrained problem as follows according to by Rao, 1984, [10].

$$\tau(w_c, r_b, \hat{h}_{ijk}) = \sum_{k\varepsilon b_i} \frac{\left(w_{ijk} - z_{ijk}\right)^2}{q_{ijk} z_{ijk}} + H(r_b)\left(\sum_{k=1}^{n_i} w_{ijk}\hat{h}_{ijk} - \sum_{k=1}^{M_i}\hat{h}_{ijk}\right)^2 + H(r_b)\left(\sum_{k=1}^{n_i} w_{ijk} - M_i\right)^2 . \quad (14)$$

Differentiating equation (14) partially with respect to w_{ijk} we get

$$\tau^1(w_{ijk}, r_b, \hat{h}_{ijk}) = 2\frac{(w_{ijk} - z_{ijk})}{q_{ijk} z_{ijk}} + 2H(r_b)\hat{h}_{ijk}\left(\sum_{k=1}^{n_i} w_{ijk}\hat{h}_{ijk} - \sum_{k=1}^{M_i}\hat{h}_{ijk}\right) + 2H(r_b)\left(\sum_{k=1}^{n_i} w_{ijk} - M_i\right). \quad (15)$$

Equating equation (15) to zero and solving for w_{ijk} the study obtained;

$$w_{ijk} = \frac{z_{ljk} - H(r_b)q_{ijk}z_{ljk}\left(\sum_{\substack{k=1 \\ p\neq k}}^{n_i} w_{ijp}\left[\hat{h}_{ijk}\hat{h}_{ijp} + 1\right] - \sum_{p=1}^{M_i}\left[\hat{h}_{ijk}\hat{h}_{ijp} - 1\right]\right)}{1 + H(r_b)\left(((\hat{h}_{ijk})^2 + 1)q_{ijk}z_{ijk}\right)}. \quad (16)$$

Thus, a nonparametric estimator of the cluster total is given as;

$$\hat{t}_i = \sum_{i\varepsilon c} w_{ijk}\hat{y}_{ijk} = \sum_{k=1}^{n_i} \frac{y_{ijk}z_{ijk}}{1 + H(r_b)\left(((\hat{h}_{ijk})^2 + 1)q_{ijk}z_{jkl}\right)} - \sum_{k=1}^{n_i} \frac{H(r_b)q_{ijk}z_{ljk}y_{ijk}\left(\sum_{\substack{k=1 \\ p\neq k}}^{n_i} w_{ijp}\left[\hat{h}_{ijk}\hat{h}_{ijp} + 1\right] - \sum_{p=1}^{M_i}\left[\hat{h}_{ijk}\hat{h}_{ijp} - 1\right]\right)}{1 + H(r_b)\left(((\hat{h}_{ijk})^2 + 1)q_{ijk}z_{ijk}\right)}. \quad (17)$$

Secondly, the nonparametric fit of the cluster totals based on penalized splines in this study was defined as;

$$\hat{h}_{t_i} = J_{psi}^T \hat{t}_c . \quad (18)$$

where J_{psi}^T is as defined by Nthiwa Janiffer Mwende et al. (2020), [8] as

$$J_{psi}^T = X_{ri}(X_{rc}{}^T W_c X_{rc} + A\alpha)^{-1} X_{rc}{}^T W_c$$

in which a matrix X_r has the rows

$$X_{ri}^T = \left\{1, x_i, \ldots, x_i^q, \left(x_i - k_1\right)_+^q, \ldots, \left(x_i - k_l\right)_+^q\right\} \quad (19)$$

for $i\varepsilon G$, further, X_{rc} is the sub matrix of X_r which consists of the rows X_{ri}^T for which the cluster $i\varepsilon c$, while the diagonal matrix of inverse inclusion probabilities was defined the same way as Nthiwa Janiffer Mwende et al. (2020), [8] as; $W = diag, i \in G\left\{\frac{1}{\pi_i}\right\}$ and its sample sub matrix defined as $W_c = diag, i \in c\left\{\frac{1}{\pi_i}\right\}$.

The study then proposed a nonparametric penalized spline model calibrated population total estimator as;

$$\hat{y}_{PS} = \sum_{i\varepsilon c} w_i \hat{t}_i . \quad (20)$$

The weight w_i in this equation (20) was obtained by minimizing the chi square distance measure given as;

$$\delta = \sum_{i\in c} \frac{\left(w_i - z_i\right)^2}{q_i z_i} \quad (21)$$

subject to

$$\begin{cases} \sum_{i\varepsilon c} w_i = C \ and \\ \sum_{i=1}^{c} w_i \hat{h}_{t_i} = \sum_{i=1}^{C} \hat{h}_{t_i} \end{cases} \tag{22}$$

where $z_i = \dfrac{1}{\pi_i}$ and q_i are some known positive constants uncorrelated with z_i. This, therefore, gave an optimization problem of;
minimize

$$\delta = \sum_{i \in c} \frac{(w_i - z_i)^2}{q_i z_i}$$

subject to

$$\begin{cases} S_1(w) = \sum_{i=1}^{c} w_i - C = 0 \\ S_2(w) = \sum_{i=1}^{c} w_i \hat{h}_{t_i} - \sum_{i=1}^{C} \hat{h}_{t_i} = 0 \end{cases} \tag{23}$$

which gave a penalty function of the form;

$$\tau(w, r_b, \hat{h}_{t_i}) = \sum_{i \in c} \frac{(w_i - z_i)^2}{q_i z_i} + H(r_b)\left(\sum_{i=1}^{c} w_i \hat{h}_{t_i} - \sum_{i=1}^{C} \hat{h}_{t_i} \right)^2 + H(r_b)\left(\sum_{i=1}^{c} w_i - C \right)^2 . \tag{24}$$

Following same penalty procedure as in estimating the optimal weights w_{ijk} by penalty method above, the weight w_i becomes;

$$w_i = \frac{z_i - H(r_b) q_i z_l \left(\sum_{\substack{i=1 \\ j \neq i}}^{c} w_j \left[\hat{h}_{t_i} \hat{h}_{t_j} + 1 \right] - \sum_{j=1}^{C} \left[\hat{h}_{t_i} \hat{h}_{t_j} - 1 \right] \right)}{1 + H(r_b)\left(((\hat{h}_{t_i})^2 + 1) q_i z_l \right)} . \tag{25}$$

The weighted nonparametric estimator of population total based on p-splines when information is available at both cluster and element level as defined in equation (20) was obtained as;

$$\hat{y}_{PSB} = \sum_{i\varepsilon c} w_i \hat{t}_i = \sum_{i \in c} \frac{\hat{t}_i z_i}{1 + H(r_b)\left(((\hat{h}_{t_i})^2 + 1) q_i z_l \right)} - \sum_{i \in c} \frac{H(r_b) q_i z_l \hat{t}_i \left(\sum_{\substack{i=1 \\ j \neq i}}^{c} w_j \left[\hat{h}_{t_i} \hat{h}_{t_j} + 1 \right] - \sum_{j=1}^{C} \left[\hat{h}_{t_i} \hat{h}_{t_j} - 1 \right] \right)}{1 + H(r_b)\left(((\hat{h}_{t_i})^2 + 1) q_i z_l \right)} \tag{26}$$

3. Penalty Function Method of Obtaining the Weights

To obtain the within cluster weights, $W_{ijk} = (k = 1, 2, ..., n_i)$ and cluster level weights $w_i, (i = 1, 2, ..., c)$, this study applied iterative procedure. Firstly, to obtain the within cluster weights, $W_{ijk} = (k = 1, 2, ..., n_i)$ the study solved the penalty function in equation (14) as an unconstrained minimization problem. The research in this case started with some initial guess for w_{ijk} and r_b then iteratively improved on the initial values until optimal values are obtained. The present study, therefore, followed the Newton method of unconstrained optimization, according to [8] as follows;

If $W_i = (w_{ij1}, w_{ij2}, ..., w_{ijn_i})$ is let to be the set of the weights, of interest was to obtain W_i^* such that

$$\Gamma(W_i^*) = \left[\tau'(w_{ij1}, r_b), ..., \tau'(w_{ijn_i}, r_b,) \right]' = 0 . \tag{27}$$

Further if W_{iu} is let to be initial estimate of W_i^* so that $W_i^* = W_{iu} + T_i$. The Taylor's series expansion of $\Gamma(W_i^*)$ gives

$$\Gamma(W_i^*) = \Gamma(W_{iu} + T_i) = \Gamma(W_i) + J_{W_{iu}} T_i + \ldots . \tag{28}$$

By neglecting the higher-order terms in the above equation (28) and setting $\Gamma(W^*) = 0$ the study had

$$\Gamma(w_{iu}) + J_{w_{iu}} T_i = 0 \tag{29}$$

where $J_{w_{iu}}$ is a n_i by n_i matrix of second derivatives of the penalty function equation (15) evaluated at W_{iu}. Let also i and j be the row and column counters respectively with $i = (1, 2, ..., n_i)$ rows with $j = (1, 2, ..., n_i)$ columns. The matrix

$J_{w_i u}$ has elements

$\frac{2}{q_{ijk} z_{ijk}} + 2H(r_b)\left((\hat{h}_{ijk})^2 + 1\right)$ in the main diagonal and elements $2H(r_b)\left(\hat{h}_{ijk}\hat{h}_{ijp} + 1\right)$ elsewhere.

If $J_{w_{ui}}$ is a nonsingular matrix, then, from the set of linear equations (29) we have for vector T

$$T_i = J^{-1}{}_{w_{ui}} \Gamma(w_{iu}) . \quad (30)$$

The following iterative procedure is used to find the improved approximations of W_i^*

$$W_{i+1} = W_i + T_i = W_i - J^{-1}{}_{w_i} \Gamma(W_i) . \quad (31)$$

The sequence of the points $W_{ij1}, W_{ij2}, ..., W_{ij(u+1)}$ eventually converges to the actual solution W_i^* . Now, if we let $W_{ib}{}^*$ be the minimum of W_i^* obtained for a particular penalty, r_b the study obtained a sequence of minimum points $W_{ij1}{}^*, W_{ij2}^*, ..., W_{ij(b+1)}^*$ for the penalties $r_1, r_2 ..., r_{b+1}$ until $W_{ib}{}^* = W_{ij(b+1)}^*$ or $\tau(w_{ijk}, r_b, \hat{h}_{ijk}) = \tau(w_{ijk}, r_{b+1}, \hat{h}_{ijk})$ for some specified accuracy level. The accuracy level may for example be, to certain decimal points or significance level. In addition, the penalty values may be set such that the starting point $r_1 > 0$ and $r_{b+1} = sr_b$, where $s < 1$, $H(r_b) \to \infty$ as $r_b \to 0$.

This study applied iterative procedure to obtain the cluster level weights $w_i, (i = 1, 2, ..., c)$. The penalty function in equation (24) was solved as an unconstrained minimization problem in a similar manner as within cluster weights; W_{ijk} discussed in this section.

4. Empirical Analysis and Discussions

In the simulation study, a population of size 10,000 (200*50=10,000) was simulated from a population structure containing 200 clusters each of size 50. Each cluster had 5 strata of size 10 each. At stage one 10, 20, 30,…, 190 clusters were sampled from the 200 clusters by simple random sampling while at stage two, 5 elements where drawn from each stratum by proportional allocation. This gave sample of 25 elements from each of the sampled clusters. 10 replications per each sample size were generated. For penalized spline method, the number of knots and the Spline penalty were optimally generated.

Using R program, a population of independent and identically distributed variable x was simulated using uniform (0, 1). This study used penalized spline equation (18) to fit cluster totals and equation (7) to fit for elements within a cluster. Using the auxiliary variable x, five populations for the dependent random variable for cluster element, and cluster total; t_i were generated with the five functions as:

Linear function (Lin); $t_i = 3 + 6x$

Quadratic function (Qd); $t_i = (20 + 6x)^2$

Exponential function (Exp); $t_i = 20 - \exp(x/40)$

Cycle 4 function (C4); $t_i = 0.5 - \sin(8\pi x)$

Cycle 2 function (C2); $t_i = 0.5 - \sin(2\pi x)$

where t_i is the ith cluster total and x_i is cluster auxiliary variable known at the cluster level. The s k^{th} unit in j^{th} stratum of i^{th} cluster was given by;

$$y_{ijk} = \frac{t_i}{cluster\ size} + error\ term(e_i)/\sqrt{cluster\ size}. \quad (32)$$

This study differentiated the five strata from each other by the following errors;

$e_1 = runif(-0.001, +0.001)$ for stratum 1, $e_2 = runif(-0.002, +0.002)$ for stratum 2, $e_3 = runif(-0.003, +0.003)$ for stratum 3, $e_4 = runif(-0.004, +0.004)$ for stratum 4 and $e_5 = runif(-0.005, +0.005)$ for stratum 5.

On the other hand, the respective kth auxiliary variable in the jth stratum of ith cluster was given as;

$$x_{ijk} = \frac{(3 - y_{ijk})}{6} + e_j \text{ for linear} \quad (33)$$

$$x_{ijk} = \frac{20 - \sqrt{y_{ijk}}}{6} + e_j \text{ for quadratic} \quad (34)$$

$$x_{ijk} = (\log(20 - y_{ijk}))40 + e_j \text{ for exponential} \quad (35)$$

$$x_{ijk} = \frac{\sin^{-1}(y_{ijk} - 0.5)}{8} + e_j \text{ for cycle 4} \quad (36)$$

$$x_{ijk} = \frac{\sin^{-1}(y_{ijk} - 0.5)}{2} + e_j \text{ for cycle 2.} \quad (37)$$

This study reports on the performance penalized spline model calibrated estimator and its efficiency in comparison with Horvitz Thompson estimator. The performance of the nonparametric estimator; $\hat{y}_{PS}$ was evaluated using its relative bias R_B and relative efficiency R_E. The relative bias was defined as

$$R_B = \frac{\sum_{r=1}^{R} (y_{PS} - Y_{AT})}{R * Y_{AT}} \quad (38)$$

where, Y_{AT} is the actual total and R is the replicate number of samples. The relative efficiency was defined as

$$R_E = \frac{MSE(y_{PS})}{MSE(y_{HT})}. \quad (39)$$

Large values of relative efficiencies represent higher efficiency for the design estimator y_{HT} over the estimator $\hat{y}_{PS}$ and vice versa. The $\hat{y}_{HT}$ estimator was defined as $\hat{y}_{HT} = \sum z_i t_i$ where z_i is the inverse of the inclusion probability given by $z_i = \frac{\mathrm{C}}{\mathrm{c}}$ for class total and $z_{ijk} = \frac{\text{cluster size}}{\text{sample size}}$ for cluster element. The estimator $\hat{y}_{HT}$ was used as the baseline comparison.

4.1. Normality Test

A One-sample Kolmogorov-Smirnov test was carried out in this study to test for normality of the nonparametric estimators; $\hat{y}_{PS}$ and the design estimator; Horvitz Thompson estimator; $\hat{y}_{HT}$. The p values at α = 0.05 for the five population mean functions obtained are as in table 1 below. A p-value greater than the set α = 0.05 significance level means normality is established. The results show that at α = 0.05 the proposed estimators are normal for all the five functions.

Table 1. *Normality test.*

Estimator	$\hat{y}_{PS}$	$\hat{y}_{HT}$
Linear	0.9928	0.8869
Quadratic	0.3763	0.3934
Exponential	0.7678	0.6876
Cycle 4	0.8850	0.5584
Cycle 2	0.9920	0.4319

4.2. Results for Population Total Estimates

The results in tables 2, 3, 4, 5 and 6 below shows the actual total and the estimates of the penalized splines and the Horvitz Thompson for the respective mean functions with sample sizes; 50, 100 and 150. From the results the estimator $\hat{y}_{PS}$ are seen to give estimates that are close the actual total and also to those of Horvitz Thompson design estimator for all the 5 population functions.

Table 2. *Linear Population total estimates for samples of sizes, 20, 50, 100 and 150.*

Replication Number		1	2	3	4	5	6	7	8	9	10
y_{AT}	Sample size 50/100/150	4543.206	4543.206	4543.206	4543.206	4543.206	4543.206	4543.206	4543.206	4543.206	4543.206
$\hat{y}_{PS}$	50	4550.709	4508.628	4568.743	4598.951	4526.772	4580.126	4506.070	4555.463	4486.108	4551.931
	100	4525.000	4568.034	4568.753	4515.870	4568.810	4531.189	4563.055	4615.942	4512.589	4568.554
	150	4537.296	4537.509	4522.132	4561.317	4515.272	4546.944	4528.760	4560.889	4535.451	4521.857
$\hat{y}_{HT}$	50	4543.147	4505.097	4559.542	4598.436	4531.062	4576.261	4503.652	4551.989	4488.678	4544.956
	100	4522.916	4566.846	4561.118	4516.681	4561.496	4532.834	4568.969	4620.047	4520.402	4565.650
	150	4536.349	4539.143	4518.782	4563.124	4516.328	4540.459	4526.952	4556.363	4532.178	4525.400

Table 3. *Quadratic Population total estimates for samples of sizes, 20, 50, 100 and 150.*

Replication Number		1	2	3	4	5	6	7	8	9	10
y_{AT}	Sample size 50/100/150	58508.64	58508.64	58508.64	58508.64	58508.64	58508.64	58508.64	58508.64	58508.64	58508.64
$\hat{y}_{PS}$	50	59832.17	58626.97	55916.10	59068.46	60239.68	61313.31	56243.05	56767.31	58457.40	55972.48
	100	59208.13	59121.51	58549.64	58667.37	58077.49	58327.13	58899.17	58961.40	59057.99	58351.96
	150	58930.17	58508.48	58202.29	58577.28	58844.61	58348.84	58376.42	58062.99	57828.53	58361.61
$\hat{y}_{HT}$	50	59835.06	58622.55	55912.85	59073.66	60238.51	61314.53	56252.56	56768.40	58452.97	55964.32
	100	59208.76	59120.87	58549.63	58665.68	58075.85	58331.41	58893.50	58961.65	59061.60	58348.58
	150	58932.61	58505.37	58205.57	58576.88	58851.44	58347.17	58379.05	58063.35	57824.62	58364.26

Table 4. *Exponential Population total estimates for samples of sizes, 20, 50, 100 and 150.*

Replication Number		1	2	3	4	5	6	7	8	9	10
y_{AT}	Sample size 50/100/150	3794.735	3794.735	3794.735	3794.735	3794.735	3794.735	3794.735	3794.735	3794.735	3794.735
$\hat{y}_{PS}$	50	3799.913	3800.427	3800.963	3792.627	3798.025	3795.023	3787.908	3790.535	3807.945	3783.372
	100	3790.318	3794.567	3790.789	3791.142	3794.270	3795.132	3793.077	3795.497	3797.233	3794.940
	150	3794.733	3793.256	3790.879	3795.593	3790.908	3796.495	3799.995	3792.718	3794.792	3795.983
$\hat{y}_{HT}$	50	3800.172	3800.242	3800.665	3792.678	3797.854	3795.050	3787.943	3790.483	3807.787	3783.129
	100	3790.028	3794.767	3790.476	3791.191	3794.124	3795.143	3793.297	3795.539	3797.097	3794.861
	150	3794.631	3793.194	3790.718	3795.681	3790.801	3796.415	3800.001	3792.819	3794.634	3795.881

Table 5. *Cycle 4 population total estimates for samples of sizes, 20, 50, 100 and 150.*

Replication Number		1	2	3	4	5
y_{AT}	Sample size					
	50/100/150	98.12195	98.12195	98.12195	98.12195	98.12195
	50	106.74650	132.50628	119.94744	89.20880	85.34613
$\hat{y}_{PS}$	100	96.41683	90.02801	97.57377	100.66343	94.51954
	150	99.74723	90.42254	93.29842	102.81139	104.91942
	50	113.48154	131.49654	125.62301	95.99730	79.22657
$\hat{y}_{HT}$	100	94.70033	92.00542	99.25881	102.30048	94.31615
	150	97.77412	91.93835	93.09909	101.91398	102.26683

Table 5. *Continued.*

Replication Number		6	7	8	9	10
y_{AT}	Sample size					
	50/100/150	98.12195	98.12195	98.12195	98.12195	98.12195
	50	95.97287	69.00133	137.85757	91.38317	94.34510
$\hat{y}_{PS}$	100	109.77530	85.37745	96.69058	104.97689	93.86149
	150	97.10018	98.32416	87.04543	92.83903	98.93072
	50	92.69908	68.07357	142.53958	92.80498	88.05275
$\hat{y}_{HT}$	100	110.22664	86.45460	95.15238	102.69284	93.62301
	150	99.12175	100.96804	86.27175	95.20490	100.40792

Table 6. *Cycle 2 population total estimates for samples of sizes, 20, 50, 100 and 150.*

Replication Number		1	2	3	4	5	6	7	8	9	10
y_{AT}	Sample size										
	50/100/150	116.4859	116.4859	116.4859	116.4859	116.4859	116.4859	116.4859	116.4859	116.4859	116.4859
	50	92.01059	104.7750	113.7534	99.91309	84.01602	129.1396	141.9112	132.7674	123.5790	154.1871
$\hat{y}_{PS}$	100	114.2216	114.2180	107.6319	121.8142	131.3348	121.3722	115.1419	100.68917	122.5620	112.1085
	150	114.7727	118.6039	120.1997	110.1363	112.3717	123.1632	122.7303	111.1371	120.6737	111.0792
	50	92.26469	104.9711	110.4062	100.62389	85.96008	138.7731	142.7816	116.2985	119.2736	148.2210
$\hat{y}_{HT}$	100	109.9282	118.0110	111.7657	117.6613	129.6345	118.8748	116.8948	99.04232	114.9889	109.5154
	150	115.2668	116.3559	114.0293	110.9012	114.7852	123.9057	124.0244	112.8755	121.4621	110.0413

4.3. Results of Variances and Variance Ratios for Various Sample Size

This section presents both the variance and variance ratio of the two estimators; $\hat{y}_{PS}$ based on a penalized spline and $\hat{y}_{HT}$ based on Horvitz Thompson and their respective graphs. The variance and variance ratios for different functions are summarized in table 7, and their comparative graphs in figures 1, 2, 3, 4 and 5 for the respective functions.

4.3.1. Tabular Results of Variances and Variance Ratios for Various Sample Size

The table 7 below shows both variance and variance ratios of the two estimators based on penalized splines and Horvitz. The variances of the two estimators for the 5 functions decrease as the sample size increases implying that the estimators are consistent. From the table, the $\mathrm{var}(\hat{y}_{HT})$ for linear and exponential functions are smaller than those of $\mathrm{var}(\hat{y}_{PS})$ except for the 9 sample sizes; 20, 60, 70, 90, 110, 120, 130, 140 and 160 and 8 sample sizes 40, 70, 90, 100, 110, 130, 150 and 180 respectively. This comparison is evident from the variance ratio; $\mathrm{var}\left(\hat{y}_{PS}/\hat{y}_{HT}\right)$ for the same functions. For quadratic function the $\mathrm{var}(\hat{y}_{PS})$ estimator is less variant than Horvitz Thompson estimator since $\mathrm{var}(\hat{y}_{PS})$ is consistently lower than that of Thompson estimator for all samples sizes except for 5 sample sizes; 30, 70, 140, 170 and 180 and this applies to the variance ratio; $\mathrm{var}\left(\hat{y}_{PS}/\hat{y}_{HT}\right)$ for the same function. For cycle 4 and cycle 2 the $\mathrm{var}(\hat{y}_{HT})$ is slightly less variant than that of $\mathrm{var}(\hat{y}_{PS})$ in some samples except for the 7 sample sizes; 50, 60, 110, 120, 140, 160 and 180 and the 6 sample sizes; 20, 40, 60, 130, 140 and 170 for cycle 2 and all these is evident from the respective ratios; $\mathrm{var}\left(\hat{y}_{PS}/\hat{y}_{HT}\right)$.

Table 7. *Results of Variances and variance ratios for various sample size.*

Sample size		10	20	30	40	50
	Lin	3651.667479	3951.1194439	3543.5359584	3337.6837219	1278.906399
	Qd	5,331,024	3,522,390	4,482,455	972,450.3	3,728,991
$\text{var}(\hat{y}_{PS})$	Exp	168.278885	169.0653534	51.0707619	47.3712646	52.4873479
	C4	2904.529547	978.5458207	505.7774565	557.0358257	477.5578929
	C2	835.9659326	1292.6891074	190.0423932	221.8191635	511.4902605
	Lin	3441.982023	4170.5863917	3293.8879518	3320.0194090	1171.349553
	Qd	5,351,925	3,529,463	4,477,180	973,958.5	3,731,934
$\text{var}(\hat{y}_{HT})$	Exp	163.048127	163.9132225	50.3517091	48.5950529	52.3219643
	C4	2609.500036	960.9689888	382.5380028	546.3226956	584.9590572
	C2	747.9882971	1358.9711656	169.4059765	231.5295217	459.9957258
	Lin	1.060920	0.9473774	1.0757913	1.0053205	1.091823
	Qd	0.9960946	0.9979960	1.001178	0.9984514	0.9.992114
$\text{var}(\hat{y}_{PS})/\text{var}(\hat{y}_{HT})$	Exp	1.032081	1.0314321	1.0142806	0.9748166	1.0031609
	C4	1.113060	1.0182907	1.3221626	1.0196095	0.8163954
	C2	1.1176190	0.9512263	1.1218163	0.9580600	1.1119457

Sample size		60	70	80	90	100
	Lin	1686.3531718	683.4909103	739.351333	621.8465705	1032.696102
	Qd	1,432,228	795,246.2	1,100,148	239,576.3	148,716.9
$\text{var}(\hat{y}_{PS})$	Exp	37.5194780	10.8705338	12.2096085	17.0080090	5.2460144
	C4	221.6688694	288.1115958	114.015931	84.9656554	48.9687456
	C2	235.0572714	178.749618	364.9281303	334.5377284	74.1583869
	lin	1798.2234675	721.2150960	730.283623	640.0131115	992.073248
	Qd	1,434661	793,026.8	1,101,905	239,976.3	148,954.9
$\text{var}(\hat{y}_{HT})$	Exp	36.9130210	11.5590297	12.1355654	17.2458368	5.5536149
	C4	222.6924910	285.6353110	107.396674	79.9022833	44.8840996
	C2	265.1027829	147.086222	338.3337506	332.5792995	63.1641415
	Lin	0.9377884	0.9476936	1.012417	0.9716154	1.040947
	Qd	0.9983042	1.002799	0.9984059	0.9983334	0.9984022
$\text{var}(\hat{y}_{PS})/\text{var}(\hat{y}_{HT})$	Exp	1.0164293	0.9404365	1.0061013	0.9862096	0.9446126
	C4	0.9954034	1.0086694	1.061634	1.0633696	1.0910043
	C2	0.8866647	1.215271	1.0786040	1.0058886	1.1740583

Sample size		110	120	130	140	150
	Lin	528.8463282	391.1319758	367.6803949	366.0277604	250.3687311
	Qd	552,643.4	566,478.0	397,977.0	205,447.0	111,742.5
$\text{var}(\hat{y}_{PS})$	Exp	4.0804940	12.6728286	8.8227916	4.0807057	7.6035401
	C4	53.0041054	62.3777231	49.178826	45.8120890	31.2021121
	C2	51.9719975	102.0590988	39.9900411	78.6794029	26.379535
	Lin	563.9793390	418.7877861	414.6097232	377.8959278	230.0089091
	Qd	555,015.1	569,733.0	398,328.6	205,206.4	112,9230
$\text{var}(\hat{y}_{HT})$	Exp	4.3935972	12.3957597	8.9766109 9	3.820459	7.7526387
	C4	73.9513379	65.9388513	35.895394	51.8420740	27.0439387
	C2	43.2446770	89.7462109	41.2334476	82.1997916	25.854080
	Lin	0.9377051	0.9339622	0.8868108	0.9685941	1.0885175
	Qd	0.9957268	0.9942867	0.9991173	1.001173	0.9895455
$\text{var}(\hat{y}_{PS})/\text{var}(\hat{y}_{HT})$	Exp	0.9287365	1.0223519	0.9828644	1.0681190	0.9807680
	C4	0.7167430	0.9459935	1.370060	0.8836855	1.1537562
	C2	1.2018126	1.1371967	0.9698447	0.9571728	1.020324

Sample size		160	170	180	190
	Lin	157.6309214	179.052477	65.0823883	28.0041116
	Qd	253,692.2	47,602.45	76,423.97	32,023.10
$\text{var}(\hat{y}_{PS})$	Exp	3.657975	3.4205472	2.3161035	1.922969
	C4	28.7065251	11.8576488	17.4000800	7.9386513
	C2	25.326064	23.6213578	11.009247	7.680566
	Lin	168.0263317	136.965407	61.7790150	21.0822213
	Qd	254,282.6	47,307.85	76,267.43	32262.74
$\text{var}(\hat{y}_{HT})$	Exp	3.602958	3.3717028	2.4699470	1.787780
	C4	31.0215984	8.1252690	18.8789233	4.2808000
	C2	14.801159	27.5178092	5.452016	2.981572

Sample size		160	170	180	190
$\text{var}(\hat{y}_{PS}) / \text{var}(\hat{y}_{HT})$	Lin	0.9381323	1.307282	1.0534708	1.3283283
	Qd	0.9976779	1.006227	1.002053	0.9925723
	Exp	1.015270	1.0144866	0.9377138	1.075619
	C4	0.9253722	1.4593546	0.9216670	1.8544784
	C2	1.711087	0.8584026	2.019298	2.576013

4.3.2. Graphical Results of Variances and Variance Ratio for Various Sample Size

Results in figures 1, 2, 3, 4, and 5 represent respectively, the graphical variance ratio for the five population functions; linear, quadratic, exponential, Cycle 4 and Cycle 2 respectively for the two estimators based on penalized splines and Horvitz Thompson. Results from the five figures show that the ratio $\text{var}\left(\hat{y}_{PS} / \hat{y}_{HT}\right)$ was concentrated below one for quadratic function, implying that Thompson estimator is more variant than $\text{var}(\hat{y}_{PS})$ in quadratic. On the other side Horvitz Thompson estimator is less variant than $\text{var}(\hat{y}_{PS})$ in linear, exponential cycle 4 and cycle 2 functions because their ratios concentrate slightly above one.

4.4. Relative Bias

The following table 8 shows the values of relative biases for the five population functions. The results from the table show that the relative biases for the two estimates are minimal, given that the population totals were in thousands, and this point to unbiasedness. It is also noted that the difference of the relative biases of the penalized spline estimator with its corresponding Horvitz Thompson estimators for all the five population functions is not very significant. Hence, the proposed model calibrated estimator is unbiased.

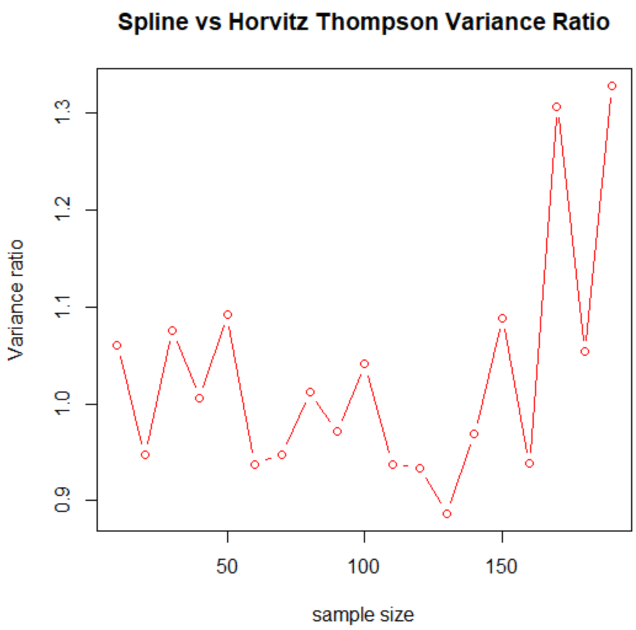

Figure 1. *Linear.*

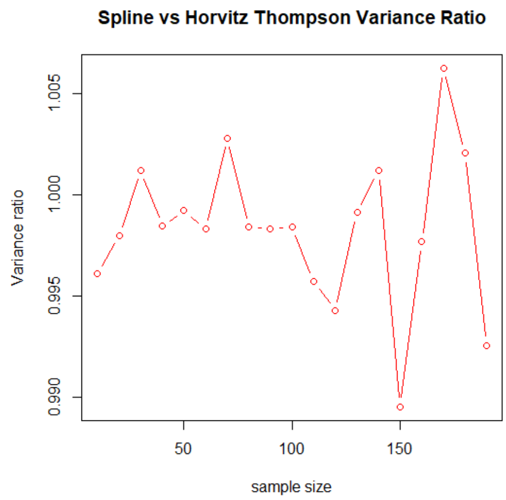

Figure 2. *Quadratic.*

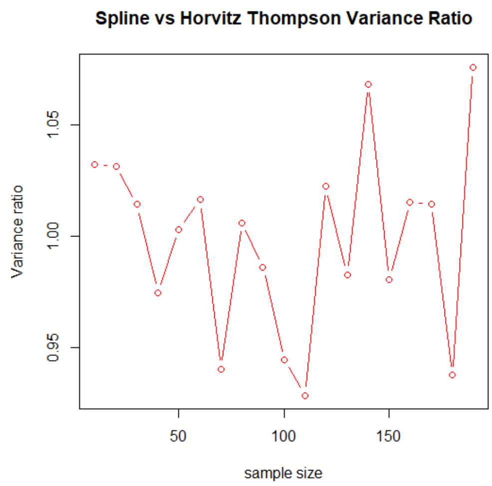

Figure 3. *Exponential.*

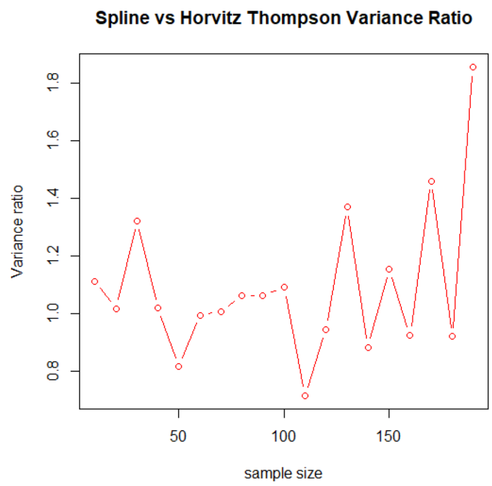

Figure 4. *Cycle 4.*

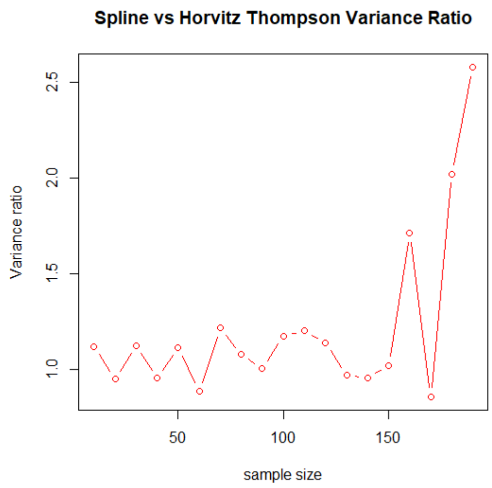

Figure 5. *Cycle 2.*

Table 8. *Relative Biases for three estimators.*

Estimator	$\hat{y}_{PS}$	$\hat{y}_{HT}$
Linear	-0.0006773947	-0.0006504365
Quadratic	0.002472271	0.002474501
Exponential	0.00005334967	0.00005276563
Cycle 4	-0.0004616429	-0.0007526457
Cycle 2	0.01028577	0.00724342

4.5. Results on Relative Efficiency for Various Sample Sizes

4.5.1. Tabular Results for MSE and MSE Ratios for Various Sample Size

The MSE and relative efficiency (MSE Ratios) for the five different functions are summarized in table 9. Generally, the estimator with a smaller MSE is regarded as the most efficient one. From table 9 MSEs of $\hat{y}_{PS}$ is smaller than that of $\hat{y}_{HT}$ in some samples sizes but from other sample sizes the $\hat{y}_{HT}$ estimator has reduced MSE than the estimator; $\hat{y}_{PS}$. The relative efficiencies (MSE Ratios) in table 9 examines the robustness of the various population functions. There doesn't appear to be a clear noticeable performance difference of $\hat{y}_{PS}$ estimator in comparison to $\hat{y}_{HT}$ estimator. In some instances $\hat{y}_{PS}$ has a smaller error margin than $\hat{y}_{HT}$, while in other samples, $\hat{y}_{HT}$ has smaller error margins than the nonparametric model calibrated estimator; $\hat{y}_{PS}$. The $\hat{y}_{PS}$ estimator is more efficient than $\hat{y}_{HT}$ estimator in all the samples except for 5 sample sizes; 30, 70, 140, 170 and 180 for quadratic and 4 sample sizes; 30, 40, 110 and 140 for cycle 4. In addition, $\hat{y}_{PS}$ estimator is slightly more efficient than $\hat{y}_{HT}$ estimator for exponential functions and cycle 2, except for 7 sample sizes; 10, 20, 70, 80, 120, 130 and 180 and 8 sample sizes; 30, 50, 60, 80, 90 110, 120 and 150 respectively. On the other hand the estimator; $\hat{y}_{HT}$ is slightly more efficient than $\hat{y}_{PS}$ estimator in all the samples except for the 9 sample sizes; 20, 40, 60, 110, 120, 130, 150, 180 and 190 for linear function.

Table 9. *Results of MSE and MSE ratios for various sample size.*

Sample size		10	20	30	40	50
MSE $\hat{y}_{PS}$	Lin	3761.901218	3997.1178405	3599.576631	3340.0779186	1278.927001
	Qd	7,446,041	4,738,647	4,550,979	984,679.2	3,799,186
	Exp	170.6475187	169.410735	52.5931846	48.8811167	53.3688749
	C4	2943.9179205	1007.5783450	568.007902	583.4024883	494.4464803
	C2	1195.9670865	1530.1841661	271.0469163	227.3060849	512.7431068
MSE $\hat{y}_{HT}$	Lin	3529.915102	4219.3669187	3345.352957	3325.2848170	1179.902056
	Qd	7,485,276	4,744,751	4,547,491	986,394.6	3,802,210
	Exp	167.6882442	159.122957	57.2071173	51.7141427	56.1322306
	C4	2903.0705831	1142.1461835	445.271374	468.4352616	641.6641625
	C2	1306.1443127	1686.9108741	202.1827495	296.3214559	414.9023874
MSE $\hat{y}_{PS}$ /MSE $\hat{y}_{HT}$	Lin	1.010684	0.9841703	1.000111	0.9907866	1.024582
	Qd	0.9947583	0.9987136	1.000767	0.9982609	0.9992047
	Exp	1.0176475	1.064653	0.9193469	0.9452176	0.9507706
	C4	0.9440421	0.8657855	1.015070	1.1705200	0.9487043
	C2	0.9441625	0.9612389	1.0125236	0.8432357	1.1093576

Sample size		60	70	80	90	100
MSE $\hat{y}_{PS}$	Lin	2213.3671469	868.6904893	941.9650729	625.4252385	1144.486644
	Qd	1,444,626	824,239.5	1,100,902	239,875.2	194,317.5
	Exp	38.555204	11.8924768	12.425321	17.0663297	6.3248485
	C4	241.8879912	288.1286386	115.2612613	101.4898390	50.2538319
	C2	285.2328794	199.2620918	396.7704457	335.674333	74.3001555
MSE $\hat{y}_{HT}$	Lin	2293.3339425	922.9854323	981.6586883	640.0297615	1102.101586
	Qd	1,447,216	822,176.1	1,102,720	240,235.0	194,374.0
	Exp	42.740574	11.8107904	11.150557	17.9337831	6.9595295
	C4	346.3966814	309.8198703	108.1009965	115.8467571	55.8920469
	C2	337.3473346	188.3122440	356.2083152	273.639732	76.4726920
MSE $\hat{y}_{PS}$ /MSE $\hat{y}_{HT}$	Lin	0.9875441	1.0345659	1.0248652	1.0786348	1.003822
	Qd	0.9982109	1.002510	0.9983508	0.9985023	0.9997094
	Exp	0.902075	1.0069163	1.114323	0.9516302	0.9088040
	C4	0.7290980	0.9229237	0.9957634	0.7655590	0.8227334
	C2	1.0160905	0.8744224	1.0139206	1.217429	0.8709300

Sample size	110	120	130	140	150

MSE $\hat{y}_{PS}$	Lin	655.8482220	477.6281396	367.6808541	379.2494284	292.1500391
	Qd	607,022.7	593,307.8	415,432.1	267,156.0	122,666.1
	Exp	4.1202120	13.494986	9.7373345	6.7647846	7.6434632
	C4	83.495772	77.3385387	50.1603724	50.3699637	33.6924933
	C2	52.271479	104.7423976	40.0147599	79.4145996	26.3795354
MSE $\hat{y}_{HT}$	Lin	700.5660838	545.2801802	415.5757672	382.8570384	289.2821260
	Qd	608,585.2	596,561.5	415,988.3	267,093.0	123,657.2
	Exp	4.4396357	13.214463	9.8181521	6.5505387	7.8190212
	C4	104.381583	91.3097778	39.0132544	51.7093668	40.4558555
	C2	39.431819	89.9134820	41.9892032	90.0818898	22.8535505
MSE $\hat{y}_{PS}$ /MSE $\hat{y}_{HT}$	Lin	0.9538936	0.9867949	0.9313259	1.0079811	0.9156409
	Qd	0.9974326	0.9945459	0.9986628	1.000236	0.9919849
	Exp	0.9107537	1.057114	1.0250755	0.9591387	0.9498795
	C4	1.106025	0.8855587	0.9287332	1.0412362	0.7055897
	C2	1.102070	1.0183366	0.9902773	0.9336432	1.1319365

Sample size		160	170	180	190
MSE $\hat{y}_{PS}$	Lin	165.6221865	179.918580	65.9200680	31.1640058
	Qd	253,768.2	62,777.18	76,511.68	36,341.06
	Exp	3.9267096	3.4288437	2.8152284	3.7338132
	C4	47.4667990	15.1352895	23.7547279	8.0433500
	C2	32.9937376	23.6264682	14.1782831	9.0175933
MSE $\hat{y}_{HT}$	Lin	179.0923562	137.646389	65.1613352	24.2407817
	Qd	254,336.6	62,561.61	76,335.32	36,612.13
	Exp	3.8768434	3.3762876	3.0369193	3.7901165
	C4	47.4120481	14.0677071	31.5174763	13.4859242
	C2	20.8959407	28.3225165	8.3965927	7.2576928
MSE $\hat{y}_{PS}$ /MSE $\hat{y}_{HT}$	Lin	1.0137243	1.050970	0.9480256	0.9487241
	Qd	0.9977651	1.003446	1.002310	9.925960
	Exp	0.9807432	0.9651552	1.0343102	0.9379330
	C4	0.9488392	0.7139708	0.7358081	0.3177479
	C2	0.8650733	0.9779437	0.6767941	0.4360844

4.5.2. MSE and MSE Ratio Graphs for Various Sample Sizes

Figure 6 up to figure 10 show graphical representation of relative efficiency for the respective population functions. The MSE ratios for the five functions are seen to mostly concentrated at a point slights below one. This implies that the nonparametric estimator; $\hat{y}_{PS}$ is more efficient than $\hat{y}_{HT}$ and that it's a robust estimator.

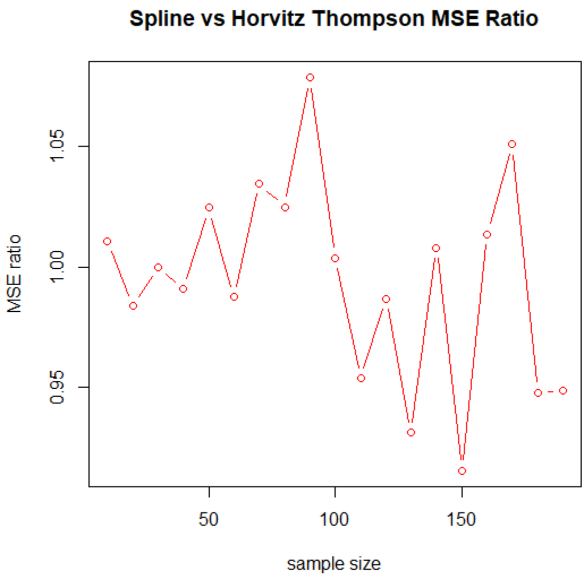

Figure 6. *Linear.*

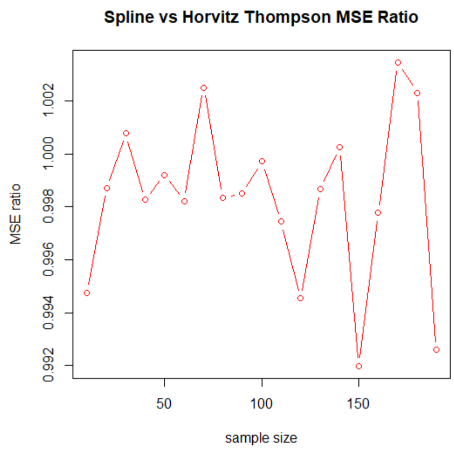

Figure 7. *Quadratic.*

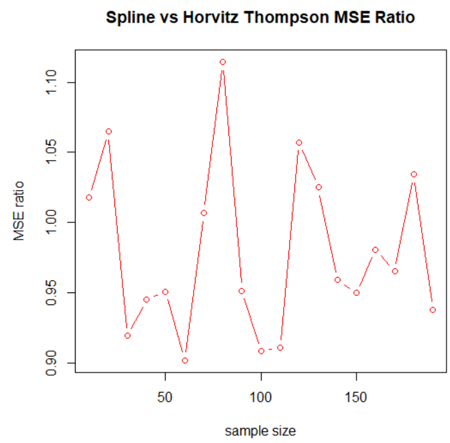

Figure 8. *Exponential.*

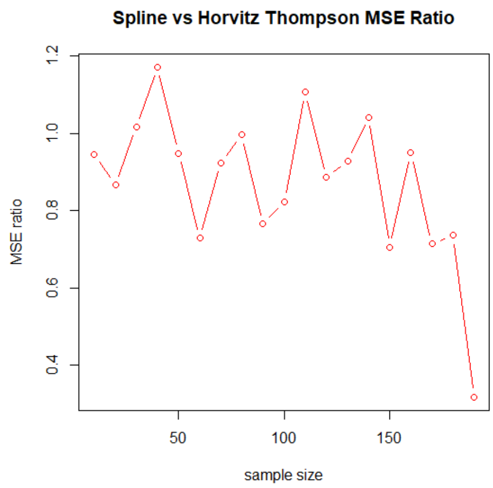

Figure 9. *Cycle 4.*

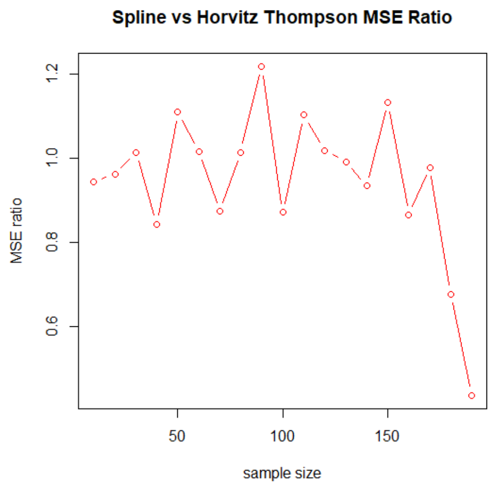

Figure 10. *Cycle 2.*

5. Conclusion

The results from this study show that, the nonparametric calibrated estimator; $\hat{y}_{PS}$ is slightly more efficient than $\hat{y}_{HT}$. The estimator; $\hat{y}_{PS}$ is also very robust because it is not failing for all the five functions compared with the design estimator $\hat{y}_{HT}$ which is known as a well-performing estimator. Generally, the results have also shown that the performance of the nonparametric estimator $\hat{y}_{PS}$ is indistinguishable from that of the design estimator in some instances, and that $\hat{y}_{PS}$ is a normal, unbiased and consistent estimator. Therefore, this study concludes that the nonparametric model calibrated estimators; it is a robust estimator since it does not fail under misspecification.

Due to these good properties of nonparametric model calibrated estimators, this study, therefore, recommends the use of such model calibrated estimators in the estimation of population total in sampling. In the present real-world problem where there are missing variables at both cluster and cluster element levels, yet there is relevant auxiliary information about the variables, model calibrated estimators would be the estimators of choice. This study has further shown that in cases where some cluster and elements within clusters are missing but auxiliary information is available at both levels, then an advantage can be taken of this auxiliary information to fit both cluster totals and cluster elements, which are then calibrated in the estimation of population total. This provides the researcher with a normal, consistent, unbiased, robust and efficient estimator of population total.

References

[1] Breidt, F. J. and Opsomer, J. D. (2000). Local Polynomial Regression Estimation in Survey Sampling. *Annals of Statistics*, 28: 1026-1053.

[2] Clair, l. (2016). Nonparametric kernel estimation methods using Complex survey data, PhD thesis, mcmaster university, Main St. West, Hamilton Ontario.

[3] De Boor, C. (2001). A Practical Guide To Splines (Revised Edition). *Springer, New York*.

[4] Deville, J. C. and Sarndal C. E. (1992). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association,* 87: 376-382.

[5] Eilers, P. H. C. and Marx, B. D. (1996). Flexible Smoothing with B-Splines and Penalties (with discussion). *Statistical Science,* 11: 89-121.

[6] Eubank, R. L. (1988). Spline smoothing and Nonparametric regression. New york and Basel: *Marcel Dekker.*

[7] Horvitz, D. G and Thompson, D. J. (1952). A Generalization of sampling without Replacement from Finite Universe. *Journal of American Statistical Association*, 47: 663-685.

[8] Nthiwa Janiffer Mwende *et al.* (2020). Population Total Estimation in a Complex Survey by Nonparametric Model Calibration Using Penalty Function Method with Auxiliary Information Known at Cluster Levels. *American Journal of Theoretical and Applied Statistics*. 4: 162-172.

[9] Pinheiro, J. C. and Bates, D. M (2000). Mixed-effects models in S and S-PLUS, Springer: New York.

[10] Rao, S. S. (1984). Optimization Theory and Applications. *Wiley Eastern Limited.*

[11] Sahar, Z. Z. (2012). Model-based methods for robust finite population inference in the presence of external information. *The University of Michigan.*

[12] Sayed, A. M. (2010). Nonparametric kernel density estimation using auxiliary information from complex survey data. Masters thesis.

[13] Wahba, G. (1990). A comparison of GCV and GLM for Choosing the Smoothing parameters in the Generalized Spline smoothing problem. *Annals of Statistics,* 4: 1378-1407.

[14] Wu, C. and Sitter, R. R. (2001). A Model Calibration Approach to Using Complete Auxiliary Information from Survey Data. *Journal of American Statistical Association*, 96: 185-193.

[15] Zheng and Little (2003): Penalized Spline Model-Based Estimation of the Finite Populations Total *Journal of Official Statistics*, 19: 99-117.