Research Journal of Mathematical and Statistical Sciences ISSN 2320-6047
Vol. **6(4),** 1-10, April **(2018)** *Res. J. Mathematical and Statistical Sci.*

Semiparametric and nonparametric calibration estimators in cluster sampling by use of penalty functions

Pius Nderitu Kihara
Department of Statistics and Actuarial Sciences, Technical University of Kenya, P.O. BOX 52428-00200 Nairobi, Kenya
piuskihara@yahoo.com

Available online at: www.isca.in, www.isca.me
Received 4th January 2018, revised 28th March 2018, accepted 10th April 2018

Abstract

The application of nonparametric model calibration estimators in multistage survey sampling has been studied by several authors with the cluster level auxiliary information assumed completely available for each cluster. The reasoning behind model calibration is that if the calibration constraints are satisfied by the auxiliary variable, then it is expected that the fitted values of the variable of interest should satisfy such constraints too. In this paper, we have considered a case of auxiliary information present at two levels. We derive estimators by treating the calibration problems at both levels as optimization problems and solving them by the method of penalty functions. We have shown that the estimators obtained are robust since they do not fail in the event the model is misspecified for the data.

Keywords: Optimization problem, semiparametric model, nonparametric model, model calibration, penalty function.

Introduction

The nonparametric and semiparametric modeling techniques have become popular due to the failings of parametric modeling when a model is misspecified. Given a sample s of n triple of observations $(Z_i, x_i, y_i), i = 1,2,....,n$ from some population U of size say N, of interest is to find an estimator for $E(y_i) = g(x_i, Z_i)$ of a missing population value. Once the missing values are imputed, an estimate of the total of the population of the dependent variable Y can then be obtained. Breidt et al[1] considered a super population regression model, ξ given by

$$E_{\xi}(y_i) = g(x_i, Z_i) = \mu(x_i) + Z_i\beta \quad (1)$$

and used a sample estimate of the form $\hat{g}_i = \hat{\mu}(x_i) + Z_i\hat{\beta}$ with $\hat{\mu}(x_i)$ obtained by local polynomial nonparametric method. Accordingly, they obtained the following estimator for population total

$$y_{reg} = \sum_U \hat{g}_i + \sum_s \frac{y_i - \hat{g}_i}{\pi_i} \quad (2)$$

They found that the estimator shares some desirable properties with the fully parametric regression estimators. It is location and scale invariant, and it is internally calibrated for both the parametric and the nonparametric components, in the sense that $\hat{X}_{reg} = \sum_U x_i$ and $\hat{Z}_{reg} = \sum_U Z_i$. The estimator was shown to be design consistent with the rate $\sqrt{n}$, in the sense that $y_{reg} = \sum_U y_i + O_p(\frac{1}{\sqrt{n}})$.

Kihara et al[2] extended the work of Breidt et al[1] to include model calibration in cluster sampling with auxiliary information available at both element and cluster levels and missing values fitted nonparametrically and semiparametrically by use of penalized splines. The work by Kihara[3] considered calibration problem as an optimization problem where missing values were fitted parametrically. Further work by Kihara[4] considered the calibration problem, in one stage sampling, as an optimization problem with missing values fitted nonparametrically and semiparametrically.

In this study, the work by Kihara et al[2] is extended by treating the two levels calibration problems, in cluster sampling, as constrained nonlinear optimization problems which we convert to unconstrained optimization problems. We solve the resulting problems by penalty function method to obtain the weights (at both cluster and cluster element levels) assigned to sample observations from some chi- square distance measures.

Two Level Model Calibration in cluster Sampling

Consider a population U partitioned into M clusters each of size N_i and let C be the population of the clusters. For all clusters included in the sample s, two independent vectors, x_i and z_i are available where z_i is a categorical vector. For simplicity, we let x_i be a scalar. At stage one, a sample s of size m consisting of clusters, is selected from C as per a fixed design $p_1(.)$. Let $\pi_i = p(i \in s)$ and $\pi_{ij} = p(i, j \in s)$ be the marginal and joint cluster inclusion probabilities respectively. From each of the sampled cluster $i \in s$, a sample s_i and size n_i consisting of cluster elements is selected as per a fixed design $p_i(.)$ with the respective marginal and joint element inclusion probabilities as $\pi_{k/i} = p(k \in s_i / i \in s)$ and $\pi_{kl/i} = p(k, l \in s_i / i \in s)$. We assume invariance and independence of the second stage design. Let $t_i = g(x_i, Z_i) + \varepsilon_i, i = 1, 2, \ldots, M$, where the smooth function $g(x_i, Z_i)$ is the fitted model mean for the ith cluster total. For simplicity we write g_i for $g(x_i, Z_i)$. Let $\hat{t}_s = \left[\hat{t}_i\right]_{i \in s}$ be the vector of cluster total estimators $\hat{t}_i$ obtained from the sampled clusters.

Now, consider the case where there is also auxiliary information known at element level such that for each element in the ith cluster, a nonparametric variable x_{ik} and a categorical vector Z_{ik} are available. Suppose that not all element values of the variable of interest in a given cluster are available and have to be imputed. We derive a model calibrated estimator of cluster total. We define the semiparametric estimator for $\mathrm{E}_{\xi_{11}}\left(y_{ik}\right)$ as

$$\hat{g}_{ik} = \hat{g}(x_{ik}, z_{ik}) = \hat{\mu}(x_{ik}) + Z_{ik}\hat{\beta} \tag{3}$$

Where: $\hat{\mu}(x_{ik})$ and x_{ik} are defined for every element k in the cluster C_i. For simplicity, we write $\hat{\mu}_{ik}$ for $\hat{\mu}(x_{ik})$. We propose a model calibrated estimator of cluster total to be

$$\hat{t}_i = \sum_{k \in s_i} w_{ik}\hat{y}_{ik} \tag{4}$$

with the weights w_{ik} derived in such a way that the chi square distance measure below is minimized as discussed by Deville and Sarndal[5].

$$\Phi_s = \sum_{k \in s_i} \frac{(w_{ik} - d_{ik})^2}{q_{ik} d_{ik}} \tag{5}$$

The distance measure is minimized subject to the constraints $\sum_{k \in s_i} w_{ik} = N_i$ and $\sum_{k \in s_i} w_{ik}\hat{g}_{ik} = \sum_{k \in C_i} \hat{g}_{ik}$ proposed by Wu and Sitter[6]. We have the optimization problem below similar to the one of Kihara[4].

$$\min imize\ \Phi_s = \sum_{k \in s_i} \frac{(w_{ik} - d_{ik})^2}{q_{ik} d_{ik}} subject\ to$$

$$\begin{cases} l_1(w_s) = \sum_{k \in s_i} w_{ik}\hat{g}_{ik} - \sum_{k \in C_i} \hat{g}_{ik} = 0\ and \\ l_2(w_s) = \sum_{k \in s_i}^{n} w_{ik} - N_i = 0 \end{cases} \tag{6}$$

We construct an unconstrained problem as given below. See Rao[7].

$$\phi(w_s, r_a) = \sum_{k \in s_i} \frac{(w_{ik} - d_{ik})^2}{q_{ik} d_{ik}} + H(r_a)\left[\sum_{k \in s_i} w_{ik}\hat{g}_{ik} - \sum_{k \in C_i} \hat{g}_{ik}\right]^2 + H(r_a)\left[\sum_{k \in s_i} w_{ik} - N_i\right]^2 \tag{7}$$

Now, $H(r_a)$ is a function of some penalty r_a.

Differentiating (7) partially with respect to w_{ik} we get

$$\phi^1(w_{ik}, r_a) = \frac{2(w_{ik} - d_{ik})}{q_{ik} d_{ik}} + 2H(r_a)\hat{g}_{ik}\left[\sum_{j \in s_i} w_{ij}\hat{g}_{ij} - \sum_{j \in C_i} \hat{g}_{ij}\right] + 2H(r_a)\left[\sum_{k \in s_i} w_{ik} - N_i\right] \tag{8}$$

Equating (8) to zero and solving for w_{ik} we have

$$w_{ik} = \frac{d_{ik} - H(r_a) q_{ik} d_{ik}\left(\sum_{\substack{j \in s_i \\ j \neq k}} w_{ij}[\hat{g}_{ik}\hat{g}_{ij} + 1] - \sum_{j \in C_i} [\hat{g}_{ik}\hat{g}_{ij} - 1]\right)}{1 + H(r_a)\left(((\hat{g}_{ik})^2 + 1) q_{ik} d_{ik}\right)} \tag{9}$$

Thus, a semiparametric estimator of the cluster total is given as

$$\begin{aligned} \hat{t}_i = \sum_{k \in s_i} w_{ik} y_{ik} &= \sum_{k \in s_i} \frac{y_{ik} d_{ik}}{1 + H(r_a)\left(((\hat{g}_{ik})^2 + 1) q_{ik} d_{ik}\right)} \\ &- \sum_{k \in s_i} \frac{H(r_a) q_{ik} d_{ik} y_{ik}\left(\sum_{\substack{j \in s_i \\ j \neq k}} w_{ij}[\hat{g}_{ik}\hat{g}_{ij} + 1] - \sum_{j \in C_i} [\hat{g}_{ik}\hat{g}_{ij} - 1]\right)}{1 + H(r_a)\left(((\hat{g}_{ik})^2 + 1) q_{ik} d_{ik}\right)} \end{aligned} \tag{10}$$

Now, having estimated the cluster totals, we then derive a population total estimator using the estimated cluster totals and the auxiliary information available at cluster level. With $\hat{g}_i$ and x_i defined for every $i \in C$, we propose a semiparametric model calibrated population total estimator as

$$y_{sp} = \sum_{i \in s} w_i \hat{t}_i \tag{11}$$

with w_i obtained in such a way that the chi square distance measure below is minimized.

$$\Phi = \sum_{i \in s} \frac{(w_i - d_i)^2}{q_i d_i} \tag{12}$$

The distance measure is minimized Subject to the constraints $\sum_{i\in s} w_i = M$ and $\sum_{i\in s} w_i \hat{g}_i = \sum_{i\in C} \hat{g}_i$. Again, $d_i = \pi^{-1}{}_i$ and q_i are some known positive constants uncorrelated with d_i. We therefore have the optimization problem

$$\min imize \quad \Phi = \sum_{i\in s} \frac{(w_i - d_i)^2}{q_i d_i} subject \quad to$$

$$\begin{cases} l_1(w) = \sum_{i\in s} w_i \hat{g}_i - \sum_{i\in C} \hat{g}_i = 0 \ and \\ l_2(w) = \sum_{i\in s} w_i - M = 0 \end{cases} \tag{13}$$

We convert (13) to an unconstrained optimization problem below

$$\phi(w, r_b) = \sum_{i\in s} \frac{(w_i - d_i)^2}{q_i d_i} + H(r_b)\left[\sum_{i\in s} w_i \hat{g}_i - \sum_{i\in C} \hat{g}_i\right]^2 + H(r_b)\left[\sum_{i\in s} w_i - M\right]^2 \tag{14}$$

Where: r_b is some penalty.

Differentiating (14) partially with respect to w_i we get

$$\phi^1(w_i, r_b) = \frac{2(w_i - d_i)}{q_i d_i} + 2H(r_b)\hat{g}_i\left[\sum_{j\in s} w_j \hat{g}_j - \sum_{j\in C} \hat{g}_j\right] + 2H(r_b)\left[\sum_{i\in s} w_i - M\right] \tag{15}$$

We equate (15) to zero and solve for w_i to obtain the following.

$$w_i = \frac{d_i - H(r_b) q_i d_i \left(\sum_{\substack{j\in s_i \\ j\neq i}} w_j [\hat{g}_i \hat{g}_j + 1] - \sum_{j=C} [\hat{g}_i \hat{g}_j - 1]\right)}{1 + H(r_b)\left(((\hat{g}_i)^2 + 1) q_i d_i\right)} \tag{16}$$

Now we have a semiparametric estimator of the population total obtained as

$$y_{sp} = \sum_{i\in s} w_i \hat{t}_i = \sum_{i\in s} \frac{\hat{t}_i d_i}{1 + H(r_b)\left(((\hat{g}_i)^2 + 1) q_i d_i\right)} - \sum_{i\in s} \frac{H(r_b) q_i d_i \hat{t}_i \left(\sum_{\substack{j\in s \\ j\neq i}} w_j [\hat{g}_i \hat{g}_j + 1] - \sum_{j\in C} [\hat{g}_i \hat{g}_j - 1]\right)}{1 + H(r_b)\left(((\hat{g}_i)^2 + 1) q_i d_i\right)} \tag{17}$$

When the vectors $Z_i = Z_{ik} = 0$, then $\hat{g}(x_{ik}, z_{ik}) = \hat{\mu}(x_{ik})$ and $\hat{g}(x_i, z_i) = \hat{\mu}(x_i)$. If we let $\hat{\mu}(x_{ik}) = \hat{\mu}_{ik}$ and $\hat{\mu}(x_i) = \hat{\mu}_i$, we have a nonparametric model calibrated estimator for cluster total as

$$\hat{t}_{ni} = \sum_{k\in s_i} w_{ik} y_{ik} = \sum_{k\in s_i} \frac{y_{ik} d_{ik}}{1 + H(r_a)\left((\hat{\mu}_{ik})^2 + 1\right) q_{ik} d_{ik}} - \sum_{k\in s_i} \frac{H(r_a) q_{ik} d_{ik} y_{ik} \left(\sum_{\substack{j\in s_i \\ j\neq k}} w_{ij} [\hat{\mu}_{ik} \hat{\mu}_{ij} + 1] - \sum_{j\in C_i} [\hat{\mu}_{ik} \hat{\mu}_{ij} - 1]\right)}{1 + H(r_a)\left((\hat{\mu}_{ik})^2 + 1\right) q_{ik} d_{ik}} \tag{18}$$

and the nonparametric population total estimator becomes

$$y_{np} = \sum_{i\in s} w_i \hat{t}_{ni} = \sum_{i\in s} \frac{\hat{t}_{ni} d_i}{1 + H(r_b)\left(((\hat{\mu}_i)^2 + 1) q_i d_i\right)} - \sum_{i\in s} \frac{H(r_b) q_i d_i \hat{t}_{ni} \left(\sum_{\substack{j\in s \\ j\neq i}} w_j [\hat{\mu}_i \hat{\mu}_j + 1] - \sum_{j\in C} [\hat{\mu}_i \hat{\mu}_j - 1]\right)}{1 + H(r_b)\left(((\hat{\mu}_i)^2 + 1) q_i d_i\right)} \tag{19}$$

For a semiparametric case, to obtain the within cluster weights w_{ik}, $(k = 1,2,\ldots,n_i)$, we solve the penalty function (7) as an unconstrained minimization problem. Starting with some initial guess for w_{ik} and r_a, we repetitively improve on the guess until optimal values are obtained. Given that our constraints are equality constraints, our initial guess for w_{ik} is not required to be feasible as explained in Kihara[3]. We make use of the Newton method discussed in Rao[7].

Let $W_i = \{w_{i1}, w_{i2}, \ldots, w_{in_i}\}$ be our set of weights. We wish to derive W_i^* so that

$$\vartheta(W_i^*) = \left[\phi'(w_{i1}, r_a), \ldots, \phi'(w_{in_i}, r_a))\right]' = 0 \tag{20}$$

We let W_{il} be the initial approximation of W_i^* so that $W_i^* = W_{il} + V_i$. By Taylor's series expansion of $\vartheta(W_i^*)$ we get

$$\vartheta(W_i^*) = \vartheta(W_{il} + V_i) = \vartheta(W_{il}) + J_{W_{il}} V_i + \ldots\ldots \tag{21}$$

If we ignore the higher order terms in (21) and set $\vartheta(W_i^*) = 0$, we get

$$\vartheta(W_{il}) + J_{W_{il}} V_i = 0 \tag{22}$$

The matrix $J_{W_{il}}$ consists of the second order derivatives of the penalty function (7) evaluated at W_{il}. In general, the $J_{w_{il}}$ matrix is a n_i *by* n_i matrix. Let k and j denote the rows and columns respectively.

Then, $J_{w_{il}}$ has the elements $\frac{2}{q_{ik}d_{ik}} + 2H(r_a)((\hat{g}_{ik})^2 + 1)$ in the diagonal and the elements $2H(r_a)(\hat{g}_{ik}\hat{g}_{ij} + 1)$ elsewhere. If $J_{W_{il}}$ is invertible, then, from the linear equations (22) we have

$$V_i = J_{W_{il}}^{-1}\vartheta(W_{il}) \tag{23}$$

The iterative procedure below is used in finding the enhanced approximations of W_i^*.

$$W_{i(l+1)} = W_{il} - J_{W_{il}}^{-1}\vartheta(W_{il}) \tag{24}$$

The sequence of the points $W_{i1}, W_{i2}, ..., W_{i(l+1)}$ will eventually converge to the actual solution W_i^*.

Let W_{ia}^* be the minimum value of W_i^* calculated for a given penalty r_a, we calculate a sequence of minimum points $W_{i1}^*, W_{i2}^*, ..., W_{i(a+1)}^*$ for the penalties $r_1, r_2, ..., r_{a+1}$ until $W_{ia}^* = W_{i(a+1)}^*$ or $\phi(w_s, r_a) = \phi(w_s, r_{a+1})$ to a given degree of accuracy. The penalty values are such that the initial value $r_1 > 0$ and $r_{a+1} = cr_a$, where $c < 1$. $H(r_a) \to \infty$ as $r_a \to 0$.

In nonparametric case, $\hat{\mu}_{ik}$ replaces $\hat{g}_{ik}$ so that $J_{W_{il}}$ matrix is then a n_i *by* n_i matrix with diagonal elements $\frac{2}{q_{ik}d_{ik}} + 2H(r_a)((\hat{\mu}_{ik})^2 + 1)$ and the elements $2H(r_a)(\hat{\mu}_{ik}\hat{\mu}_{ij} + 1)$ elsewhere.

We next obtain the cluster level weights w_i, $(i = 1,2,...,m)$. Considering the semiparametric case, we solve the penalty function (14) as an unconstrained minimization problem. Let the set of weights be $W = \{w_i, w_2, ..., w_m\}$. We require W^* such that

$$\vartheta(W^*) = [\phi'(w_1, r_b), ..., \phi'(w_m, r_b)]' = 0 \tag{25}$$

We let W_i be the initial estimate of W^* so that $W^* = W_i + V$. The Taylor's series expansion of $\vartheta(W^*)$ now gives

$$\vartheta(W^*) = \vartheta(W_i + V) = \vartheta(W_i) + J_{W_i}V + ... \tag{26}$$

Again, we neglect the higher order terms in (26) and set $\vartheta(W^*) = 0$ to arrive at

$$\vartheta(W_i) + J_{W_i}V = 0 \tag{27}$$

Where: J_{W_i} is the m *by* m matrix of second order derivatives of the penalty function (14) evaluated at W_i. Let i and j be the row and column counters respectively. The matrix J_{W_i} has elements $\frac{2}{q_i d_i} + 2H(r_b)((\hat{g}_i)^2 + 1)$ in the main diagonal and the elements $2H(r_b)(\hat{g}_i\hat{g}_j + 1)$ elsewhere. For nonparametric case, the matrix has $\frac{2}{q_i d_i} + 2H(r_b)((\hat{\mu}_i)^2 + 1)$ as diagonal elements and the elements $2H(r_b)(\hat{\mu}_i\hat{\mu}_j + 1)$ elsewhere.

We now have the iterative procedure below to find the improved estimates of W^*.

$$W_{i+1} = W_i - J_{W_i}^{-1}\vartheta(W_i) \tag{28}$$

Letting W_b^* be the minimum value of W^* calculated for a given penalty r_b, we again calculate a sequence of minimum points $W_1^*, W_2^*, ..., W_{b+1}^*$ for the penalties $r_1, r_2, ..., r_{b+1}$ until $W_b^* = W_{b+1}^*$ or $\phi(w, r_b) = \phi(w, r_{b+1})$ to a specified degree of accuracy. The penalty values for r_b may be set in similar manner as r_a described above.

Local Polynomial Method of Fitting the Missing Values

The aim in local polynomial regression is to minimize the degree q polynomial

$$\sum_{j=1}^{n}\{y_i - \beta_0 - \beta_1(x_j - x_i)..\beta_p(x_j - x_i)^q\}^2 K(x_j - x_i) \tag{29}$$

with respect to $\beta = (\beta_0, \beta_1, ..., \beta_p)$ where β_0 estimates $\mu(x_i) = \mu_i$ while $\beta_1, ..., \beta_p$ estimates higher order derivatives of

μ_i . The kernel function $K(.)$ is discussed in Simonof[8]. From the local polynomial smoother, the nonparametric fit of the cluster totals can be obtained as

$$\hat{\mu}_i = S_{si}^T \hat{t}_s \tag{30}$$

Where: $S_{si}^T = \varepsilon^T (X_{si}^T \varpi_{si} X_{si})^{-1} X_{si}^T \varpi_{si}, \varepsilon = (1,0,...,0)^T$, $\hat{t}_s = (\hat{t}_1, \hat{t}_2, ..., \hat{t}_n)^T$, $\varpi_{si} = diag(K((x_1 - x_i)/h), ..., K((x_n - x_i)/h))$, h is the bandwidth and the matrix X_{si} has the rows $[1, \ (x_j - x_i), \ ..., (x_j - x_i)^q], \ j = 1,2,...,n$. A discussion of this is given by Breidt and Opsomer[9].

In a manner similar to that of Breidt and Opsomer[9], we obtain a semiparametric fit for cluster totals as

$$\hat{g}_i = S_{si}^T (Y_s - Z_s^T \hat{\beta}) + Z_i (Z_s^T S_s Z_s)^{-1} Z_s^T S_s \hat{t}_s \tag{31}$$

where $S_s = [S_{si}, i = 1,2,...,n]$, $\hat{\beta} = (Z_s^T S_s Z_s)^{-1} Z_s^T S_s Y_s$ and $Z_s = [Z_1, Z_2, ...]$ is the vector of categorical variables.

The nonparametric fit of the elements within clusters is then obtained as

$$\hat{\mu}_{ik} = S_{sik}^T Y_{si} \tag{32}$$

Where: $S_{sik}^T = \varepsilon_1^T (X_{sik}^T \varpi_{sik} X_{sik})^{-1} X_{sik}^T \varpi_{sik}$, $\varepsilon_1 = (1,0,...,0)^T$, $Y_{si} = (y_{i1}, y_{i2}, ..., y_{in_i})^T$, $\varpi_{sik} = diag(K((x_{ij} - x_{ik} / h), ..., K((x_{in_i} - x_{ik}) / h_i)), \quad h_i$ is the bandwidth within the ith cluster and X_{sik} is a matrix with rows $[1, \ (x_{ij} - x_{ik}), \ ..., (x_{ij} - x_{ik})^q], \ j = 1,2,..., n_i$.

The semiparametric fit of the elements within clusters is similarly obtained as

$$\hat{g}_{ik} = S_{sik}^T (Y_{si} - Z_{si}^T \hat{\beta}_i) + Z_{ik} (Z_{si}^T S_{sx} Z_{si})^{-1} Z_{si}^T S_{sx} Y_{si} \tag{33}$$

Where $S_{sx} = [S_{sik}, k = 1,2,..., n_i]$, $\hat{\beta}_i = (Z_{si}^T S_{sx} Z_{si})^{-1} Z_{si}^T S_{sx} Y_{si}$ and $Z_{si} = [Z_{i1}, Z_{i2}, ...]$ is the vector of categorical variables in the ith cluster.

Results

We analyze the performance of the derived estimators in comparison to the performance of Horvitz Thompson design estimator $y_{ht} = \sum_{i=1}^{n} \hat{t}_{hi} d_i$ of the population total, where $\hat{t}_{hi} = \sum_{k \in s_i} d_{ik} y_{ik}$ is the cluster total estimator. In Figures-1 to 4, the sample sizes given are for one stage sampling. That is, sizes m of the samples of clusters. The size n_i of the sub sample within a cluster was set as 0.25 of m.

Semiparametric Estimator Results: We simulated a population size 300 of independent and identically distributed variable X using uniform (0.1) and a categorical matrix Z. For each generated x_i and vector Z_i , $N_i = 100$ element values were generated as follows.

$$y_{ik} = \frac{g(x_i, Z_i)}{\sqrt{N_i}} + \frac{\varepsilon_{ik}}{\sqrt{N_i}}, \{\varepsilon_{ik}\} iid \ N(0, 0.1) \tag{34}$$

Where: y_{ik} is the kth element in the ith cluster and $g(x_i, Z_i)$, which we simply write g_i is the mean function for the cluster total t_i. This generating function is an adaptation to semiparametric modeling of the generating function by Montanari and Ranalli[10]. We considered the linear mean function $Z\beta' + 2 + 5x$ and the function $Z\beta' + (2 + 5x)^2$ which is quadratic, for auxiliary information at cluster level. For simplicity, within each cluster, the auxiliary information x_{ik} at element level was generated using the linear and quadratic mean functions and working backward in a similar manner as in Kihara[11] to obtain the following.

$$x_{ik} = \frac{y_{ik} - 2 - z_{ik} \beta'}{5} \tag{35}$$

And

$$x_{ik} = \frac{-2 + \sqrt{y_{ik} - z_{ik} \beta'}}{5} \tag{36}$$

Where: Z_{ik} is the matrix (Z_{i1}, Z_{i2}, Z_{i3}), Z_{i1} is a matrix of 1s, Z_{i2} is a matrix of 2s, 3s and 4s, while Z_{i3} is a matrix of 5s,6s, and 7s. β is the matrix $(1, 2, 3)$.

At stage one, samples of clusters of size m were generated by simple random sampling. At stage two, within each of the selected clusters, sub samples of size n_i were generated by simple random sampling. For any combination of sample sizes m and n_i, 5 samples were generated at stage one and 10 samples at stage two. We used local polynomial equation (33) in fitting cluster elements and equation (31) in fitting cluster totals and in each case the bandwidths are chosen to be a quarter of the respective range in the data. In our study we used a polynomial of degree 1(local linear) and used the standard kernel defined as $K(v)=0.75(1-v^2),\ v\leq 1$.

In estimating cluster totals by the penalty function method, our initial penalty constant for r_a was set at $r_1=0.00010$. The convergence criteria considered was $W_{ia}^{*}=W_{i(a+1)}^{*}$ and $\phi(w_s,r_a)=\phi(w_s,r_{a+1})$ to six decimal places. Using the estimated cluster totals, we generated estimates of the population total. Again, we set the initial penalty value for r_b at $r_1=0.00010$ and the convergence criteria as $W_b^{*}=W_{b+1}^{*}$ and $\phi(w,r_b)=\phi(w,r_{b+1})$ to six decimal places. We compared the performance of our estimator y_{sp} with the Horvitz Thompson estimator y_{ht}.

We let $y_t=\sum_{i\in C} t_i$ be the actual population total where $t_i=\sum_{k\in C_i} y_{ik}$ is the actual cluster total. The errors in the estimation are the differences y_t-y_{sp} and y_t-y_{ht}.

Results on Linear Data: Looking at table (1), the errors indicate that the performances of both estimators y_{sp} and y_{ht} are indistinguishable, which indicates y_{sp} is as reliable as the popular design estimator y_{ht}. Convergence at both stage 1 and stage 2 occurs at the initial values of the penalties. From Figure-1, it can be seen that, from the ratio $\text{var}iance(y_{sp})/\text{var}iance(y_{ht})$, y_{sp} is a bit more variable than Horvitz Thompson estimator y_{ht}.

Table-1: Results of y_{sp} on Linear Data.

sample serial number	1	2	3	4	5
sample sizes m and n_i	100 and 50	100 and 50	100 and 50	100 and 50	100 and 50
y_t	10637.07767	10637.07767	10637.07767	10637.07767	10637.07767
y_{sp}	10655.64312	10591.80901	10776.77585	10657.06451	10616.70892
y_{ht}	10713.64452	10551.51204	10587.80962	10579.51231	10722.72513
y_t-y_{sp}	-18.56545	45.26866	-139.69818	-19.98684	20.36875
y_t-y_{ht}	-76.56685	85.56563	49.26805	57.56536	-85.64746
r_a	0.00010	0.00010	0.00010	0.00010	0.00010
r_b	0.00010	0.00010	0.00010	0.00010	0.00010

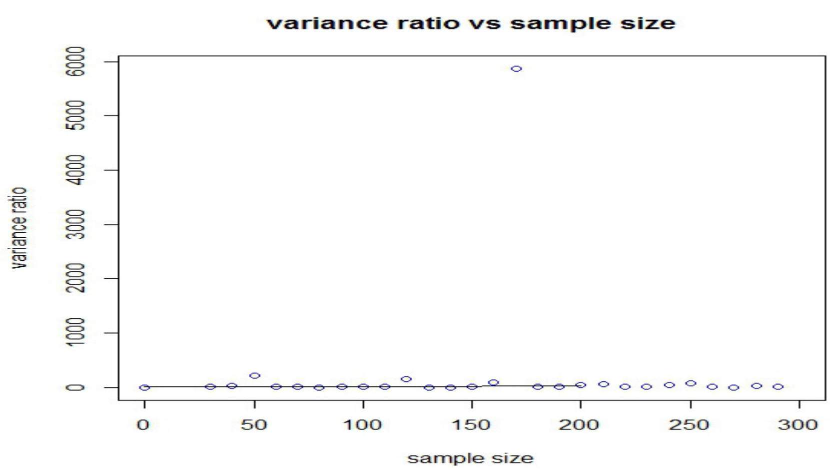

Figure-1: Fraction of $\text{var}iance(y_{sp})/\text{var}iance(y_{ht})$ on Linear Data

Results on Quadratic Data: Looking at table (2), again the errors in the estimation indicate that the performances of y_{sp} and y_{ht} are indistinguishable. This serves to show robustness of the estimator y_{sp} which is in fact a misspecified model for quadratic data. In figure (2), we see that, from the ratio $\mathrm{var}iance(y_{sp})/\mathrm{var}iance(y_{ht})$, the estimator y_{sp} has bigger variance than y_{ht}. This is expected since the data is from a quadratic function, while y_{sp} uses a local linear function in fitting the values.

Nonparametric Estimator Results: Using R software program, and using uniform (0, 1), a population of the variable x was simulated. Using the auxiliary variable x, two populations for the dependent random variable y were generated as $y = 2+5x$ and $y = (2+5x)^2$. We used local polynomial equation (30) to fit cluster totals and equation (32) to fit element values within a cluster. The cluster element values were generated as

$$y_{ik} = \frac{\mu(x_i)}{N_i} + \frac{\varepsilon_{ik}}{\sqrt{N_i}}, \{\varepsilon_{ik}\} iid\ N(0,0.1) \tag{37}$$

Table-2: Results of y_{sp} on Quadratic Data.

sample serial number	1	2	3	4	5
sample size m and n_i	100 and 50	100 and 50	100 and 50	100 and 50	100 and 50
y_t	16054.39204	16054.39204	16054.39204	16054.39204	16054.39204
y_{sp}	16077.78872	16389.73764	15525.60252	15845.52393	15936.92781
y_{ht}	15706.03516	16389.95682	16259.14548	16174.82744	16081.07106
$y_t - y_{sp}$	-23.39668	-335.3456	521.78952	208.86811	117.46423
$y_t - y_{ht}$	347.60332	-335.56478	-204.75344	-120.4354	-26.67902
r_a	0.00010	0.00010	0.00010	0.00010	0.00010
r_b	0.00010	0.00010	0.00010	0.00010	0.00010

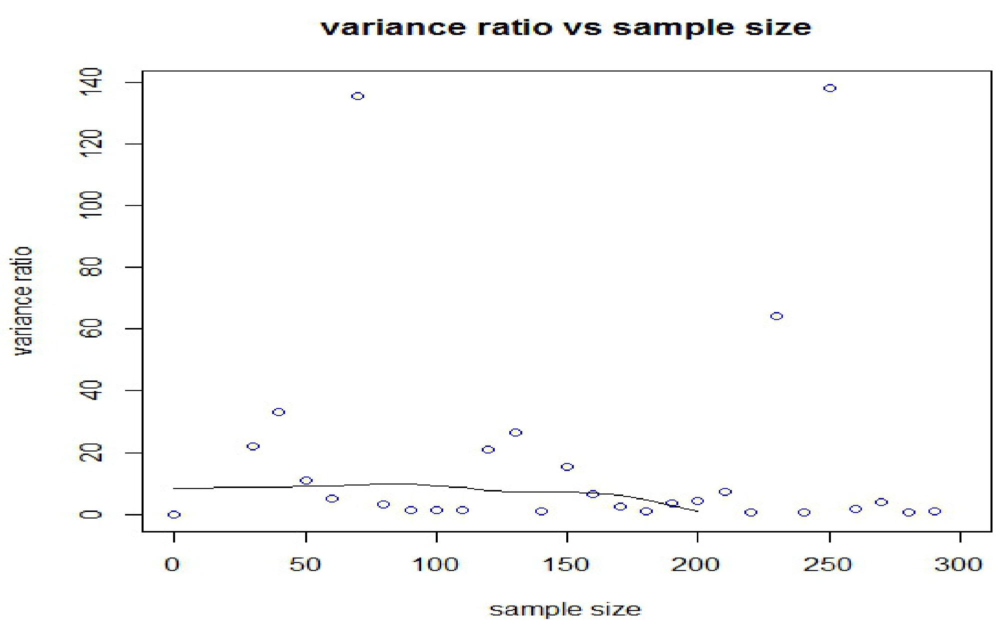

Figure-2: Fraction of $\mathrm{var}iance(y_{sp})/\mathrm{var}iance(y_{ht})$ on Quadratic Data.

The respective auxiliary information was regenerated as shown below for the linear and quadratic mean functions.

$$x_{ik} = \frac{y_{ik} - 2}{5} \tag{38}$$

And

$$x_{ik} = \frac{\sqrt{y_{ik}} - 2}{5} \tag{39}$$

Results on Linear Data: From Table-3, both estimators y_{np} and y_{ht} have small errors and consistently, y_{np} has the smaller error margins. This can be explained by the fact that the data is linear which implies that y_{np} is correctly specified for the data. From Figure-3, we see that the ratio $\mathrm{var}iance(y_{np})/\mathrm{var}iance(y_{ht})$ increases as the sample size grows up to a constant of about 0.37. Thus, the variance for y_{np} is consistently lower than that of y_{ht}. This can be explained by the fact that y_{np} is correctly specified for the given data.

Table-3: Results of y_{np} on Linear Data.

Sample serial number	1	2	3	4	5
Sample size m and n_i	100 and 50	100 and 50	100 and 50	100 and 50	100 and 50
y_t	1344.531793	1344.531793	1344.531793	1344.531793	1344.531793
y_{np}	1345.95725	1340.25334	1327.40832	1349.00000	1350.49969
y_{ht}	1347.04198	1337.97500	1318.56434	1351.24476	1353.21979
$y_t - y_{np}$	-1.425457	4.278453	17.123461	-4.468207	-5.967897
$y_t - y_{ht}$	-2.510187	6.556793	25.967453	-6.712967	-8.687997
r_a	0.00010	0.00010	0.00010	0.00010	0.00010
r_b	0.00010	0.00010	0.00010	0.00010	0.00010

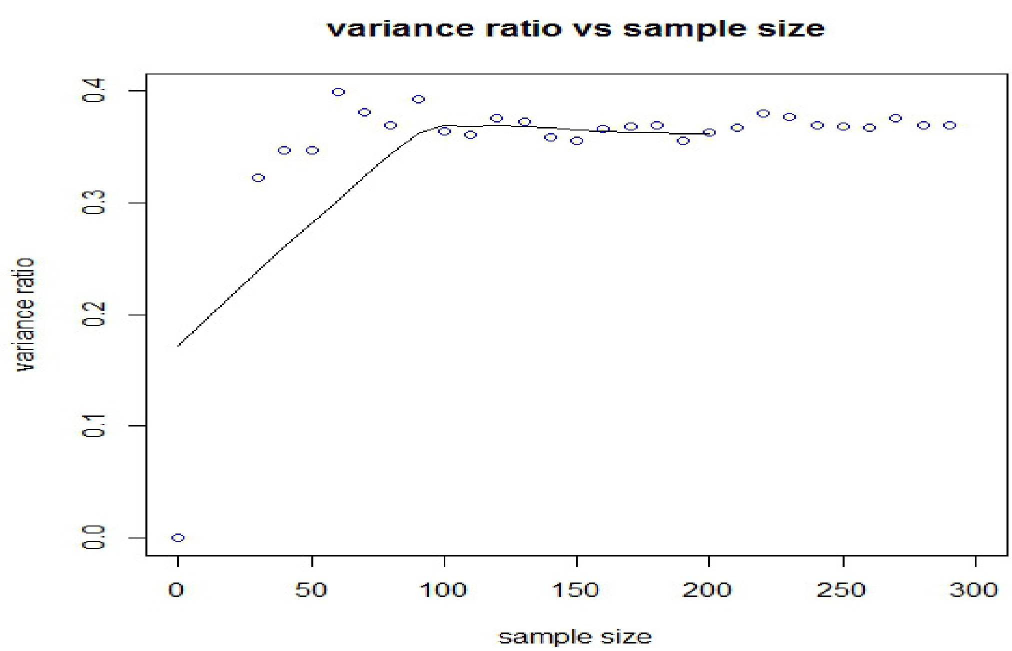

Figure-3: Fraction of $\mathrm{var}iance(y_{np})/\mathrm{var}iance(y_{ht})$ on Linear Data

Results on Quadratic Data: Looking at Table-4, the errors in estimation indicates that the performances of the estimators y_{np} and y_{ht} are indistinguishable. This points to the robustness of the estimator y_{np} which is misspecified tor the quadratic data. The ratio $\mathrm{var}iance(y_{np})/\mathrm{var}iance(y_{ht})$ seem to tend to a constant as seen in figure (4), though the ratio is a bit wild for small samples. Also, the variance for y_{np} is larger than the variance for y_{ht} .

Conclusion

From the results, it is clear that when the nonparametric estimator y_{np} is correctly specified for the data, it is more efficient than the popular Horvitz Thompson design estimator y_{ht} and that y_{np} is only slightly less efficient when it is misspecified for the data. Also, the performance of the semiparametric estimator y_{sp} is indistinguishable from that of the design estimator. We conclude that the semiparametric and nonparametric estimators are robust estimators since they do not fail under misspecification.

In a real world problem where we may not have, or may not be sure that we have all the relevant auxiliary information about a variable, model calibrated estimators would therefore be the estimators of choice. We have shown that in cases where some elements within clusters are unreachable but auxiliary information is available at element level, we can take advantage of this auxiliary information to obtain cluster totals, which are then used in the estimation of population total. We note that if there is a possibility that some clusters may be unreachable, it means there is also the possibility that some cluster elements may be unreachable.

Table-4: Results of y_{np} on Quadratic Data.

Sample serial number	1	2	3	4	5
Sample size m and n_i	100 and 50	100 and 50	100 and 50	100 and 50	100 and 50
y_t	6702.63067	6702.63067	6702.63067	6702.63067	6702.63067
y_{np}	6989.35523	6579.98771	7013.19892	6677.42846	6716.28391
y_{ht}	6411.78004	6589.61917	6853.44946	6655.73623	6802.89124
$y_t - y_{np}$	-286.72456	122.64296	-310.56825	25.20221	-13.65324
$y_t - y_{ht}$	290.85063	113.0115	-150.81879	46.89444	-100.26057
r_a	0.00010	0.00010	0.00010	0.00010	0.00010
r_b	0.00010	0.00010	0.00010	0.00010	0.00010

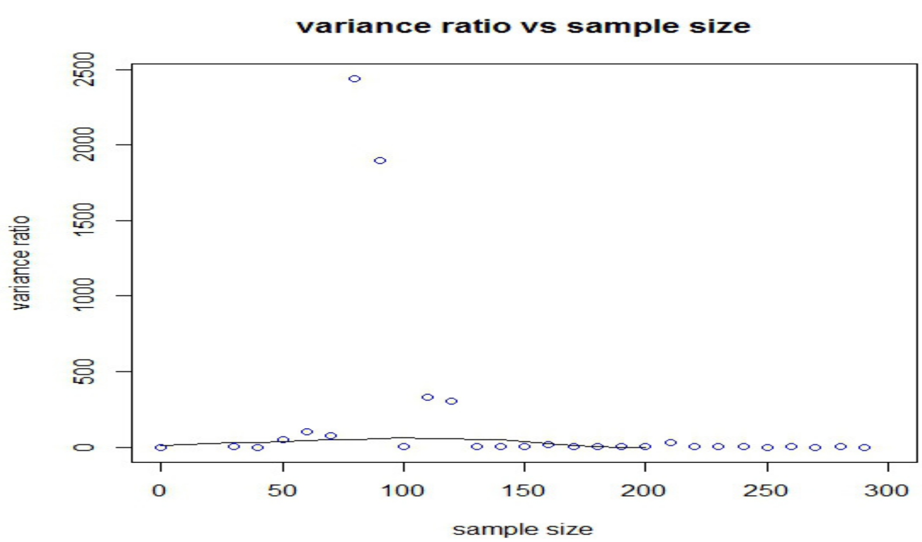

Figure-4: Fraction of $\mathrm{var}iance(y_{np})/\mathrm{var}iance(y_{ht})$ on Quadratic Data.

References

1. Breidt F.J., Opsomer J.D., Johnson A.A. and Ranalli M.G. (2007). Semiparametric model-assisted estimation for natural resource surveys. *Survey Methodology*, 33(1), 35.
2. Kihara P.N., Otieno R.O. and Kihoro J.M. (2015). Two Levels Model Calibration in Cluster Sampling; Use of Penalized Splines in Semiparametric Estimation. *Mathematical Theory and Modeling,* 5(4), 94-104.
3. Kihara P.N. (2017). Calibration Estimators by Penalty Function Method. *Mathematical Theory and Modeling,* 7(6), 22-32.
4. Kihara P.N. (2017). Robust Nonparametric and Semiparametric Model Calibration Estimators by Penalty Function Method. *Mathematical Theory and Modeling,* 7(8), 22-39.
5. Deville J.C. and Sarndal C.E. (1992). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association,* 87(418), 376-382.
6. 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(453), 185-193.
7. Rao S.S. (1984). Optimization Theory and Applications. Wiley Eastern Limited, 390-424, ISBN: 0-85226-756-8.
8. Simonof J. (1996). Smoothing Methods in Statistics. New York: Springer, 40-93, ISBN: 0-387-94716-7
9. Breidt F.J. and Opsomer J.D. (2000). Local Polynomial Regression Estimation in Survey Sampling. *Annals of Statistics,* 28(4), 1026-1053.
10. Montanari G.E. and Ranalli M.G. (2003). Nonparametric Model Calibration Estimation in Survey Sampling. *Journal of American Statistical Association,* 100(472), 1429-1442.
11. Kihara P.N. (2012). Estimation of Finite Population Total in the Face of Missing Values Using Model Calibration and Model Assistance on Semiparametric and Nonparametric Models (Unpublished doctoral thesis). Jomo Kenyatta University of Agriculture and Technology, Kenya.