

Robust Nonparametric and Semiparametric Model Calibration Estimators by Penalty Function Method

Pius Nderitu Kihara
Department of Statistics and Actuarial Sciences, Technical University of Kenya
Email: piuskihara@yahoo.com

Abstract

Use of nonparametric model calibration estimators for population total and mean has been considered by several authors. In model calibration, a distance measure defined on some design weights thought to be close to the inclusion probabilities, is minimized subject to some calibration constraints imposed on the fitted values of the study variable. The minimization is usually by way of introducing langrage equation whose solution gives the optimal design weights to be used in estimation of population total. Sometimes a solution to the langrage constants does not exist. Numerical approaches are some of the alternatives to the langrage approach. In this paper, we have derived nonparametric and semiparametric model calibration estimators by treating the calibration problem as a nonlinear constrained minimization problem, which we transform to an unconstrained optimization problem using penalty functions. We show that the resulting nonparametric and semiparametric estimators are robust in the sense that they are quite efficient when the model is correctly specified for the data and that the estimators do not fail even when the model is misspecified for the data. When the penalty constant approaches zero, the estimators reduce to the Horvitz Thompson design estimator.

Keywords: model calibration, nonparametric model, semiparametric model, penalty function

1. Introduction

Use of auxiliary information in estimation of missing values and descriptive parameters of a survey variable in a finite population has become fairly common. A simple way to incorporate known population totals of auxiliary variables is through ratio and regression estimation. More general situations are handled by means of generalized regression estimation as discussed by Sarndal [10] and calibration estimation discussed by Deville and Sarndal [4]. The processes of estimation of population total and mean starts first with the point estimation of the missing values based on auxiliary variable. Then, techniques like calibration and model assistance are employed on the fitted values to estimate population parameters and or any other required analysis of the data are carried out. The reasoning towards use of nonparametric and semiparametric modeling techniques for the missing values includes the following. First, an initial nonparametric estimate may well suggest a suitable parametric model such as linear regression. That is, it may give the data more of a chance to speak for themselves in choosing the model to be fitted (Silverman [11]). Secondly, known facts suggest a tentative model which in turn suggest a particular examination and analysis of data or the need to acquire further data or suggest a modified model resulting in an iterative procedure (Box[1], Hastie and Tibshirani [6], Simonof[12]). It is very important to note that parametric models would be very efficient if the model is correctly specified. However, if the assumed model is misspecified, inferences can lead to misleading interpretations of data.

Considered is a super population regression model which is denoted by ξ and given as

$$y_i = \mu(x_i) + \varepsilon \quad (1)$$

where $\mu(x_i)$ is a smooth function. Given n pair of observations $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ from a population of size N, of interest is the estimator $\hat{\mu}(x_i)$ of $\mu(x_i) = E_\xi(y/x)$. A nonparametric method like local polynomial or splines could be used for this estimation.

In some circumstances, the auxiliary information is such that it contains a component whose parametric structure is known and a component that need to enter the estimation nonparametrically. Consider case where auxiliary information consists of a single univariate term x that is to enter estimation nonparamtrically and a vector Z composed of an arbitrary number of linear terms.

Consider super population regression model given by

$$y_i = g(x_i, z_i) + \varepsilon \quad (2)$$

where z_i is a vector of the categorical or continuous auxiliary variable. The interest is to find an estimator

$$\hat{g}(x_i, z_i) \text{ of } g(x_i, z_i) = E_\xi(y/x, z) \quad (3)$$

This is semiparametric estimation. Breidt et al [3] uses a sample estimator of the form

$$\hat{g}(x_i, z_i) = \hat{\mu}(x_i) + z_i\hat{\beta} \quad (4)$$

Once the missing data has been modeled, a nonparametric estimator $\hat{y}_t = \sum_{i=1}^{n} w_i y_i$ for the population total $\sum_{i=1}^{N} y_i$ is then obtained where given the sample inclusion probability π_i, the weights $w_i^{'}s$ are design weights which are as close as possible to $d_i = \pi_i^{-1}$ and are obtained by minimizing a given distance measure between $w_i^{'}s$ and $d_i^{'}s$ subject to some constraints. Wu and Sitter [14] considered the two constraints below

$$\sum_{i=1}^{n} w_i x_i = \sum_{i=1}^{N} x_i \quad (5)$$

$$\sum_{i=1}^{n} w_i = N \quad (6)$$

In a parametric setting, Kihara [7] considered the conversion of the above calibration problem into an optimization problem. He has considered reducing the chis square distance measure below

$$\Phi = \sum_{i \in s} \frac{(w_i - d_i)^2}{q_i d_i} \quad (7)$$

subject to constraints (5) and (6) to obtain a penalty function

$$\phi(w,r_k,x)=\sum_{i=1}^{n}\frac{(w_i-d_i)^2}{q_i d_i}+r_k\left[\sum_{i=1}^{n}w_i x_i-\sum_{i=1}^{N}x_i\right]^2+r_k\left[\sum_{i=1}^{n}w_i-N\right]^2 \quad (8)$$

Differentiating (8) partially with respect to w_i he got

$$\phi^1(w_i,r_k,x)=\frac{2(w_i-d_i)}{q_i d_i}+2r_k x_i\left[\sum_{j=1}^{n}w_j x_j-\sum_{j=1}^{N}x_j\right]+2r_k\left[\sum_{i=1}^{n}w_i-N\right] \quad (9)$$

Equating (9) to zero and solving for w_i we have

$$w_i=\frac{d_i-r_k q_i d_i\left(\sum_{\substack{j=1\\ j\neq i}}^{n}w_j(x_i x_j+1)-\sum_{j=1}^{N}(x_i x_j-1)\right)}{1+r_k\left((x_i^2+1)q_i d_i\right)} \quad (10)$$

He therefore derived the following estimator of population total

$$\hat{y}_t=\sum_{i=1}^{n}w_i y_i=\sum_{i=1}^{n}\frac{y_i d_i}{1+r_k\left((x_i^2+1)q_i d_i\right)}-\sum_{i=1}^{n}\frac{r_k q_i d_i y_i\left(\sum_{\substack{j=1\\ j\neq i}}^{n}w_j(x_i x_j+1)-\sum_{j=1}^{N}(x_i x_j-1)\right)}{1+r_k\left((x_i^2+1)q_i d_i\right)} \quad (11)$$

To obtain the weights w_i, $(i=1,2,...,n)$, the penalty function (8) is solved as an unconstrained minimization problem in which case we only require to start with some initial guess for w_i and r_k and then iteratively improve on the initial values until we have optimal values. Since the constraints (5) and (6) are equality constraints, we need not start with a feasible guess for w_i. In this paper we extend the work of Kihara [7] to nonparametric and semiparametric regression modeling. We also consider model calibration in which case calibration is done with respect to the fitted values.

2. Penalty Function Method for Nonparametric and Semiparametric Estimators

Let there be a population of size N for our variable of interest y from which we draw a sample of size n. Let the auxiliary value x_i be available for every element of the population of variable y. We wish to estimate the population total $y_t=\sum_{i=1}^{N}y_i$ from a sample of size n and incorporating the auxiliary information present. The penalty function method transforms the basic constrained optimization problem into an unconstrained optimization problem. In nonparametric model calibration estimation, we consider an optimization problem of the form

$$\min imize\ \Phi = \sum_{i\in s}\frac{(w_i-d_i)^2}{q_i d_i} subject\ to \qquad (12)$$
$$\begin{cases} l_1(w) = \sum_{i=1}^{n} w_i\hat{\mu}(x_i) - \sum_{i=1}^{N}\hat{\mu}(x_i) = 0\ and \\ l_2(w) = \sum_{i=1}^{n} w_i - N = 0 \end{cases}$$

where $\hat{\mu}(x_i)$ is a nonparametric fit of the missing value y_i. Here, calibration constraint $\sum_{i=1}^{n} w_i\hat{\mu}(x_i) - \sum_{i=1}^{N}\hat{\mu}(x_i) = 0$ is defined on the fitted values. We call this model calibration. We construct an unconstrained problem as follows.

$$\phi(w,r_k) = \sum_{i\in s}\frac{(w_i-d_i)^2}{q_i d_i} + \psi\ (r_k, l_j(w)), \quad j = 1,2 \qquad (13)$$

where $\psi\ (r_k, l_j(X))$ is a penalty function which is continuous and which is such that $\psi\ (r_k,t) \geq 0$ for all r_k and $t \in \Re^n$. Also, $\psi\ (r_k,t)$ is strictly increasing for $r_k > 0$ and $t > 0$. In a form similar to the one discussed in Rao [6], we have the function

$$\phi(w,r_k) = \sum_{i\in s}\frac{(w_i-d_i)^2}{q_i d_i} + H(r_k)\sum_{j=1}^{2} l_j^q(w) \qquad (14)$$

where $H(r_k)$ is some function of the parameter r_k tending to infinity as r_k tends to zero and so that $\sum_{j=1}^{2} l_j^q(w)$ also tend to zero. A common choice for value of q is 2. Also, the function ϕ will always be greater than f. The penalty terms are chosen such that their values will be small at points away from the constraint boundaries and will tend to infinity as the constraint boundaries are approached. Hence, the value of ϕ will also blow up as the constraint boundaries are approached. Frank and Jorge [5] have discussed flexible ways of choosing the penalty. In an iterative process, the unconstrained minimization of ϕ does not have to start with a feasible solution since we have equality constraints. The subsequent points generated will always lie within the feasible region since the constraint boundaries act as barriers during the minimization process. The rationale of the penalty terms as described by Ozgur [8] is that if the constraint is violated, that means $l_j(w) \neq 0$, a term will be added to ϕ function such that the solution is pushed back towards to the feasible region.

In the minimization of ϕ, for the solution to be the global, $\sum_{i \in s} \frac{(w_i - d_i)^2}{q_i d_i}$ and $\sum_{j=1}^{2} l_j^q(w)$ should be convex and one of the functions $\sum_{i \in s} \frac{(w_i - d_i)^2}{q_i d_i}$, $l_1^q(w)$ and $l_2^q(w)$ be strictly convex. See Rao [9]. If we let $q = 2$ then, from equations (12) and (14), we have the penalty function

$$\phi(w, r_k, \hat{\mu}(x)) = \sum_{i=1}^{n} \frac{(w_i - d_i)^2}{q_i d_i} + H(r_k)\left[\sum_{i=1}^{n} w_i \hat{\mu}(x_i) - \sum_{i=1}^{N} \hat{\mu}(x_i)\right]^2 + H(r_k)\left[\sum_{i=1}^{n} w_i - N\right]^2 \tag{15}$$

Differentiating (15) partially with respect to w_i we get

$$\phi^1(w_i, r_k, \hat{\mu}(x)) = \frac{2(w_i - d_i)}{q_i d_i} + 2H(r_k)\hat{\mu}(x_i)\left[\sum_{j=1}^{n} w_j \hat{\mu}(x_j) - \sum_{j=1}^{N} \hat{\mu}(x_j)\right] + 2H(r_k)\left[\sum_{i=1}^{n} w_i - N\right] \quad (\tag{16}$$

Equating (16) to zero and solving for w_i we have

$$w_i = \frac{d_i - H(r_k) q_i d_i \left(\sum_{\substack{j=1 \\ j \neq i}}^{n} w_j [\hat{\mu}(x_i)\hat{\mu}(x_j) + 1] - \sum_{j=1}^{N} [\hat{\mu}(x_i)\hat{\mu}(x_j) - 1] \right)}{1 + H(r_k)\left((\hat{\mu}(x_i)^2 + 1) q_i d_i\right)} \tag{17}$$

A weighted nonparametric estimator of population total is therefore obtained as

$$\hat{y}_{np} = \sum_{i=1}^{n} w_i y_i = \sum_{i=1}^{n} \frac{y_i d_i}{1 + H(r_k)\left((\hat{\mu}(x_i)^2 + 1) q_i d_i\right)} - \sum_{i=1}^{n} \frac{H(r_k) q_i d_i y_i \left(\sum_{\substack{j=1 \\ j \neq i}}^{n} w_j [\hat{\mu}(x_i)\hat{\mu}(x_j) + 1] - \sum_{j=1}^{N} [\hat{\mu}(x_i)\hat{\mu}(x_j) - 1] \right)}{1 + H(r_k)\left((\hat{\mu}(x_i)^2 + 1) q_i d_i\right)} \tag{18}$$

In semiparametric estimation, we have an optimization problem of the form

$$\min imize\ \Phi = \sum_{i\in s} \frac{(w_i - d_i)^2}{q_i d_i} subject\ to$$
$$\begin{cases} l_1(w) = \sum_{i=1}^{n} w_i \hat{g}(x_i) - \sum_{i=1}^{N} \hat{g}(x_i) = 0\ and \\ l_2(w) = \sum_{i=1}^{n} w_i - N = 0 \end{cases} \quad (19)$$

where $\hat{g}(x_i)$ is a semiparametric fit of the missing value y_i. We have the penalty function as

$$\phi(w, r_k, \hat{g}(x)) = \sum_{i=1}^{n} \frac{(w_i - d_i)^2}{q_i d_i} + H(r_k)\left[\sum_{i=1}^{n} w_i \hat{g}(x_i) - \sum_{i=1}^{N} \hat{g}(x_i)\right]^2 + H(r_k)\left[\sum_{i=1}^{n} w_i - N\right]^2 \quad (20)$$

This yields the following semiparametric estimator of the population total

$$\hat{y}_{sp} = \sum_{i=1}^{n} w_i y_i = \sum_{i=1}^{n} \frac{y_i d_i}{1 + H(r_k)\left((\hat{g}(x_i)^2 + 1) q_i d_i\right)} - \sum_{i=1}^{n} \frac{H(r_k) q_i d_i y_i \left(\sum_{\substack{j=1 \\ j\neq i}}^{n} w_j [\hat{g}(x_i)\hat{g}(x_j) + 1] - \sum_{j=1}^{N} [\hat{g}(x_i)\hat{g}(x_j) - 1] \right)}{1 + H(r_k)\left((\hat{g}(x_i)^2 + 1) q_i d_i\right)} \quad (21)$$

From equations (18) and (21), we see that as $H(r_k) \to 0$, the estimators reduce to the Horvitz Thompson design estimator $\sum_{i=1}^{n} y_i d_i$.

To obtain the weights w_i, $(i = 1,2,..., n)$, we solve the penalty functions (15) and (20) as unconstrained minimization problems in which case we only require to start with some initial guess for w_i and r_k and then iteratively improve on the initial values until we have optimal values. Since the constraints in our case are equality constraints, we need not start with a feasible guess for w_i as discussed in Kihara [7]. We appeal to Newton method of unconstrained optimization. See Rao [9].

Considering the nonparametric case, let $W = \{w_i, w_2,..., w_n\}$ be the set of the weights. We need to obtain W^* such that

$$\vartheta(W^*) = [\phi'(w_1, r_k, \hat{\mu}(x)),..., \phi'(w_n, r_k, \hat{\mu}(x))]^{\prime} = 0 \quad (22)$$

We first start with some initial approximation W_i of W^* so that $W^* = W_i + Z$. The Taylor's series expansion of

$\vartheta(W^*)$ gives

$$\vartheta(W^*) = \vartheta(W_i + Z) = \vartheta(W_i) + J_{W_i} Z + ... \quad (23)$$

By neglecting the higher order terms in (23) and setting $g(W^*) = 0$ we obtain

$$\vartheta(W_i) + J_{W_i} Z = 0 \quad (24)$$

Where J_{W_i} is the matrix of second derivatives evaluated at W_i. In general, J is a n by n matrix with $i = 1,2,...,n$ rows and $j = 1,2,...,n$ columns with diagonal elements $\frac{2}{q_i d_i} + 2H(r_k)(\hat{\mu}(x_i)^2 + 1)$ and

elements $2H(r_k)(\hat{\mu}(x_i)\hat{\mu}(x_j) + 1)$ elsewhere. If J_{W_i} is nonsingular, then, from the set of linear equations (24) we have for vector Z

$$Z = J_{W_i}^{-1} \vartheta(W_i) \quad . \quad (25)$$

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

$$W_{i+1} = W_i + Z_i = W_i - J_{W_i}^{-1} \vartheta(W_i) \quad (26)$$

The sequence of the points $W_1, W_2, ..., W_{i+1}$ eventually converges to the actual solution W^*.

Now, if we let W_k^* be the minimum of W^* obtained for a particular penalty r_k, we obtain a sequence of minimum points $W_1^*, W_2^*, ..., W_{k+1}^*$ for the penalties $r_1, r_2, ..., r_{k+1}$ until $W_k^* = W_{k+1}^*$ or $\phi(w, r_k, \hat{\mu}(x)) = \phi(w, r_{k+1}, \hat{\mu}(x))$ for some specified accuracy level. The accuracy level may for example be, to certain decimal points or significance level. The penalty values may be set such that the starting point $r_1 > 0$ and $r_{k+1} = cr_k$, where $c < 1$. $H(r_k) \to \infty$ as $r_k \to 0$.

The Newton solution process in semiparametric case is similar to that of nonparametric case described above but with $\hat{\mu}(x_i)$ replaced by $\hat{g}(x_i)$. The J matrix is a n by n matrix with diagonal elements $\frac{2}{q_i d_i} + 2H(r_k)(\hat{g}(x_i)^2 + 1)$ and elements $2H(r_k)(\hat{g}(x_i)\hat{g}(x_j) + 1)$ elsewhere.

3. Fitting the Missing Values by Local Polynomial Method

The objective in polynomial regression is to minimize

$$\sum_{j=1}^{n} \left\{ y_i - \beta_0 - \beta_1(x_j - x_i) ... \beta_p(x_j - x_i)^q \right\}^2 K(x_j - x_i) \quad (27)$$

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

$\mu(x_i)$. Also, q is the degree of the polynomial and $K(.)$ is some kernel function, a discussion of which is given by Simonof [12]. The corresponding nonparametric fit can be obtained from the local polynomial smoother as

$$\hat{\mu}(x_i) = S_{si}^T Y_s \tag{28}$$

where $S_{si}^T = \varepsilon_1^T (X_{si}^T \varpi_{si} X_{si})^{-1} X_{si}^T \varpi_{si}$, $\varepsilon = (1,0,...,0)^T$, $Y_s = (y_1, y_2, ..., y_n)^T$, $\varpi_{si} = (K((x_i - x_i)/h),..., K((x_i - x_i)/h))$, h is the bandwidth and X_{si} is a matrix with rows $[1,\ (x_j - x_i),\ ...,(x_j - x_i)^q],\ j = 1,2,...,n$. See Breidt and Opsomer[2] .

A semiparametric fit for the missing values similar to that derived by Breidt and Opsomer [2] may be obtained as

$$\hat{g}_i = S_{si}^T (Y_s - Z_s^T \hat{\beta}) + Z_i (Z_s^T S_s Z_s)^{-1} Z_s^T S_s Y_s \tag{29}$$

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.

4. Empirical Results

In section 4.1, we report on the performance of the nonparametric estimator y_{np}. In subsection 4.1.1, we have results of the nonparametric estimator y_{np} on the linear model data and a comparison of its performance with that of Horvitz Thompson estimator $y_{ht} = \sum_{i=1}^{n} y_i d_i$ discussed in Thompson [13]. In

subsection 4.1.2, we report on the results for estimator y_{np} on the quadratic model data and again compare with Horvitz Thompson estimator. In section 4.2, we discuss the performance of the semiparametric estimator y_{sp} where in subsection 4.2.1, we have results of the estimator y_{sp} on the linear model data and a comparison of its performance with that of Horvitz Thompson estimator y_{ht}. In subsection 4.2.1, we report on the results for estimator y_{sp} on the quadratic model data.

4.1. Analysis of the Nonparametric Estimator Results

Using R program, we simulated a population of independent and identically distributed variable x using uniform (0, 1). Using x as the auxiliary variable we generated the populations of size 300 for random variable y as a linear function $y = 2 + 5x$ and quadratic function $y = (2 + 5x)^2$. For each of different sample sizes n, 5 samples were generated. Our initial penalty constant was set at $r_1 = 0.00010$. The convergence criteria considered was $W_k^* = W_{k+1}^*$ and $\phi(w, r_k, x) = \phi(w, r_{k+1}, x)$ to six decimal places.

We used local polynomial method described in section (3.0) to fit the missing values. in particular ,we have considered local polynomial of degree 1, that is local linear function. We have used the standard Epernecknikov kernel $K(u) = 3/4(1-u^2), u \leq 1$ with a bandwidth 0f 0.25. The choice of the bandwidth is based on the ad hoc rule of a quarter of the range of the data.

4.1.1. Results for Nonparametric Estimator y_{np} on Linear Model Data

We let $y_t = \sum_{i=1}^{N} y_i$ be the actual population total, r_k be the penalty parameter, and $y_t - y_{np}$ and $y_t - y_{ht}$ be the errors in the estimation.

Table 1: Nonparametric Estimates for Linear Model Data					
sample number	1	2	3	4	5
sample size n	100	100	100	100	100
y_t	1344.531793	1344.531793	1344.531793	1344.531793	1344.531793
y_{np}	1345.865888	1341.235027	1330.40555	1348.019108	1348.805556
y_{ht}	1346.733668	1339.116040	1321.57077	1350.289785	1351.609775
$y_t - y_{np}$	-1.334095	3.296766	14.12624	-3.487315	-4.273763
$y_t - y_{ht}$	-2.201875	5.415753	22.96103	-5.757992	-7.077982
r_k	0.00010	0.00010	0.00010	0.00010	0.00010

From table (1), the estimators y_{np} and y_{ht} have small error margins. Consistently, y_{np} has a smaller error margin. This is expected because the data is linear and y_{np} is obtained from a linear local polynomial model. We say the nonparametric model is correctly specified for the data. For all the samples, convergence is achieved at the same penalty value of 0.00010 and which was the initial penalty value.

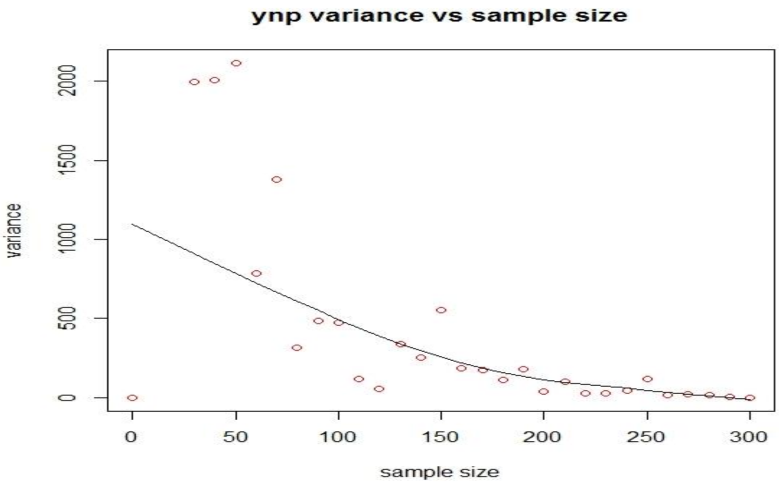

Figure 1: Variance for Estimator y_{np} *on Linear Model Data*

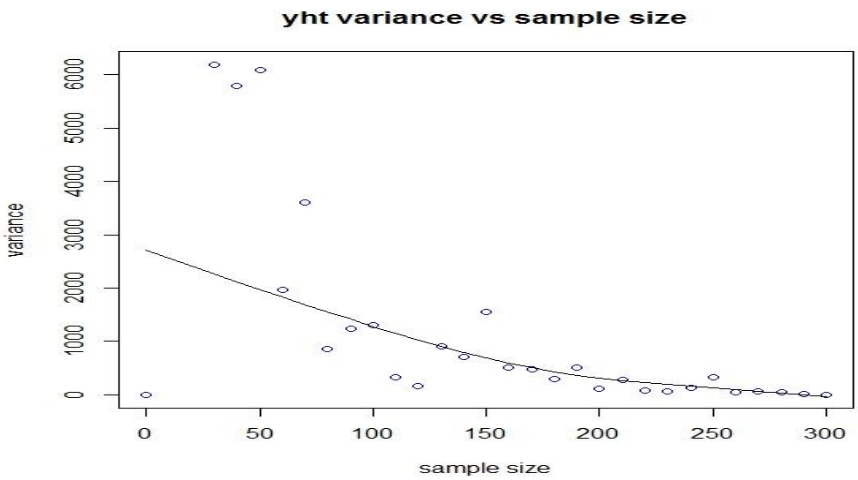

Figure 2: Variance for Horvitz Thompson Estimator y_{ht} *on Linear Model Data*

In figure (1) and figure (2), the variances for y_{np} and y_{ht} decrease as the sample size increases. From figure (3), the ratio $\text{var}(y_{np})/\text{var}(y_{ht})$ settles almost to a constant, estimated to be 0.37, as the sample size increases. That is, y_{np} consistently has a lower variance than y_{ht}. This is expected since y_{np} is correctly specified for the data.

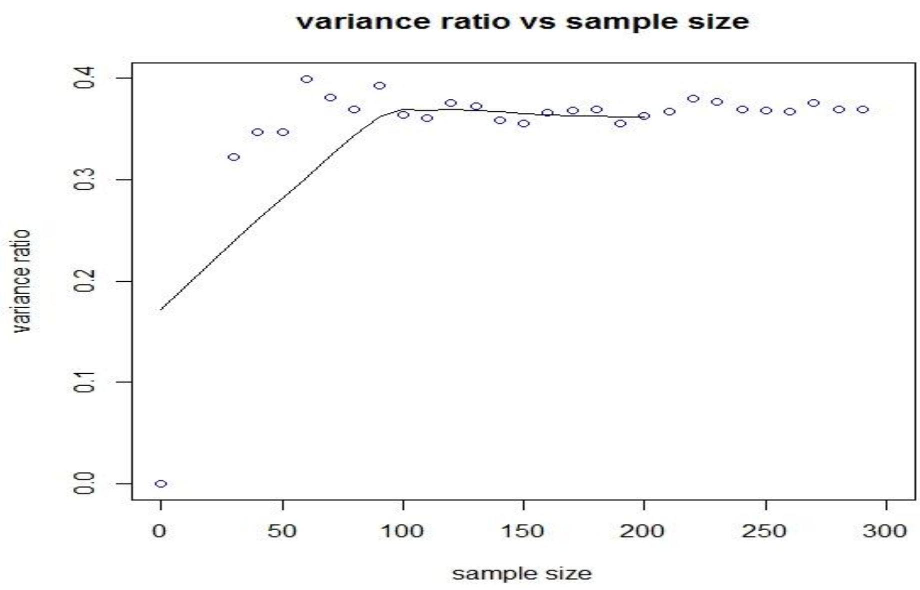

Figure 3: Variance Ratio $\mathrm{var}(y_{np})/\mathrm{var}(y_{ht})$ ***on Linear Model Data***

4.1.2. Results for Nonparametric Estimator y_{np} *on Quadratic Model Data*

Table 2: Nonparametric Estimates for Quadratic Model Data					
sample number	1	2	3	4	5
sample size n	100	100	100	100	100
y_t	6702.63067	6702.63067	6702.63067	6702.63067	6702.63067
y_{np}	6991.0552	6582.4742	7021.3978	6679.41066	6714.58378
y_{ht}	6409.9701	6592.1200	6858.8175	6654.76502	6805.59084
$y_t - y_{np}$	-288.4246	120.1564	-318.7671	23.22000	-11.95311
$y_t - y_{ht}$	292.6606	110.5106	-156.1868	47.86565	-102.96018
r_k	0.00010	0.00010	0.00010	0.00010	0.00010

From table (2), there does not appear to be a noticeable difference in the performances of y_{np} and y_{ht}. In some instances y_{np} has smaller error margins than y_{ht}, while in other samples, y_{ht} has smaller error margins.

This lack of noticeable difference in the performances may point to the robustness of the estimator y_{np}. This is because for quadratic data, y_{np} is actually as misspecified model since y_{np} is obtained from a local linear polynomial model. We note also that the penalty value is 0.00010 for all the samples.

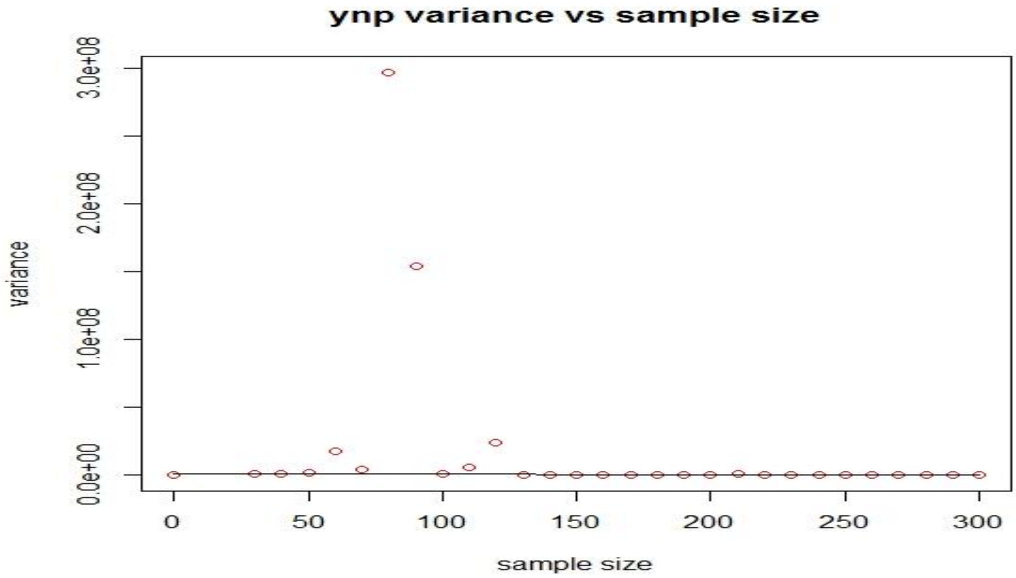

Figure 4: Variance for Estimator y_{np} on Quadratic Model Data

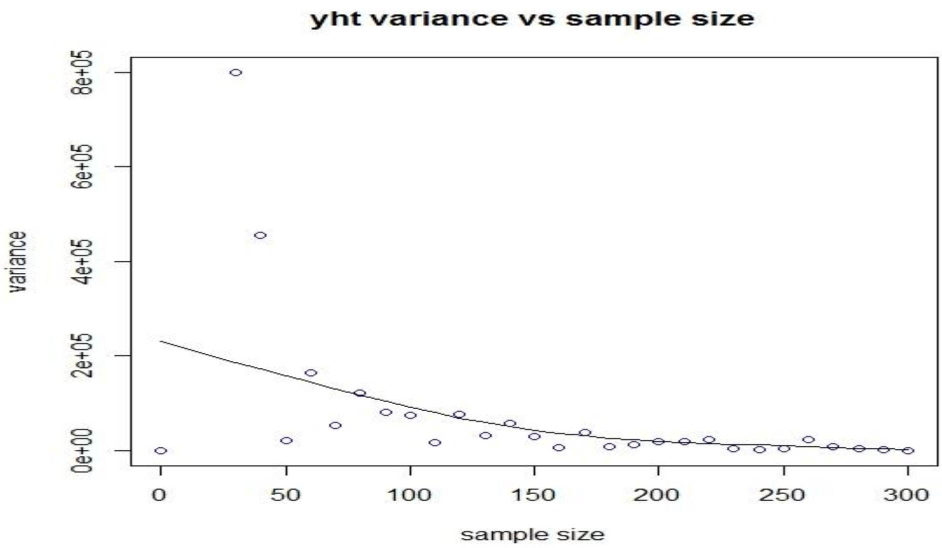

Figure 5: Variance for Horvitz Thompson Estimator y_{np} on Quadratic Model Data

From figure (4), variance for y_{np} does not appear to significantly change as the sample size increases. But for small samples, the variance is more erratic as opposed to large samples. From figure (5), the variance for y_{ht} steadily decrease as the sample size increases. Looking at the scales in figure (4) and figure (5), it can be seen that y_{np} has higher variance than y_{ht} .From figure (6), the ratio $\text{var}(y_{np})/\text{var}(y_{ht})$ tends to a constant, though more erratic for smaller samples. Looking at the scale, we can see that variance for y_{np} dominates variance for y_{ht} .

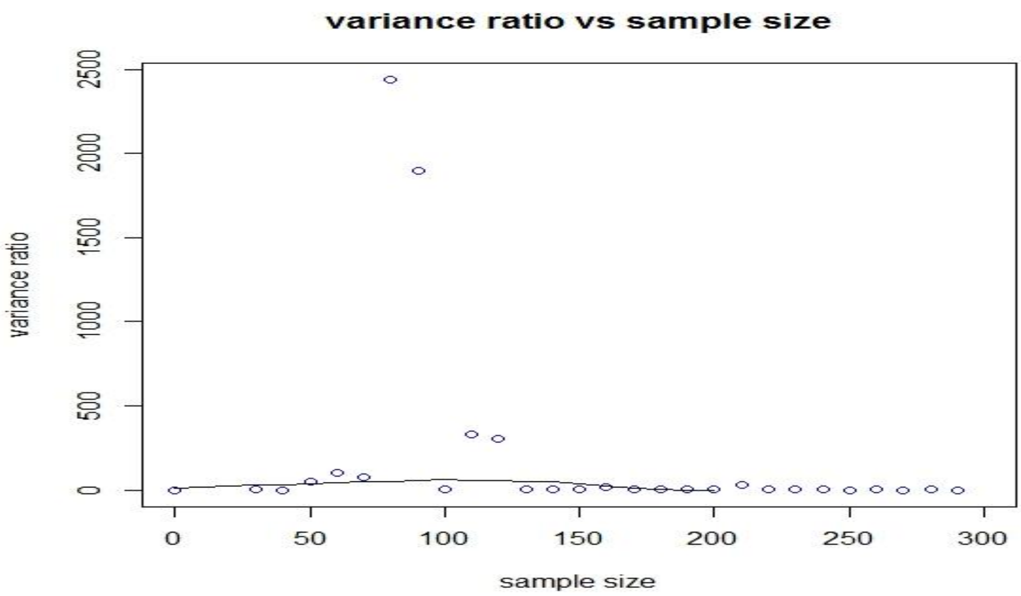

Figure 6: Variance Ratio $\text{var}(y_{np})/\text{var}(y_{ht})$ on Quadratic Model Data

4.2. Analysis of Semiparametric Estimator Results

For semiparametric estimation, the dependent population values y were generated from the linear function $Z\beta' + 2 + 5x$ and quadratic function $Z\beta' + (2+5x)^2$. Z is the matrix (Z_1, Z_2, Z_3) ,where Z_1 is a matrix of 2s with dimension N, the population size. Z_2 is a matrix of alternating 3s,4s and 5s with dimension N, while Z_3 is a matrix of alternating 6s,7s and 8s with dimension N. The vector of coefficients $\beta = (1,2,3)$. We let $y_t = \sum_{i=1}^{N} y_i$ be the actual population total, r_k be the penalty parameter,

and $y_t - y_{sp}$ and $y_t - y_{ht}$ be the errors in the estimation

4.2.1 Results for Semiparametric Estimator y_{sp} on Linear Model Data

Table 3: Semiparametric Estimates for Linear Model Data					
sample number	1	2	3	4	5
sample size n	100	100	100	100	100
y_t	10637.07767	10637.07767	10637.07767	10637.07767	10637.07767
y_{sp}	10653.33589	10592.83642	10772.50201	10656.61600	10620.68442
y_{ht}	10710.11132	10553.27255	10591.87632	0579.65450	10718.60134
$y_t - y_{sp}$	-16.25822	44.24125	-135.42434	-19.53834	16.39324
$y_t - y_{ht}$	-73.03365	83.80512	45.20135	57.42317	-81.52367
r_k	0.00010	0.00010	0.00010	0.00010	0.00010

From table (3), in some samples y_{sp} has larger error margins than y_{ht} , while in other samples, the reverse is true. Convergence is achieved at the same penalty value of 0.00010 and which was the initial penalty value.

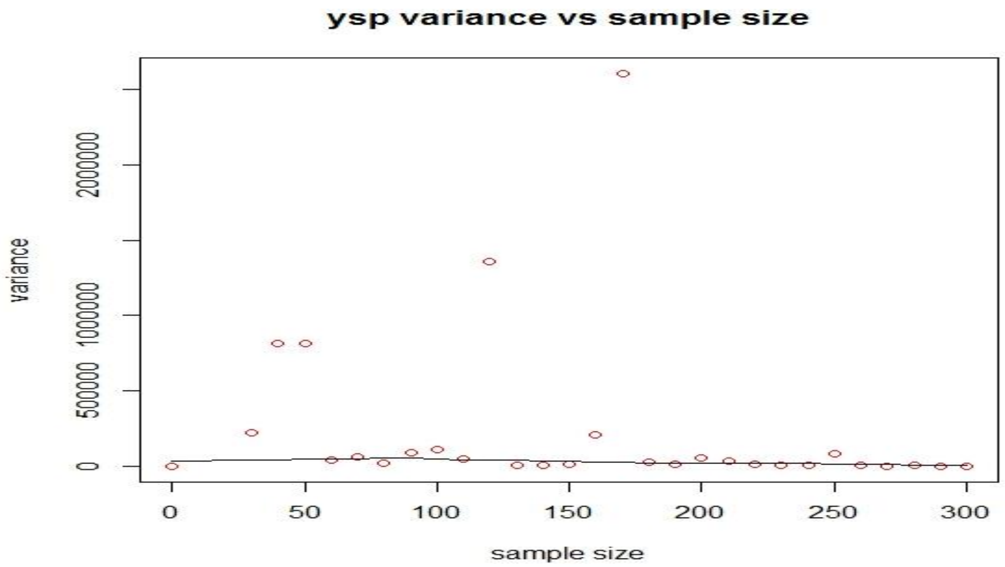

Figure 7: Variance for Estimator y_{sp} on Linear Model Data

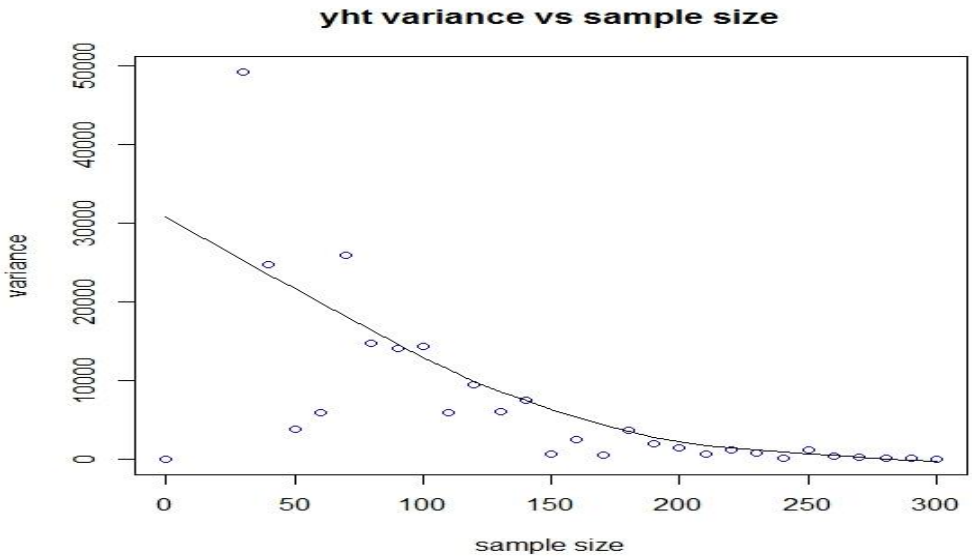

Figure 8: Variance for Horvitz Thompson Estimator y_{ht} on Linear Model Data

In figure (7), though variance for y_{sp} appear to be largely constant when a Lowess line if fitted, a look at individual plots shows higher and more erratic variance for small samples before stabilizing for larger samples. In figure (8), the variance and y_{ht} steadily decrease as the sample size increases. From figure (9), the ratio $\mathrm{var}(y_{sp})/\mathrm{var}(y_{ht})$ is more than one, indicating that y_{np} has higher variance than y_{ht} .

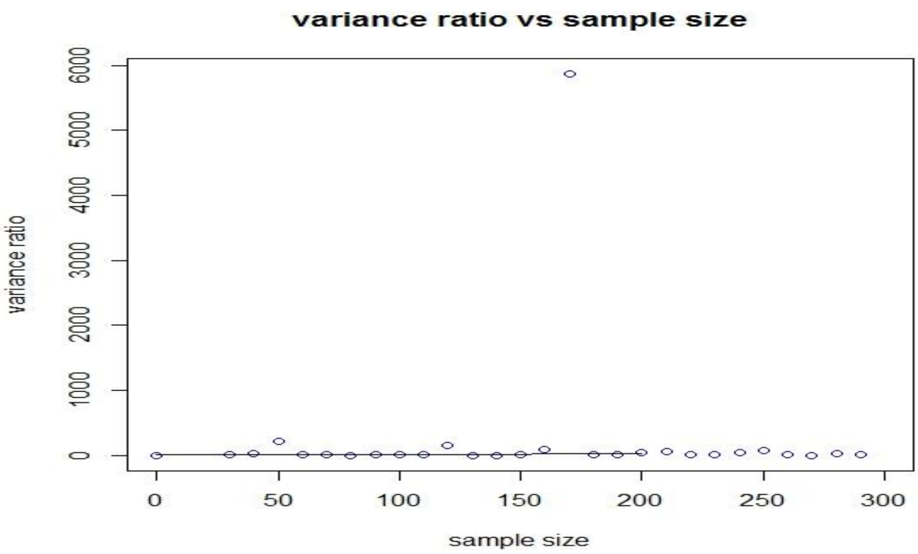

Figure 9: Variance Ratio $\mathrm{var}(y_{sp})/\mathrm{var}(y_{ht})$ *on Linear Model Data*

4.2.2 Results for Semiparametric Estimator y_{sp} *on Quadratic Model Data*

Table 4: Semiparametric Estimates for Quadratic Model Data					
sample number	1	2	3	4	5
sample size n	100	100	100	100	100
y_t	16054.39204	16054.39204	16054.39204	16054.39204	16054.39204
y_{sp}	16083.00298	16386.4447	15530.1848	15850.0349	15939.60759
y_{ht}	15711.98381	16386.9231	16254.8241	16169.8247	16073.76477
$y_t - y_{sp}$	-28.61094	-332.0527	524.2072	204.3571	114.78445
$y_t - y_{ht}$	342.40823	-332.5310	-200.4320	-115.4326	-19.37273
r_k	0.00010	0.00010	0.00010	0.00010	0.00010

From table (4), there is no noticeable difference in the performances of y_{sp} and y_{ht}. In some instances y_{sp} has smaller error margins than y_{ht}, while in other samples, y_{ht} has smaller error margins. This lack of noticeable difference in the performances is evidence to the robustness of the semiparametric estimator y_{sp}. We note also that the penalty value is 0.00010 for all the samples.

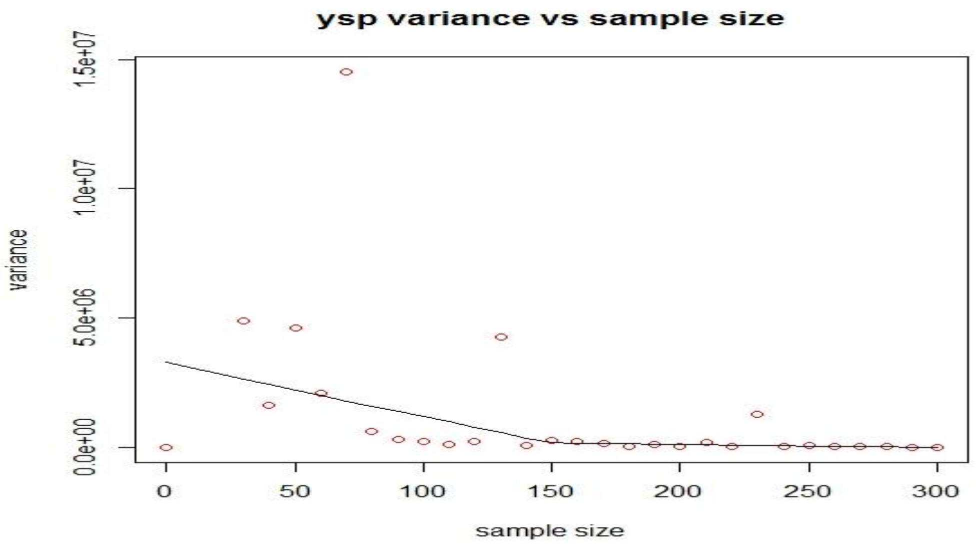

Figure 10: Variance for Estimator y_{sp} on Quadratic Model Data

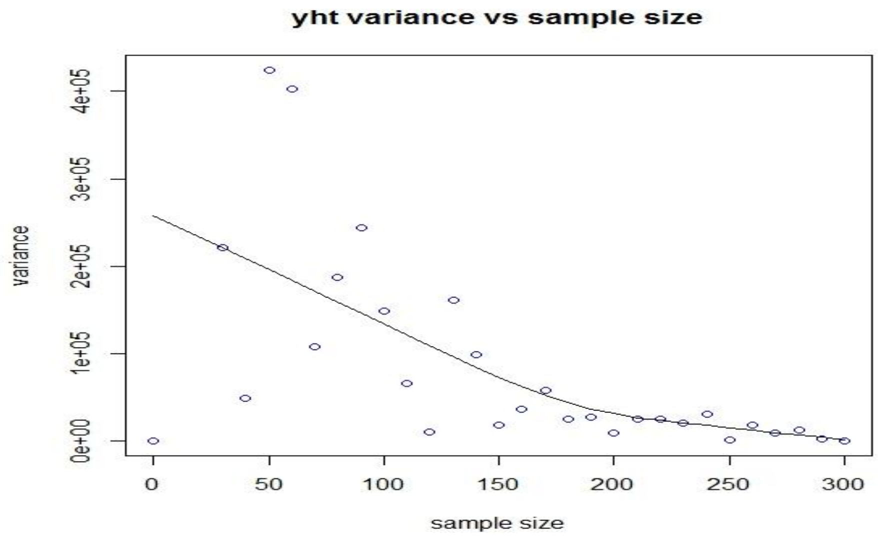

Figure 11: Variance for Horvitz Thompson Estimator y_{ht} on Quadratic Model Data

From figure (10), variance for the semiparametric estimator y_{sp} is higher and more erratic for small samples before stabilizing almost to a constant for lager samples. From figure (11), variance for y_{ht} steadily decrease as the sample size increases. From figure (12), the ratio $\text{var}(y_{sp})/\text{var}(y_{ht})$ show Cleary y_{sp} has larger variance than y_{ht}.

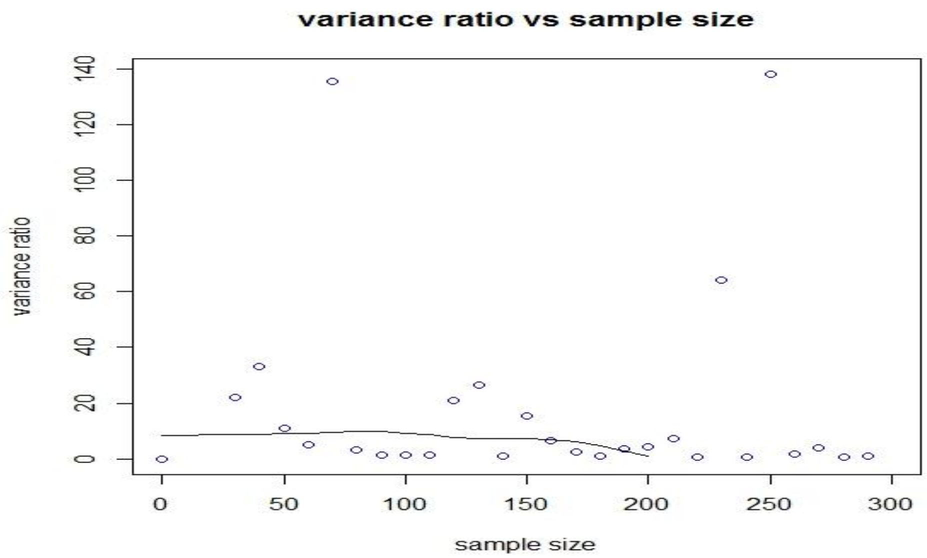

Figure 12: Variance Ratio $\text{var}(y_{np})/\text{var}(y_{ht})$ on Quadratic Model Data

5. Conclusion

We conclude that when the nonparametric model is correctly specified for the data, the nonparametric estimator y_{np} is quite accurate, more than the Horvitz Thompson design estimator y_{ht}. When the nonparametric model is misspecified for the data, the nonparametric estimator y_{np}, though a bit less efficient than the Horvitz Thompson design estimator y_{ht}, still yields quite reliable estimates. This shows that y_{np} is a robust estimator. The semiparametric estimator y_{sp} is also a very robust estimator giving estimates that are very close to those of Horvitz Thompson design estimator even when the nonparametric model component of the semiparametric estimator is misspecified.

References

[1] Box, G.E.P. (1980). Sampling and Baye's Inference in Scientific Modeling and Robustness. *Journal of the Royal Statistical Society, series A, 143,383-40.*

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

[3] Breidt, F.J. Opsomer, J.D. Alicia, A.J. and Ranalli, G. (2007). Semiparametric Model Assisted Estimation for Natural Resource Surveys. *Statistics Canada, Catalogue No. 12-001.*

[4] Deville, J.C. & Sarndal C.E. (1992), "Calibration Estimators in Survey Sampling", *Journal of the American Statistical Association,* 87,376-82.

[5] Frank E.C. Jorge N. (2007). "Flexible Penalty Functions for Nonlinear Constrained Optimization*", IMA Journal of Numerical Analysis*

[6]Hastie, T. and Tibshirani, R. (1987). Generalized Addidative Models: Some Applications. *Journal of the American Statistical Association, 82,371-386*

[7] Kihara, P.N.(2017). Calibration Estimators by Penalty Function Method. *Mathematical Theory and Modeling, Vol 7, No 6*

[8] Ozgur Y. (2005) Penalty Function Methods for Constrained Optimization with Generic Algorithms*", mathematical and Computation Applications, Vol 10, No 1, Page 45-56*

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

[10] Sarndal, C.E. (1980). On -Inverse weighting versus best Linear Unbiased Weighting in Probability Sampling. *Biometrika, 67,639-650*

[11] Silverman, B.W. (1985). Some Aspects of the Spline Smoothing Approach to Nonparametric Regression Curve Fitting. *Journal of the Royal Statistical Society,Series B, 47, 1-21*

[12] Simonof, J. (1996). Smoothing Methods in Statistics. *New York: Springer*

[13] Thompson M.E. (1997), "Theory of Sample Surveys", *Chapman Hall, London*

[14] Wu, C, & Sitter, R.R. (2001), "A Model Calibration Approach to Using Complete Auxiliary Information from Survey Data", Journal *of American Statistical Association,* 96, 185-93.