Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.7, No.6, 2017

Calibration Estimators by Penalty Function Method

Pius Nderitu Kihara
Department of Statistics and Actuarial Sciences, Technical University of Kenya

Abstract

Estimation of finite population total using calibration has been considered by several authors. A distance measure is minimized subject to some calibration constraints, usually by way of introducing langrage equation whose solution gives the design weights used in estimation of population total. Sometimes a solution to the langrage constants does not exist. In this paper, we have considered the calibration problem as a nonlinear constrained minimization problem, which we transform to an unconstrained optimization problem using penalty functions. The design weights are obtained iteratively in a numerical manner. We show that the resulting estimator is more accurate than the popular Horvitz Thompson design estimator

Keywords: calibration, interior penalty function, exterior penalty function

1. Introduction

The notion of calibration was introduced by Deville and Sarndal [1] in the context of using auxiliary information from survey data. Suppose $U = \{1,2,...,N\}$ is the set of labels for the finite population. Let (y_i, x_i) be the respective values of the study variable y and the auxiliary variable x attached to the ith unit. If we let $s = \{1,2,...,n\}$ be the set of sampled units under a general sampling design p, and let $\pi_i = p(i \in s)$ be the first order inclusion probabilities, then the conventional calibration estimator for the population total y_t is defined by $\hat{y}_t = \sum_{i=1}^{n} w_i y_i$ where $w_i's$ are design weights which are as close as possible to $d_i = \pi^{-1}$ and are obtained by minimizing a given distance measure between $w_i's$ and $d_i's$ subject to some constraints. A common distance measure is the chi-square distance measure below.

$$\Phi = \sum_{i \in s} \frac{(w_i - d_i)^2}{q_i d_i} \tag{1}$$

where $q_i's$ are some constants unrelated to $d_i's$. Other distance functions were considered by Deville et al. [2], Singh and Mohl [7] as well as Stukel et al. [8]. Deville and Sarndal [1] considered the calibration constraint

$$\sum_{i=1}^{n} w_i x_i = \sum_{i=1}^{N} x_i \tag{2}$$

Minimizing (1) subject to (2) by way of Lagrange equation, they obtained the equation

$$w_i = d_i = \left\{ \frac{d_i q_i x_i}{\sum_{i=1}^{n} d_i q_i x_i^s} \right\} \left\{ \sum_{i=1}^{N} x_i - \sum_{i=}^{n} d_i x_i \right\} \tag{3}$$

Wu and Sitter [10] introduced yet another calibration constraint

$$\sum_{i=1}^{n} w_i = N \tag{4}$$

and minimizing (1) subject to (2) and (4) and by way of Lagrange equation, they obtained

$$w_i = d_i + (N - \sum_{i\in s} d_i)\left\{\frac{d_i q_i}{\sum_{i\in s} d_i q_i} - \left\{\frac{d_i q_i\left(x_i - \frac{\sum_{i\in s} d_i q_i x_i}{\sum_{i\in s} d_i q_i}\right)\left(1 - \frac{\sum_{i\in s} d_i q_i x_i}{\sum_{i\in s} d_i q_i}\right)}{\sum_{i\in s} d_i q_i\left(x_i - \frac{\sum_{i\in s} d_i q_i x_i}{\sum_{i\in s} d_i q_i}\right)^2}\right\}\right\}$$

$$+\left(\sum_{i\in U} x_i - \sum_{i\in s} d_i x_i\right)\left\{\frac{d_i q_i\left(x_i - \frac{\sum_{i\in s} d_i q_i x_i}{\sum_{i\in s} d_i q_i}\right)\left(1 - \frac{\sum_{i\in s} d_i q_i x_i}{\sum_{i\in s} d_i q_i}\right)}{\sum_{i\in s} d_i q_i\left(x_i - \frac{\sum_{i\in s} d_i q_i x_i}{\sum_{i\in s} d_i q_i}\right)^2}\right\}$$

Obtaining the weights $w_i's$ as derived in (3) and (5) and hence obtaining the estimator $\hat{y}_t = \sum_{i=1}^{n} w_i y_i$ is quite tedious and may not be feasible in day to day applications. Also, the solution for λ may not always exist in which case Deville and Sarndal [1] recommend that λ be set to 0. Ralf et al [5] considered transforming the calibration problem for general functions f into a nonlinear equation depending on the Lagrange multiplier λ and since the mapping *was* no longer differentiable, they used, semismooth Newton method to solve the resulting equation numerically. We propose use of penalty function to obtain the design weights $w_i's$, a procedure that does not require introduction of langrage multipliers.

2. Penalty Function Method

The penalty function methods transform the basic constrained optimization problem into an unconstrained optimization problem. Consider an optimization problem of the form

$$\begin{aligned} &\min imize \;\; f(X) \;\; subject\ to \\ &\begin{cases} g_j(X) \le 0, \;\; j = 1,2,..., m \;\; and \\ h_j(X) = 0, \;\; j = 1,2,..., p \end{cases} \end{aligned} \tag{6}$$

By the interior penalty function method (also called barrier method), an unconstrained problem may be constructed as follows.

$$\phi(X, r_k) = f(X) + \psi_1(r_k, g_j(X)) + \psi_2(r_k, l_j(X)) \tag{7}$$

where $\psi_1(r_k, g_j(X))$ and $\psi_2(r_k, l_j(X))$ are penalty functions and which are such that $\psi_i, (i = 1,2)$ is continuous, $\psi_i(r_k, t) \ge 0$ for all r_k and $t \in \Re^n$,and $\psi_i(r_k, t)$ is strictly increasing for $r_k > 0$ and $t > 0$. A common form similar to the one discussed in Rao [6] is given below

$$\phi(X, r_k) = f(X) - r_k \sum_{j=1}^{m} \frac{1}{g_j(X)} + H(r_k) \sum_{j=1}^{p} l_j^q(X) \tag{8}$$

where $H(r_k)$ is some function of the parameter r_k tending to infinity as r_k tends to zero and so that $\sum_{j=1}^{p} l_j^q(X)$ also tend to zero. A common choice for value of q is 2. Also, the function ϕ will always be greater than f since $g_j(X)$ is negative for all feasible points X. 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 [3] have discussed flexible ways of choosing the penalty. In an iterative process, the unconstrained minimization of ϕ is started from any feasible solution for the inequality constraint but not necessarily so for the 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 [4] is that if the constraint is violated, that means $g_j(X) > 0$ or $l_j(X) \neq 0$, a big term will be added to ϕ function such that the solution is pushed back towards the feasible region. In the minimization of ϕ, for the solution to be the global minimum, we must have that $f(X)$, $g_j(X)$ $j = 1,2,..., m$, and $\sum_{j=1}^{p} l_j^q(X)$ being convex and we must also have one of the functions $f(X)$, $g_j(X)$ $j = 1,2,..., m$ and $\sum_{j=1}^{p} l_j^q(X)$ being strictly convex. See Rao [6].

Using the exterior penalty function method, a solution to the constrained problem (6) would be given by

$$\phi(X, r_k) = f(X) + r_k \sum_{j=1}^{m} \langle g_j(X) \rangle^q + H(r_k) \sum_{j=1}^{p} l_j^q(X) \tag{9}$$

where $\langle g_j(X) \rangle = \max(g_j(X), 0)$. Also, as $k \to \infty$, $r_k \to \infty$ and $H(r_k) \to \infty$. For exterior penalty function method, in the iterative minimization of ϕ, the starting point X does not have to be feasible. Looking at the equations (8) and (9), we see that, when the optimization problem has only the equality constraints, both interior and exterior penalty functions yield a function of the form

$$\phi(X, r_k) = f(X) + H(r_k) \sum_{j=1}^{p} l_j^q(X) \tag{10}$$

Setting $H(r_k) = r_k$, where $r_k \to \infty$ as $k \to \infty$, and setting $q = 2$ we have from (10)

$$\phi(X, r_k) = f(X) + r_k \sum_{j=1}^{p} l_j^2(X) \tag{11}$$

3. Penalty Function Method of Estimating Population Total

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. To obtain design weights, we reduce the chi-square distance measure (1) subject to the constraints (2) considered by Deville and Sarndal [1]. Using the penalty function method we obtain the penalty function

$$\phi_1(w, r_k, x) = \sum_{i=1}^{n} \frac{(w_i - d_i)^2}{q_i d_i} + r_k \left[\sum_{i=1}^{n} w_i x_i - \sum_{i=1}^{N} x_i \right]^2 \tag{12}$$

where r_k is some penalty. We need to find the weights w_i that minimize the penalty function (12) above.

Differentiating (12) partially with respect to w_i we have

$$\phi_1^1(w_i, r_k, x) = \frac{2(w_i - d_i)}{q_i d_i} + 2 r_k x_i \left[\sum_{j=1}^{n} w_j x_j - \sum_{j=1}^{N} x_j \right] \tag{13}$$

We equate (13) to zero and solve for w_i to obtain

$$w_i = \frac{d_i - r_k x_i q_i d_i \left(\sum_{\substack{j=1 \\ j \neq i}}^{n} w_j x_j - \sum_{j=1}^{N} x_j \right)}{1 + r_k \left(x_i^2 q_i d_i \right)} \tag{14}$$

We have the following estimator of population total

$$\hat{y}_{t1} = \sum_{i=1}^{n} w_i y_i = \sum_{i=1}^{n} \frac{y_i d_i}{1 + r_k \left(x_i^2 q_i d_i \right)} - \sum_{i=1}^{n} \frac{r_k x_i q_i d_i y_i \left(\sum_{\substack{j=1 \\ j \neq i}}^{n} w_j x_j - \sum_{j=1}^{N} x_j \right)}{1 + r_k \left(x_i^2 q_i d_i \right)} \tag{15}$$

Minimizing (1) subject to both (2) and (4) as considered by Wu and Sitter [10], we have the penalty function

$$\phi_2(w, r_k, x) = \sum_{i=1}^{n} \frac{\left(w_i - d_i \right)^2}{q_i d_i} + r_k \left[\sum_{i=1}^{n} w_i x_i - \sum_{i=1}^{N} x_i \right]^2 + r_k \left[\sum_{i=1}^{n} w_i - N \right]^2 \tag{16}$$

Differentiating (16) partially with respect to w_i we have

$$\phi_2^1(w_i, r_k, x) = \frac{2\left(w_i - d_i \right)}{q_i d_i} + 2 r_k x_i \left[\sum_{j=1}^{n} w_j x_j - \sum_{j=1}^{N} x_j \right] + 2 r_k \left[\sum_{i=1}^{n} w_i - N \right] \tag{17}$$

Equating (17) to zero and solving for w_i we have

$$w_i = \frac{d_i - r_k q_i d_i \left(\sum_{\substack{j=1 \\ j \neq i}}^{n} w_j \left(x_i x_j + 1 \right) - \sum_{j=1}^{N} \left(x_i x_j - 1 \right) \right)}{1 + r_k \left(\left(x_i^2 + 1 \right) q_i d_i \right)} \tag{18}$$

We therefore have the following estimator of population total

$$\hat{y}_{t2} = \sum_{i=1}^{n} w_i y_i = \sum_{i=1}^{n} \frac{y_i d_i}{1 + r_k\left((x_i^2 + 1) q_i d_i\right)} - \sum_{i=1}^{n} \frac{r_k q_i d_i y_i \left(\sum_{\substack{j=1 \\ j \neq i}}^{n} w_j (x_i x_j + 1) - \sum_{j=1}^{N} (x_i x_j - 1) \right)}{1 + r_k\left((x_i^2 + 1) q_i d_i\right)} \tag{19}$$

The beauty with this approach is that to obtain the weights w_i, $(i = 1,2,..., n)$, we solve the penalty functions (12) and (16) 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 (2) and (4) are equality constraints, we need not start with a feasible guess for w_i. We appeal to Newton method of unconstrained optimization. See Rao [6].

Let $W = \{w_1, w_2,..., w_n\}$ be the set of the weights. We need to obtain W^* such that

$$g(W^*) = \left[\phi'(w_1, r_k, x),..., \phi'(w_n, r_k, x)\right]' = 0 \tag{20}$$

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

$$g(W^*) = g(W_i + Z) = g(W_i) + J_{W_i} Z + ... \tag{21}$$

By neglecting the higher order terms in (21) and setting $g(W^*) = 0$ we obtain

$$g(W_i) + J_{W_i} Z = 0 \tag{22}$$

Where J_{W_i} is the matrix of second derivatives evaluated at W_i. In general, when we consider the constraint (2) alone, then J is a n by n matrix with $i = 1,2,..., n$ rows and $j = 1,2,..., n$ columns. It has diagonal elements $\frac{2}{q_i d_i} + 2r_k x_i^2$ and elements $2r_k x_i x_j$ elsewhere. If we consider both constraints (2) and (4), then J has diagonal elements $\frac{2}{q_i d_i} + 2r_k (x_i^2 + 1)$ and elements $2r_k (x_i x_j + 1)$ elsewhere. If J_{W_i} is nonsingular, then, from the set of linear equations (22) we have for vector Z

$$Z = J_{W_i}^{-1} g(W_i) \quad . \tag{23}$$

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

$$W_{i+1} = W_i + S_i = W_i - J_{W_i}^{-1} g(W_i) \quad (24)$$

The sequence of the points $W_1, W_2, ..., W_{i+1}$ eventually converges to the actual solution W^*. Since our penalty functions (12) and (16) are quadratic, we find the minimum in a single step using equation (24) since the Taylor's series expansion is exact.

Now, if we let w_κ^* 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 $\phi(w, r_k, x) = \phi(w, r_{k+1}, x)$ for some specified accuracy level. The accuracy level may for example be, to certain decimal points or significance level. The penalty values are set such that the starting point $r_1 > 0$ and $r_{k+1} = cr_k$, where $c > 1$. We can now generalize our estimator for the population total as

$$\hat{y}_{tv} = \sum_{i=1}^{n} w_i y_i = W^* Y_s, \ v = 1,2 \quad (25)$$

where $Y_s = (y_1, y_2, ..., y_n)$ is the sample from the population of y and $v = 1$ if $w_i's$ are obtained as defined in (14) and $v = 2$ if $w_i's$ are obtained as defined in (18).

4. Empirical Analysis

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 both populations, the estimators exhibited same properties. We will therefore report the results for the linear function $y = 2 + 5x$. 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. In section 4.1, we report on the performance of estimator y_{t1} and compare its performance with that of Horvitz Thompson estimator $y_{ht} = \sum_{i=1}^{n} y_i d_i$ discussed in Thompson [9], while in section 4.2, we report on the results for estimator y_{t2} and again compare with Horvitz Thompson estimator.

4.1 Results for Estimator y_{t1}

We let $y_t = \sum_{i=1}^{N} y_i$ be the actual population total, r_k be the penalty parameter, and $y_t - y_{t1}$ and $y_t - y_{ht}$ be the errors in the estimation.

Table 1: Population Total Estimates and the Errors in the Estimation for y_{t1} and y_{ht}					
sample number	1	2	3	4	5
sample size n	100	100	100	100	100
y_t	1361.13529	1361.13529	1361.13529	1361.13529	1361.13529
y_{t1}	1349.07154	1376.48127	1360.3058151	1400.78331	1364.304392
y_{ht}	1348.87572	1376.73391	1360.2924510	1401.46633	1364.356816
$y_t - y_{t1}$	12.06375	-15.34597	0.8294757	-39.64802	-3.169101
$y_t - y_{ht}$	12.25957	-15.59861	0.8428398	-40.33104	-3.221525
r_k	0.00010	0.00010	0.00010	0.00010	0.00010

Looking at table (1), we see that the estimators y_{t1} and y_{ht} have almost equal error margins, but consistently, y_{t1} has a smaller error margin. For all the samples, convergence is achieved at the same penalty value of 0.00010 and which was the initial penalty value. For different sample sizes, we observed that the penalty value ranged between 0.00010 and 0.0013 with no particular pattern that could be attributed to the sample size. In most of the cases however, the penalty was 0.00010.

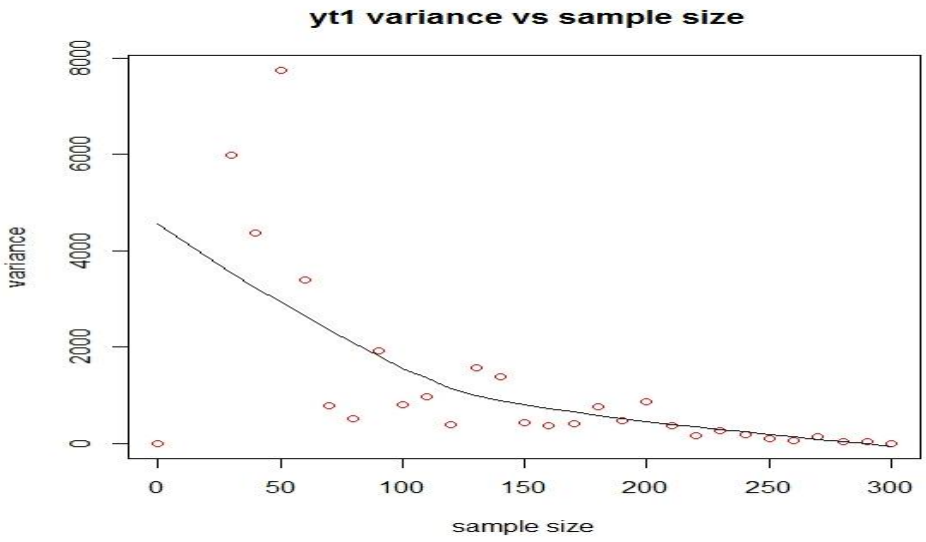

Fig 1: Variance for Estimator y_{t1}

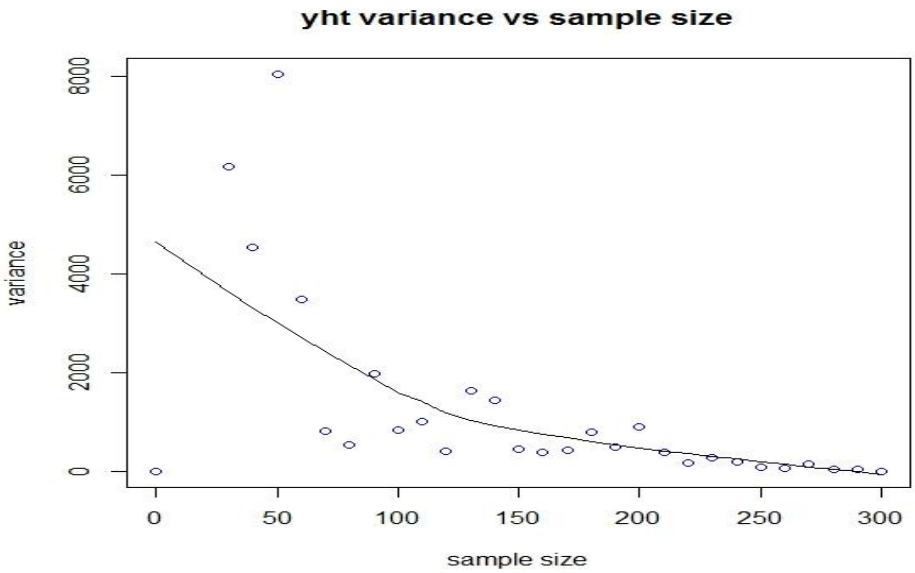

Fig 2: Variance for Horvitz Thompson Estimator y_{ht}

In Fig (1) and Fig (2), the variances for y_{t1} and y_{ht} have a similar pattern. As the sample size increases, the variance decreases. From Fig (3), the ratio $\text{var}(y_{t1})/\text{var}(y_{ht})$ settles almost to a constant as the sample size increases. The constant is found to be about 0.97, which indicates that y_{t1} has a smaller variance than y_{ht}, and which is consistent with the smaller error margin for y_{t1} as seen in table (1).

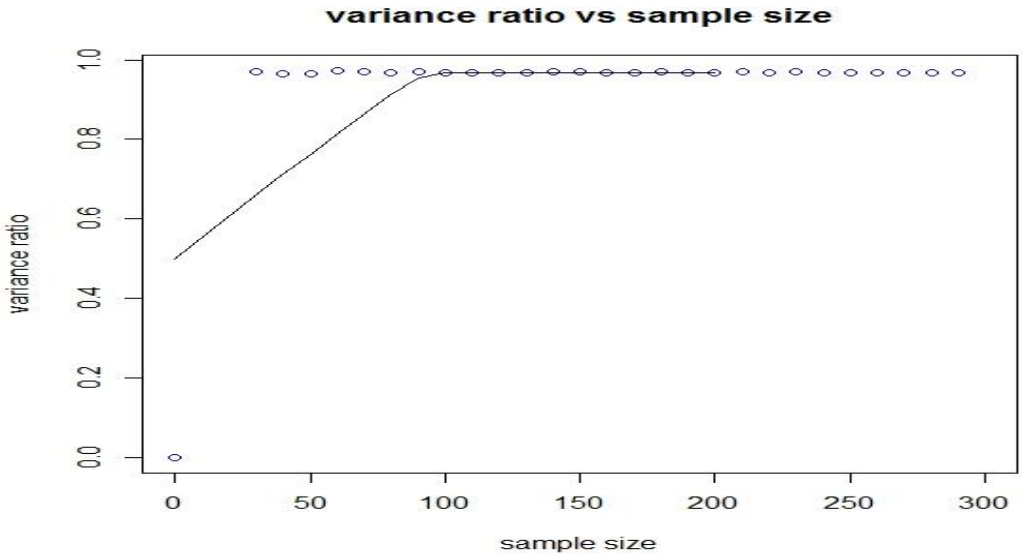

Fig 3: Variance Ratio $\text{var}(y_{t1})/\text{var}(y_{ht})$

4.2 Results for Estimator y_{t2}

Table 2: Population Total Estimates and the Errors in the Estimation for y_{t2} and y_{ht}

sample number	1	2	3	4	5
sample size n	100	100	100	100	100
y_t	1384.49498	1384.49498	1384.49498	1384.49498	1384.49498
y_{t2}	1400.01439	1406.98567	1398.03903	1321.13222	1330.37056
y_{ht}	1400.27413	1407.37208	1398.26738	1320.20986	1329.55309
$y_t - y_{t2}$	-15.51940	-22.49068	-13.54405	63.36276	54.12442
$y_t - y_{ht}$	-15.77915	-22.87709	-13.77240	64.28512	54.94189
r_k	0.00010	0.00010	0.00010	0.00010	0.00010

In table (2) $y_t - y_{t2}$ and $y_t - y_{ht}$ are the errors in the estimation. From the table, y_{t2} and y_{ht} error margins are quite close, but with y_{t2} consistently, having the smaller error margin. Also the penalty value is 0.00010 for all the samples.

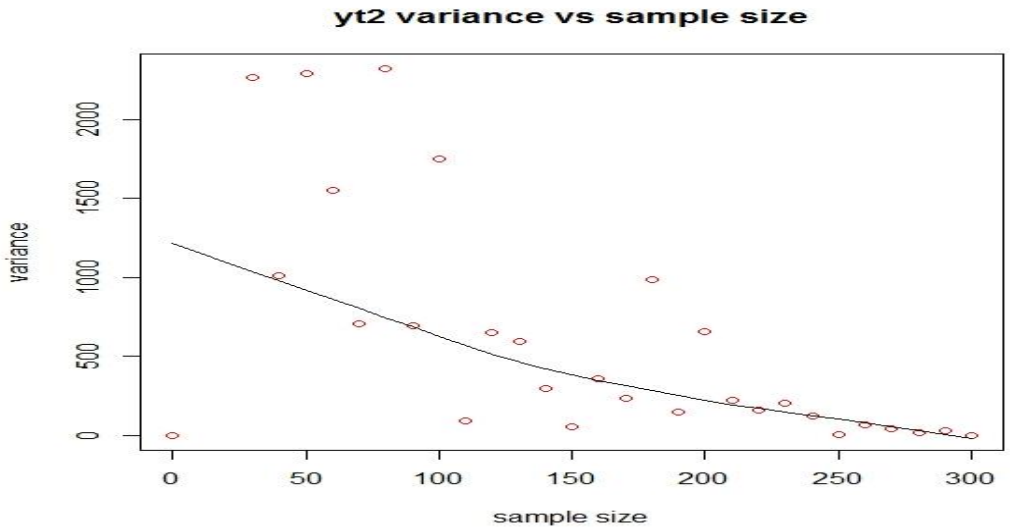

Fig 4: Variance for Estimator y_{t2}

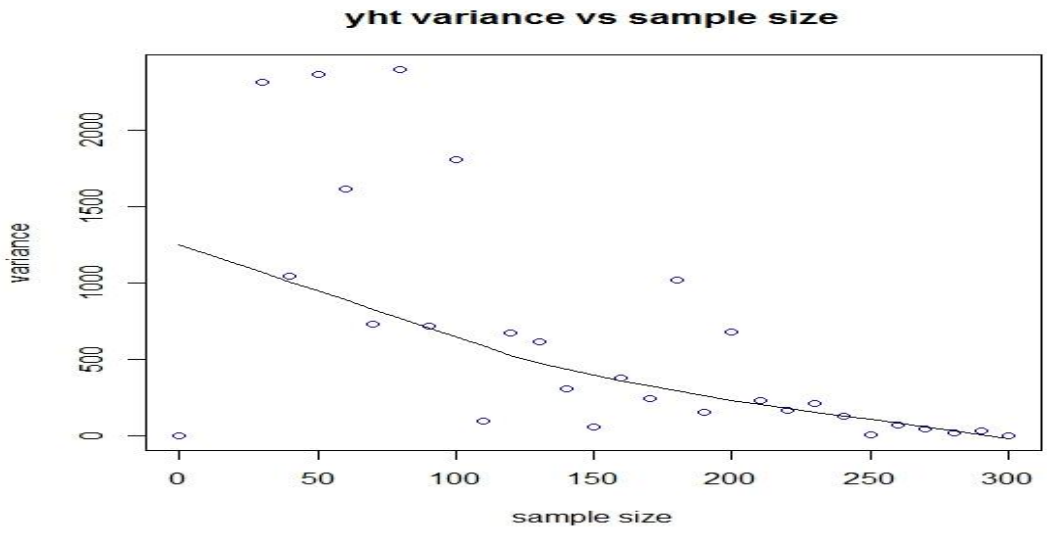

Fig 5: Variance for Horvitz Thompson Estimator y_{ht}

Fig (4) and Fig (5), show similar patterns for the variances of y_{t2} and y_{ht}. As the sample size increases, the variances are decreasing. From Fig (6), the ratio $\text{var}(y_{t2})/\text{var}(y_{ht})$ tends to a constant, estimated to about 0.97 and which indicates that y_{t2} has a smaller variance than y_{ht}.

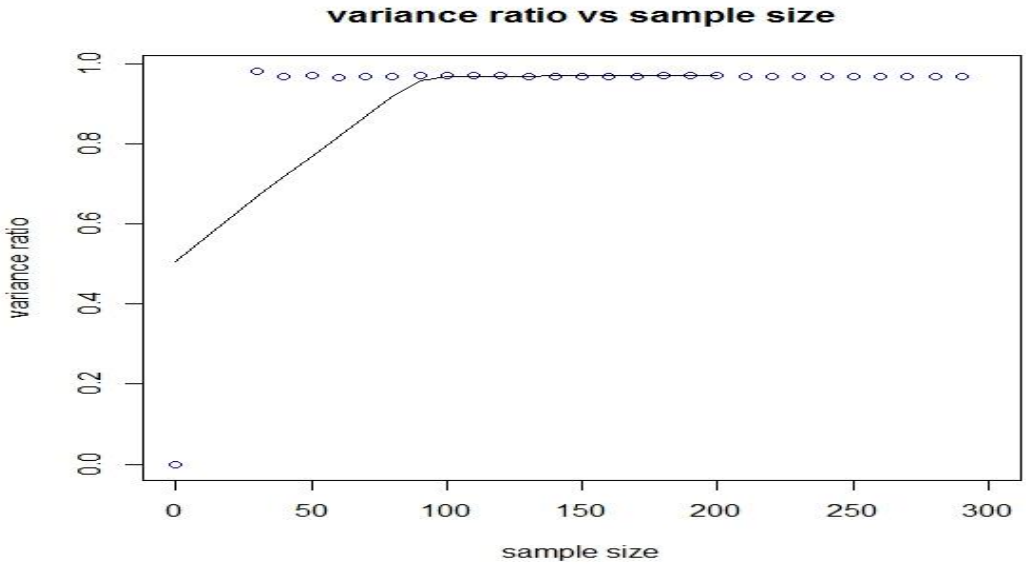

Fig 6: Variance Ratio $\text{var}(y_{t2})/\text{var}(y_{ht})$

5. Conclusion

We conclude that both estimators y_{t1} and y_{t2} are more accurate than the Horvitz Thompson design estimator y_{ht} since they both have smaller margin of errors and smaller variance than y_{ht}. From the variance ratios $\text{var}(y_{t1})/\text{var}(y_{ht})$ and $\text{var}(y_{t2})/\text{var}(y_{ht})$ both of which are about 0.97, we conclude that $\text{var}(y_{t1})$ and $\text{var}(y_{t2})$ are not significantly different and that estimators y_{t1} and y_{t2} are not different in terms of the

accuracy in estimation. We conclude that the estimators y_{t1} and y_{t2} are consistent in the sense that as the sample size increases, their variances tend to zero.

.

References

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

[2] Deville, J.C., Särndal, C.E., Sautory, O (1993). "Generalized raking procedures in survey sampling", *J. Am. Stat .Assoc*. **88**, 1013–1020

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

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

[5] Ralf T. M., Ekkehard W. S., Matthias W. (2012), "Calibration of estimator-weights via semismooth Newton method**",** *J Glob Optim,* 52:471–485

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

[7] Singh, A., Mohl, C. (1996), "Understanding Calibration Estimators in Survey Sampling", *Surv. Methodol*. **22**, 107–115

[8] Stukel, D., Hidiroglou, M., Särndal, C.E. (1996), "Variance Estimation for Calibration Estimators: A Comparison of Jackknifing Versus Taylor Linearization", *Surv. Methodol*, **22**, 117–125 (1996)

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

[10] 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.