

Comparison of Various Methods for Estimating Finite Population Total in Survey Sampling When Study Variable and Auxiliary Variable are Inversely Related

S. Kumar, B. V. S. Sisodia*

Department of Agricultural Statistics, Narendra Deva University of Agriculture and Technology, Kumarganj, Faizabad – 224229, (U.P.), India

Received 6 June 2018, accepted in final revised form 26 September 2018

Abstract

In the present paper, a model based calibration estimator of population total has been developed when study variable y and auxiliary variable x are inversely related. The relative performance of the proposed model based calibration estimator in comparison to model based estimator, the usual regression estimator and calibration based regression estimator have been examined by conducting a limited simulation study. In view of the results of the simulation study, it has been found that model based calibration estimator has outperformed the other estimators. However, calibration based regression estimator was found to be close to the model based calibration estimator.

Keywords: Auxiliary information; Super population model; Calibration estimator; Model based calibration estimator; Model based estimator.

© 2019 JSR Publications. ISSN: 2070-0237 (Print); 2070-0245 (Online). All rights reserved.
doi: <http://dx.doi.org/10.3329/jsr.v11i1.36922>

J. Sci. Res. **11** (1), 15-22 (2019)

1. Introduction

The auxiliary information is used to improve the precision of the estimates of the population parameters such as population mean, population total, population variance etc. in finite population survey sampling. Various estimation approaches for estimating finite population total using information on auxiliary variables have been resorted. Most common methods of estimation are ratio and regression estimators, model based estimator by Royall and Herson [2], calibration estimator by Deville and Särndal [5] and model based calibration estimator by Wu and Sitter [6]. Recently, some research workers like Sud *et al.* [11], Mourya *et al.* [12] and Sandeep Kumar *et al.* [13,14] have contributed significantly in calibration approach based estimation in finite population survey sampling. Following Royall and Herson [2], a model based unbiased estimator of

$Y = \sum_{i=1}^N y_i$ under the model

* Corresponding author: bvssisodia@gmail.com

$$Y_i = \alpha + \frac{\beta}{x_i} + e_i, \quad i = 1, 2, 3, \dots, N, \quad V(Y_i) = \sigma^2, \tag{1}$$

is given by

$$\hat{Y}_M = N\bar{y} + N\hat{\beta} \left(\frac{1}{N} \sum_{i=1}^N \frac{1}{x_i} - \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} \right) \tag{2}$$

where $\hat{\beta} = \frac{n \sum_{i=1}^n \frac{y_i}{x_i} - \sum_{i=1}^n \frac{1}{x_i} \sum_{i=1}^n y_i}{n \sum_{i=1}^n \frac{1}{x_i^2} - \left(\sum_{i=1}^n \frac{1}{x_i} \right)^2}$, y_i are realized values of independent random variables

y_i 's and \bar{y} is sample mean for given sample s of size n . The estimator in (2) is in fact the usual regression estimator of Y when $\frac{1}{x_i}$ is considered as auxiliary variable instead of

x_i . The model variance of \hat{Y}_M is obtained following Sukhatme *et al.* [4]

$$V(\hat{Y}_M) = \sigma^2 \left[\sum_{i \in s} \left\{ \frac{N-n}{n} - \frac{N \left(\frac{1}{N} \sum_{i=1}^N \frac{1}{x_i} - \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} \right) \left(\frac{1}{x_i} - \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} \right)}{\sum_{i \in s} \left(\frac{1}{x_i} - \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} \right)^2} \right\} + (N-n) \right] \tag{3}$$

Sud *et al.* [11] developed calibration based regression type estimator of finite population total of the study variate y , when study variable is inversely related to an auxiliary variable x . Their estimator and its variances etc. are briefly presented here.

Consider that a sample s of size n is drawn from the population $U = [U_1, U_2, U_3, \dots, U_N]$ according to sampling design $P(\cdot)$. Let π_i and π_{ij} be the inclusion probability of i^{th} unit and joint inclusion probability of i^{th} and j^{th} unit, respectively, in the sample s . Suppose that the information on an auxiliary variable x related to the study variate y is known for all $i = 1, 2, 3, \dots, N$. Hence $X = \sum_{i=1}^N x_i$ and

$X^* = \sum_{i=1}^N \frac{1}{x_i}$ are also known. Replacing design weight $d_i = \frac{1}{\pi_i}$ in the Horvitz-Thompson

estimator of Y i.e. $\hat{Y}_{HT} = \sum_{i \in s} d_i y_i$, by a calibration weight w subject to the calibration

constraints $\sum_{i=1}^n \frac{w_i}{x_i} = \sum_{i=1}^N \frac{1}{x_i}$ and $\sum_{i=1}^n w_i = \sum_{i=1}^n d_i$

They developed calibration based regression type estimator as

$$\hat{Y}_C = \hat{Y}_{HT} + \hat{B} \left(\sum_{i=1}^N \frac{1}{x_i} - \sum_{i=1}^n \frac{d_i}{x_i} \right), \quad \hat{B} = \frac{\left(\sum_{i=1}^n \frac{d_i y_i}{x_i} \sum_{i=1}^n d_i - \sum_{i=1}^n d_i y_i \sum_{i=1}^n \frac{d_i}{x_i} \right)}{\left[\sum_{i=1}^n \frac{d_i}{x_i} \sum_{i=1}^n d_i - \left(\sum_{i=1}^n \frac{d_i}{x_i} \right)^2 \right]} \tag{4}$$

Under simple random sampling without replacement (SRSWOR) design, say SI, the estimator in (4) reduces to

$$\hat{Y}_{C(SI)} = N\bar{y} + Nb \left(\frac{1}{N} \sum_{i=1}^N \frac{1}{x_i} - \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} \right) \tag{5}$$

where
$$b = \frac{\sum_{i=1}^n (y_i - \bar{y}) \left(\frac{1}{x_i} - \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} \right)}{\sum_{i=1}^n \left(\frac{1}{x_i} - \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} \right)^2}, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i,$$

Note that the usual regression estimator of y as per Cochran [3] under SRSWOR is given by

$$\hat{Y}_{reg(SI)} = N\bar{y} + b' N \left(\frac{1}{N} \sum_{i=1}^N x_i - \frac{1}{n} \sum_{i=1}^n x_i \right) \tag{6}$$

where
$$b' = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

Up to the first order of approximation following Sukhatme *et al.* [4], the variance of $\hat{Y}_{C(SI)}$ and $\hat{Y}_{reg(SI)}$ are given by

$$V(\hat{Y}_{C(SI)}) = N^2 \left(\frac{1}{n} - \frac{1}{N} \right) S_y^2 \left(1 - \rho_{y, \frac{1}{x}}^2 \right) \tag{7}$$

and
$$V(\hat{Y}_{reg(SI)}) = N^2 \left(\frac{1}{n} - \frac{1}{N} \right) S_y^2 (1 - \rho_{y,x}^2) \tag{8}$$

where $\rho_{y, \frac{1}{x}}$ is the correlation coefficient between y and $\frac{1}{x}$, and $\rho_{y,x}$ is the correlation coefficient between y and x , and $S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{Y})^2, \bar{Y} = \frac{Y}{N}$. Obviously, it can be remarked from the expression (7) and (8) that

if $\rho_{y, \frac{1}{x}} \geq \rho_{y,x}$.

An estimate of variance of $\hat{Y}_{C(SI)}$ up to the first order of approximation according to Sud *et al.* [11], is given by

$$\hat{V}(\hat{Y}_{C(SI)}) = \frac{N(N-n)}{n(n-1)} \frac{1}{M^2} \sum_{i=1}^n \left\{ M + \left(\frac{1}{N} \sum_{i=1}^N \frac{1}{x_i} - \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} \right) L_i \right\}^2 \{y_i - \bar{y} - bL_i\}^2 \tag{9}$$

where $L_i = \left(\frac{1}{x_i} - \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} \right)$ and $M = \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i^2} - \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} \right)^2$.

An approximate estimate of variance of $\hat{Y}_{reg(SI)}$ up to the first order of approximation is given by

$$\hat{V}(\hat{Y}_{reg(SI)}) = \frac{N(N-n)}{n} s_y^2 (1 - r_{y,x}^2) \tag{10}$$

where $s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$, $r_{y,x}$ is the estimate of $\rho_{y,x}$ obtained from the data contained in the sample S .

An attempt has been made in the present paper to first develop a model based calibration estimator under model (1) following Wu and Sitter [6] in section-2. A limited simulation study has been conducted to make the comparison of relative performance of various estimators described in preceding section and proposed model based calibration estimator in section-3.

2. Proposed Model Based Calibration Estimator

Following Wu and Sitter [6], we develop model based calibration estimator of Y under the following model

$$Y_i = \alpha + \frac{\beta}{x_i} + e_i, \quad i = 1, 2, 3, \dots, N, \tag{11}$$

with variance $V(Y_i) = \sigma^2$, $E(e_i) = 0$ and $Cov(e_i, e_j) = 0$, for $i = 1, 2, 3, \dots, N$. It is assumed that x_i 's are known for all $i = 1, 2, 3, \dots, N$. The usual Horvitz-Thompson estimator of Y is given by

$$\hat{Y}_{HT} = \sum_{i \in S} d_i y_i \tag{12}$$

We propose a model based calibration estimator of Y under the model (11) as

$$\hat{Y}_{MC} = \sum_{i \in S} w_i y_i \tag{13}$$

where w_i is calibrated weight, w_i is obtained by minimizing a distance measure function

$\sum_{i \in S} (w_i - d_i)^2 / d_i q_i$, subject to calibration constraint

$$\sum_{i \in S} w_i \hat{Y}_i = \sum_{i=1}^N \hat{Y}_i \tag{14}$$

and $\sum_{i \in S} w_i = \sum_{i=1}^N d_i$ to $\sum_{i=1}^n w_i = \sum_{i=1}^n d_i$ (15)

where \hat{Y}_i is fitted value of Y_i by least square technique.

The following function

$$\phi(w_i, \lambda_1, \lambda_2) = \sum_{i \in S} \frac{(w_i - d_i)^2}{d_i q_i} - 2\lambda_1 \left(\sum_{i \in S} w_i \hat{Y}_i - \sum_{i=1}^N \hat{Y}_i \right) - 2\lambda_2 \left(\sum_{i \in S} w_i - \sum_{i \in S} d_i \right) \tag{16}$$

is minimized with respect to w_i , where λ_1 and λ_2 are Langrangian multipliers. This yield w_i as

$$w_i = d_i + \frac{\left(\sum_{i=1}^N \hat{Y}_i - \sum_{i \in s} d_i \hat{Y}_i \right) \left(d_i q_i \hat{Y}_i \sum_{i \in s} d_i q_i - d_i q_i \sum_{i \in s} d_i q_i \hat{Y}_i \right)}{\sum_{i \in s} d_i q_i \hat{Y}_i^2 \sum_{i \in s} d_i q_i - \left(\sum_{i \in s} d_i q_i \hat{Y}_i \right)^2} \tag{17}$$

Substituting W_i from (17) into (13), the model based calibration estimator of Y is obtained as

$$\hat{Y}_{MC} = \hat{Y}_{HT} + \hat{B}^* \left(\sum_{i=1}^N \hat{Y}_i - \sum_{i \in s} d_i \hat{Y}_i \right) \tag{18}$$

where
$$\hat{B}^* = \frac{\left(\sum_{i \in s} d_i q_i \hat{Y}_i y_i \sum_{i \in s} d_i q_i - \sum_{i \in s} d_i q_i y_i \sum_{i \in s} d_i q_i \hat{Y}_i \right)}{\sum_{i \in s} d_i q_i \hat{Y}_i^2 \sum_{i \in s} d_i q_i - \left(\sum_{i \in s} d_i q_i \hat{Y}_i \right)^2}$$

An approximate variance of \hat{Y}_{MC} is obtained following Wu and Sitter [6] as follows

$$V(\hat{Y}_{MC}) = \sum_{i < j}^N (\pi_i \pi_j - \pi_{ij}) \left(\frac{E_i}{\pi_i} - \frac{E_j}{\pi_j} \right)^2 \tag{19}$$

where $E_i = y_i - B^* \hat{Y}_i$ and
$$B^* = \frac{\left[\sum_{i=1}^N d_i q_i \hat{Y}_i y_i \sum_{i=1}^N d_i q_i - \sum_{i=1}^N d_i q_i y_i \sum_{i=1}^N d_i q_i \hat{Y}_i \right]}{\left[\sum_{i=1}^N d_i q_i \hat{Y}_i^2 \sum_{i=1}^N d_i q_i - \left(\sum_{i=1}^N d_i q_i \hat{Y}_i \right)^2 \right]}$$

An approximate unbiased estimator of variance $V(\hat{Y}_{MC})$ is obtained as

$$\hat{V}(\hat{Y}_{MC}) = \sum_{i < j}^n \left(\frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \right) \left(\frac{e_i}{\pi_i} - \frac{e_j}{\pi_j} \right)^2, \text{ where } e_i = y_i - \hat{B}^* \hat{Y}_i, \tag{20}$$

For simple random sampling without replacement (SRSWOR), $\pi_i = \frac{n}{N}$ and

$$\pi_{ij} = \frac{n(n-1)}{N(N-1)},$$

For $q_i = 1$, we get the model based calibration estimator under SRSWOR, denoted as $\hat{Y}_{MC(SI)}$, as

$$\hat{Y}_{MC(SI)} = \frac{N}{n} \sum_{i \in s} y_i + \hat{B}_1^* \left(\sum_{i=1}^N \hat{Y}_i - \frac{N}{n} \sum_{i \in s} \hat{Y}_i \right), \text{ where } \hat{B}_1^* = \frac{\left(n \sum_{i \in s} \hat{Y}_i y_i - \sum_{i \in s} y_i \sum_{i \in s} \hat{Y}_i \right)}{n \sum_{i \in s} \hat{Y}_i^2 - \left(\sum_{i \in s} \hat{Y}_i \right)^2} \tag{21}$$

An approximate variance of $\hat{Y}_{MC(SI)}$ is obtained as

$$V(\hat{Y}_{MC(SI)}) = \left[\frac{(N-n)}{n(N-1)} \right] \sum_{i < j}^N (E_i - E_j)^2 \tag{22}$$

where $E_i = y_i - B_1^* \hat{Y}_i$ and $B_1^* = \frac{\left[N \sum_{i=1}^N \hat{Y}_i y_i - \sum_{i=1}^N y_i \sum_{i=1}^N \hat{Y}_i \right]}{\left[N \sum_{i=1}^N \hat{Y}_i^2 - \left(\sum_{i=1}^N \hat{Y}_i \right)^2 \right]}$,

An approximate unbiased estimator of variance $v(\hat{Y}_{MC(SI)})$ is obtained as

$$\hat{v}(\hat{Y}_{MC(SI)}) = \left[\frac{N(N-n)}{n^2(n-1)} \right] \sum_{i < j}^n (e_i - e_j)^2, \text{ where } e_i = y_i - \hat{B}_1^* \hat{Y}_i \tag{23}$$

3. Simulation Study

A limited simulation study has been conducted to examine the performance of the various estimators of population total i.e. usual regression estimator, calibration based regression type estimator due to Sud *et al.* [11], model based estimator and proposed model based calibration estimator. The performance of the estimators has been examined by their average estimates of variances obtained.

To examine the performance of the estimators through simulation, we generate hypothetical population using the following super population model

$$Y_i = \alpha + \frac{\beta}{x_i} + e_i, i = 1, 2, 3, \dots, N, v(Y_i) = \sigma^2, \tag{24}$$

We assume the value of $\alpha = 2.5$ and $\beta = 1.5$ and consider that error term e_i follows normal distribution with mean 0 and variance 2. Using the above model, populations of size $N = 500$ were generated to get y_i 's values, by considering that x_i follows chi-square distribution with 5 degree of freedom. 30000 samples of each size $n = 75$ and $n = 100$ were drawn independently from population generated of size $N = 500$ by SRSWOR design. R Software was used for selection of 30000 samples and for computing estimates of variances of the estimators under consideration.

The estimates of variances of various estimators have been computed for each sample of size $n = 75$ and $n = 100$. Therefore, 30000 estimates of the variance of each estimator have been obtained for each sample size. Average estimate of variance of each estimator has been computed as follows

$$\text{Average estimate of Variance} = \frac{1}{30000} \sum_{i=1}^{30000} V_i \tag{25}$$

where V_i is the estimate of variance of the estimator corresponding to i^{th} sample ($i = 1, 2, \dots, 30000$). The results are presented in the Tables 1.

Table 1. Average estimate of variance of the estimators.

Sample size	$\hat{v}(\hat{Y}_{reg(SI)})$	$\hat{v}(\hat{Y}_{C(SI)})$	$\hat{v}(\hat{Y}_M)$	$\hat{v}(\hat{Y}_{MC(SI)})$
n=75	120.76	98.54	2123.43	88.78
n=100	89.53	53.46	1727.84	45.67

Note that $\hat{V}(\hat{Y}_{reg(SI)})$ is the estimate of variance of usual regression estimator, $\hat{V}(\hat{Y}_{C(SI)})$ is the estimate of variance of calibration estimator, $\hat{V}(\hat{Y}_M)$ is the estimate of variance of model based estimator and $\hat{V}(\hat{Y}_{MC(SI)})$ is the estimate of variance of model based calibration estimator.

It can be observed from the results of the Table 1 that the model based calibration estimator has outperformed the other estimators. The calibration based regression type estimator has performed better than the model based estimator and usual regression estimator. It may also be noted that the performance of the calibration based regression estimator is close to model based calibration estimator. This result justifies the argument of Wu and Sitter [6] that the calibration based regression estimator and model based calibration estimator perform almost equally if model is linear. However, in view of the results in the Table 1, proposed model based calibration estimator $\hat{Y}_{MC(SI)}$ can be recommended to use in practice for estimating population total when the study variate y and the auxiliary variable x are inversely related.

4. Conclusion

The proposed model based calibration estimator has been compared with model based estimator, calibration based regression estimator and the usual regression estimator by conducting a limited simulation study. The overall results indicate that model based calibration estimator has outperformed other estimators, and it can be recommended for use in practice.

References

1. D. G. Horvitz and D. J. Thompson, *J. Am. Statist. Assoc.* **47**, 663 (1952).
<https://doi.org/10.1080/01621459.1952.10483446>
2. R. M. Royall and J. Herson, *J. Am. Statist. Assoc.* **68**, 880 (1973).
<https://www.tandfonline.com/doi/abs/10.1080/01621459.1973.10481440>
3. W. G. Cochran, *Sampling Techniques* (John Wiley & Sons, New York, 1977).
4. P. V. Sukhatme, *Sampling Theory of Surveys with Applications* (Ind. Soc. Agricult. Statist. New Delhi, 1984).
<http://krishikosh.egranth.ac.in/bitstream/1/20300/1/5632.pdf>
5. J. C. Deville and C. E. Särndal, *J. Am. Statist. Assoc.* **87**, 376 (1992).
<https://amstat.tandfonline.com/doi/abs/10.1080/01621459.1992.10475217>
6. C. Wu and R. R. Sitter, *J. Am. Statist. Assoc.* **96**, 185 (2001).
<https://doi.org/10.1198/016214501750333054>
7. P. S. Kott and C. D. Day, *J. Off. Statist.* **30**, 521 (2003). <http://doi.org/10.2478/jos-2014-0032>
8. P. J. Farrell and S. Singh, *Aust. New Zealand J. Statist.* **47**, 375 (2005).
<https://doi.org/10.1111/j.1467-842X.2005.00402.x>
9. V. M. Estevao and C. E. Särndal, *Int. Statist. Rev.* **74**, 127 (2006).
<https://doi.org/10.1111/j.1751-5823.2006.tb00165.x>
10. J.K. Kim and M. Park, *Int. Statist. Rev.* **78**, 21 (2010). <https://doi.org/10.1111/j.1751-5823.2010.00099.x>
11. U. C. Sud, H. Chandra, and V. K. Gupta, *J. Statist. Theory Pract.* **8**, 707 (2014).
<https://doi.org/10.1080/15598608.2013.832643>

12. K. K. Mourya, B. V. S. Sisodia, and H. Chandra, *J. Statist. Theory Pract.* **10**, 550 (2016).
<https://doi.org/10.1080/15598608.2016.1189861>
13. S. Kumar, B. V. S. Sisodia, D. Singh, and P. Basak, *J. Reliab. Statist. Studies* **10**, 83 (2017).
<https://www.jrss.in.net/assets/06%20JRSS%20100206.pdf>
14. S. Kumar, B. V. S. Sisodia, S. Kumar, D. Singh, and P. Basak, *Int. J. Agricult. Statist. Sci.* **13**, 677 (2017).