Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online): 2229-5666
Vol. 10, Issue 2 (2017): 83-93

CALIBRATION APPROACH BASED ESTIMATION OF FINITE POPULATION TOTAL IN SURVEY SAMPLING UNDER SUPER POPULATION MODEL WHEN STUDY VARIABLE AND AUXILIARY VARIABLE ARE INVERSELY RELATED

Sandeep Kumar[1], B.V.S. Sisodia[2]*, Dhirendra Singh[3] and Pradip Basak[4]
[1,2,3]Department of Agricultural Statistics,
Narendra Deva University of Agriculture and Technology, Faizabad, India
[4]ICAR - Indian Agricultural Statistics Research Institute, New Delhi, India
E Mail: sandippundir1982@gmail.com; bvssisodia@gmail.com*;
dhirendra.teotia@gmail.com; pradipbasak99@gmail.com
*Corresponding author

Received June 01, 2017
Modified November 07, 2017
Accepted December 05, 2017

Abstract

In the present paper we have developed calibration and model based calibration estimators of finite population total when study variable and auxiliary variable are inversely related. It has been shown that calibration, model based and model based calibration approaches provided the same estimators under certain conditions but their variances are different. A limited simulation study has been conducted to examine the relative performance of the estimators based on the aforesaid three approaches. The results of simulation study indicated that regression type model based calibration estimator is the best among all the estimators.

Key Words: Auxiliary Information, Super Population Model, Calibration Estimator, Model Based Estimator, Model Based Calibration Estimator.

1. Introduction

The information on the auxiliary variable x related to study variable y is generally used to improve the precision of the estimators of the population parameters of y such as population mean, population total, population variance etc. Various estimation approaches for estimating finite population total in survey sampling using information on auxiliary variables have been reported in the past. Most classical methods of estimation are ratio and ratio-type, product and product-type, and regression estimators. Prediction approach of Royall (1970) advocated for model based estimator (see, also Royall & Herson, 1973) which had certain advantages over classical ratio and regression estimators. Calibration approach of Deville and Sarndal (1992) also used auxiliary information related to study variable in developing calibration estimators but without assuming any underlying model. Wu and sitter (2001) developed model based calibration estimators assuming relationship between y and an auxiliary variable x, may be linear and non-linear. They argued that if the relationship is linear, then calibration estimator developed by Deville and Sarndal is the best, otherwise it performs badly if model is non-linear. However, model based calibration approach of Wu and Sitter (2001) perform precisely in both situations of linear and non-linear model.

A lot of work have been carried out on calibration estimators by various research workers. Notably among them are Kott and Day (2003), Farrel and Singh (2005), Estevao and Sarndal (2006), Kim and Park (2010), Mourya et al. (2016) etc.

Agrawal and Jain (1989) have developed model based estimator of population total $Y = \sum_{i=1}^{N} y_i$, under the model

$$Y_i = \frac{\beta}{x_i} + \frac{e_i}{x_i^{1/2}} \ , i = 1,2,3,.........N \ , \tag{1}$$

where $V(Y_i) = \frac{\sigma^2}{x_i}$ and $Y_1, Y_2, Y_3, ...Y_N$, are random variables, and y_i is realized value of Y_i.

A model based unbiased estimator of Y under the model (1) for given a sample s of size n drawn from the population of size N is given by

$$\hat{Y}_{MR} = \frac{\sum_{i\in s} y_i}{\sum_{i\in s} \frac{1}{x_i}} \sum_{i=1}^{N} \frac{1}{x_i} \tag{2}$$

with model variance

$$V(\hat{Y}_{MR}) = \sigma^2 \frac{\sum_{i\notin s} \frac{1}{x_i}}{\sum_{i\in s} \frac{1}{x_i}} \sum_{i=1}^{N} \frac{1}{x_i} \tag{3}$$

Note that the estimator in (2) is the usual ratio estimator when $\frac{1}{x_i}$ is considered as auxiliary variable instead of x_i.

Sud et al. (2014) developed calibration approach based regression type estimator for inverse relationship between study and auxiliary variable. Recently, Sandeep Kumar et al. (2017) have studied the relative performance of various approaches of estimating population total under ratio super population model.

In view of above discussion, an attempt has been made in the present paper to examine the relative performance of the estimators based on various approaches, i.e. (i) model based (ii) calibration approach based and (iii) model based calibration approach, when study and auxiliary variables are inversely related. Calibration estimator is developed in section-2. Model based calibration estimator is developed in section-3. A limited simulation study has been conducted to find out relative performance of model based estimator, calibration estimator and model based calibration estimator. A concluding remark is given in section- 5.

2. Development of Calibration Estimator

Consider that finite population $U = (U_1, U_2, U_3, ...U_N)$ consists of N identifiable units. Horvitz-Thompson estimator of population total Y is given by

$$\hat{Y}_{HT} = \sum_{i \notin s} d_i y_i \tag{4}$$

where $d_i = \dfrac{1}{\pi_i}$ is design weight, π_i being inclusion probability of i[th] unit in sample of size n drawn from the population according to sampling design $P(\cdot)$. Assuming that the auxiliary information on an auxiliary variable x_i is known for all $i = 1,2,3,.........N$, and therefore $\sum_{i=1}^{N} \dfrac{1}{x_i}$ is also known. Following Deville and Sarndal (1992), a calibration estimator of Y is defined as

$$\hat{Y}_C = \sum_{i \in s} w_i y_i \tag{5}$$

where w_i is calibrated weight obtained by minimizing a distance measure $\sum_{i \in s} (w_i - d_i)^2 \big/ d_i q_i$, where q_i is positive quantity unrelated to y, subject to calibration constraint

$$\sum_{i \in s} w_i \left(\frac{1}{x_i} \right) = \sum_{i=1}^{N} \frac{1}{x_i} \tag{6}$$

The following function

$$\phi(w_i, \lambda) = \sum_{i \in s} \frac{(w_i - d_i)^2}{d_i q_i} - 2\lambda \left(\sum_{i \in s} \frac{w_i}{x_i} - \sum_{i=1}^{N} \frac{1}{x_i} \right) \tag{7}$$

is minimized with respect to w_i, where λ is Langrangian multiplier. This yields w_i as

$$w_i = d_i + \frac{\dfrac{d_i q_i}{x_i}}{\sum_{i \in s} \dfrac{d_i q_i}{x_i^2}} \left(\sum_{i=1}^{N} \frac{1}{x_i} - \sum_{i \in s} \frac{d_i}{x_i} \right) \tag{8}$$

Substituting w_i in equation (5), the calibration estimator of Y is given by

$$\hat{Y}_C = \hat{Y}_{HT} + \hat{B}\left(\sum_{i=1}^{N}\frac{1}{x_i} - \sum_{i\in s}\frac{d_i}{x_i}\right), \text{ where } \hat{B} = \frac{\sum_{i\in s}\frac{d_i q_i y_i}{x_i}}{\sum_{i\in s}\frac{d_i q_i}{x_i^2}}, \tag{9}$$

An approximate variance of $\hat{Y}_C$ of Y for a large sample is obtained following Deville and Särndal (1992) ,i.e

$$V\left(\hat{Y}_C\right) = \frac{1}{2}\sum\sum_{i\neq j\in U} D_{ij}\pi_{ij}\left(d_i E_i - d_j E_j\right)^2 \tag{10}$$

where $D_{ij} = \frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}}$, $E_i = y_i - B\frac{1}{x_i}$ and $B = \frac{\sum_{i=1}^{N}\frac{d_i q_i y_i}{x_i}}{\sum_{i=1}^{N}\frac{d_i q_i}{x_i^2}}$, π_{ij} is joint inclusion probability of i[th] and j[th] units in the sample.

An approximate unbiased estimator of variance $V\left(\hat{Y}_C\right)$ is obtained as

$$\hat{V}\left(\hat{Y}_C\right) = \frac{1}{2}\sum\sum_{i\neq j\in s} D_{ij}\left(w_i e_i - w_j e_j\right)^2, \text{ where } e_i = y_i - \hat{B}\frac{1}{x_i}, \tag{11}$$

In case of simple random sampling without replacement (SRSWOR), we have

$\pi_i = \frac{n}{N}$, i.e. $d_i = \frac{N}{n}$, $\pi_{ij} = \frac{n(n-1)}{N(N-1)}$ and $D_{ij} = \frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} = \frac{N-n}{N(n-1)}$

For $q_i = 1$, we have regression type estimator

$$\hat{Y}_{C1} = \frac{N}{n}\sum_{i\in s} y_i + \hat{B}_1\left(\sum_{i=1}^{N}\frac{1}{x_i} - \frac{N}{n}\sum_{i\in s}\frac{1}{x_i}\right), \text{ where } \hat{B}_1 = \frac{\sum_{i\in s}\frac{y_i}{x_i}}{\sum_{i\in s}\frac{1}{x_i^2}}, \tag{12}$$

An approximate variance of $\hat{Y}_{C1}$ for a large sample under SRSWOR is obtained as

$$V\left(\hat{Y}_{C1}\right) = \frac{(N-n)}{2n(N-1)}\sum_{i\neq j=1}^{N}\sum^{N}\left[\left(y_i - y_j\right) - B_1\left(\frac{1}{x_i} - \frac{1}{x_j}\right)\right]^2 \tag{13}$$

where $B_1 = \dfrac{\sum_{i=1}^{N} \dfrac{y_i}{x_i}}{\sum_{i=1}^{N} \dfrac{1}{x_i^2}}$,

An approximate unbiased estimator of variance $V(\hat{Y}_{C1})$ is obtained as

$$\hat{V}(\hat{Y}_{C1}) = \frac{(N-n)}{2N(n-1)} \sum\sum_{i \neq j \in s} \left[(w_i y_i - w_j y_j) - \hat{B}_1 \left(\frac{w_i}{x_i} - \frac{w_j}{x_j} \right) \right]^2 \tag{14}$$

where $w_i = \dfrac{N}{n} + \dfrac{\dfrac{1}{x_i}}{\sum_{i \in s} \dfrac{1}{x_i^2}} \left(\sum_{i=1}^{N} \dfrac{1}{x_i} - \dfrac{N}{n} \sum_{i \in s} \dfrac{1}{x_i} \right)$,

For $q_i = x_i$, the calibration estimator $\hat{Y}_C$ under SRSWOR reduces to the usual ratio estimator i.e.

$$\hat{Y}_{C2} = \frac{\sum_{i \in s} y_i}{\sum_{i \in s} \frac{1}{x_i}} . \sum_{i=1}^{N} \frac{1}{x_i} \tag{15}$$

which is exactly similar to $\hat{Y}_{MR}$, when $\dfrac{1}{x_i}$ is considered as an auxiliary variable.

The approximate variance of $\hat{Y}_{C2}$ is obtained as

$$V(\hat{Y}_{C2}) = \frac{(N-n)}{2n(N-1)} \sum_{i \neq j = 1}^{N} \sum^{N} \left[(y_i - y_j) - B_2 \left(\frac{1}{x_i} - \frac{1}{x_j} \right) \right]^2 \tag{16}$$

where $B_2 = \dfrac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} \dfrac{1}{x_i}}$,

An approximate unbiased estimator of variance $V(\hat{Y}_{C2})$ is obtained as

$$\hat{V}(\hat{Y}_{C2}) = \frac{(N-n)}{2N(n-1)} w_s^{\ 2} \sum\sum_{i \neq j \in s} \left[(y_i - y_j) - \hat{B}_2 \left(\frac{1}{x_i} - \frac{1}{x_j} \right) \right]^2 \tag{17}$$

where $\hat{B}_2 = \dfrac{\sum_{i\in s} y_i}{\sum_{i\in s}\frac{1}{x_i}}$ and $w_s = \dfrac{\sum_{i\in U}\frac{1}{x_i}}{\sum_{i\in s}\frac{1}{x_i}}$, which will be constant for given sample.

3. Model Based Calibration Estimator

Following Wu and Sitter (2001), we develop model based calibration estimator of Y under the11 model given in equation (1), i.e.

$$Y_i = \frac{\beta}{x_i} + \frac{e_i}{x_i^{1/2}} \text{ , } i = 1,2,3,\ldots\ldots\ldots..N \text{ ,} \tag{18}$$

with variance $V(Y_i) = \dfrac{\sigma^2}{x_i}$, $E(e_i) = 0$ and $Cov(e_i, e_j) = 0$,

for $i = 1,2,3,\ldots\ldots\ldots..N$,

The usual Horvitz-Thompson estimator of Y is given by

$$\hat{Y}_{HT} = \sum_{i\in s} d_i y_i \tag{19}$$

We propose a model based calibration estimator of Y under the model (18) as

$$\hat{Y}_{MC} = \sum_{i\in s} w_i y_i \tag{20}$$

where w_i is calibrated weight, w_i is obtained by minimizing a distance measure $\sum_{i\in s}(w_i - d_i)^2 \Big/ d_i q_i$, subject to calibration constraint

$$\sum_{i\in s} w_i \hat{Y}_i = \sum_{i=1}^{N} \hat{Y}_i \tag{21}$$

where $\hat{Y}_i$ is estimated value of Y_i after fitting the model (18) with sample observations by least square.

The following function

$$\phi(w_i, \lambda) = \sum_{i\in s}\frac{(w_i - d_i)^2}{d_i q_i} - 2\lambda\left(\sum_{i\in s} w_i \hat{Y}_i - \sum_{i=1}^{N}\hat{Y}_i\right) \tag{22}$$

is minimized with respect to w_i, where λ is Langrangian multiplier. This yields w_i as

$$w_i = d_i + \frac{\left(\sum_{i=1}^{N}\hat{Y}_i - \sum_{i\in s} d_i \hat{Y}_i\right) d_i q_i \hat{Y}_i}{\sum_{i\in s} d_i q_i \hat{Y}_i^2} \tag{23}$$

Substituting w_i in equation (20), we get model based calibration estimator of Y as

$$\hat{Y}_{MC} = \hat{Y}_{HT} + \hat{B}^{*}\left(\sum_{i=1}^{N}\hat{Y}_i - \sum_{i\in s} d_i\hat{Y}_i\right) \tag{24}$$

where $\hat{B}^{*} = \dfrac{\sum_{i\in s} d_i q_i y_i \hat{Y}_i}{\sum_{i\in s} d_i q_i \hat{Y}_i^2}$,

Following Wu and Sitter (2001), an approximate variance of $\hat{Y}_{MC}$ is obtained as

$$V\left(\hat{Y}_{MC}\right) = \sum_{i<j}^{N}\left(\pi_i\pi_j - \pi_{ij}\right)\left(\frac{E_i}{\pi_i} - \frac{E_j}{\pi_j}\right)^2 \tag{25}$$

where $E_i = y_i - B^{*}\hat{Y}_i$ and $B^{*} = \dfrac{\sum_{i=1}^{N} d_i q_i y_i \hat{Y}_i}{\sum_{i=1}^{N} d_i q_i \hat{Y}_i^2}$,

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

$$\hat{V}\left(\hat{Y}_{MC}\right) = \sum_{i<j}^{n}\left(\frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\right)\left(\frac{e_i}{\pi_i} - \frac{e_j}{\pi_j}\right)^2 \tag{26}$$

where $e_i = y_i - \hat{B}^{*}\hat{Y}_i$,

In case of SRSWOR and for $q_i = 1$, we get model based calibration estimator, denoted by $\hat{Y}_{MC1}$, as

$$\hat{Y}_{MC1} = \frac{N}{n}\sum_{i\in s} y_i + \hat{B}_1^{*}\left(\sum_{i=1}^{N}\hat{Y}_i - \frac{N}{n}\sum_{i\in s}\hat{Y}_i\right) \tag{27}$$

where $\hat{B}_1^{*} = \dfrac{\sum_{i\in s} y_i\hat{Y}_i}{\sum_{i\in s}\hat{Y}_i^2}$

which is regression type estimator of Y .

An approximate variance of $\hat{Y}_{MC1}$ is obtained as

$$V\left(\hat{Y}_{MC1}\right)=\left[\frac{(N-n)}{n(N-1)}\right]\ \sum_{i<j}^{N}\left(E_i-E_j\right)^2 \tag{28}$$

where $E_i = y_i - B_1^{\ *}\hat{Y}_i$ and $B_1^* = \dfrac{\sum_{i=1}^{N} y_i\hat{Y}_i}{\sum_{i=1}^{N}\hat{Y}_i^{\,2}}$,

An approximate unbiased estimator of variance $V\left(\hat{Y}_{MC1}\right)$ is obtained as

$$\hat{V}\left(\hat{Y}_{MC1}\right)=\left[\frac{N(N-n)}{n^2(n-1)}\right]\ \sum_{i<j}^{n}\left(e_i-e_j\right)^2, \text{ where } e_i = y_i - \hat{B}_1^{\ *}\hat{Y}_i, \tag{29}$$

For $q_i = \dfrac{1}{\hat{Y}_i}$, we get the model based calibration estimator, denoted as $\hat{Y}_{MC2}$, as

$$\hat{Y}_{MC2} = \frac{\sum_{i\in s} y_i}{\sum_{i\in s}\frac{1}{x_i}}\sum_{i=1}^{N}\frac{1}{x_i} \tag{30}$$

which is similar to the usual ratio estimator given in equation (2) and equation (15).

An approximate variance of $\hat{Y}_{MC2}$ is obtained as

$$V\left(\hat{Y}_{MC2}\right)=\left[\frac{(N-n)}{n(N-1)}\right]\ \sum_{i<j}^{N}\left(E_i-E_j\right)^2 \tag{31}$$

where $E_i = y_i - B_2^*\hat{Y}_i$ and $B_2^* = \dfrac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N}\frac{1}{x_i}}$,

An approximate unbiased estimator of variance $V\left(\hat{Y}_{MC2}\right)$ is obtained as

$$\hat{V}\left(\hat{Y}_{MC2}\right)=\left[\frac{N(N-n)}{n^2(n-1)}\right]\ \sum_{i<j}^{n}\left(e_i-e_j\right)^2 \tag{32}$$

where $e_i = y_i - \hat{B}_2^*\hat{Y}_i$ and $\hat{B}_2^* = \dfrac{\sum_{i\in s} y_i}{\sum_{i\in s}\frac{1}{x_i}}$,

Remark

It may be noted that all the three approaches, i.e. calibration approach, model based approach and model based calibration approach provide the same usual ratio estimator when $q_i = x_i$ in calibration approach and $q_i = \frac{1}{\hat{Y}_i}$ in model based calibration approach. However, their variances are different.

4. Simulation Study

A limited simulation study has been conducted to examine the performance of the estimators of population total under various approaches, i.e. calibration approach, model based approach and model based calibration approach. The performance of the estimators has been examined on the basis of their average estimates of variances obtained.

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

$$Y_i = \frac{\beta}{x_i} + \frac{e_i}{x_i^{1/2}}, i = 1,2,3,..........N, \tag{33}$$

We assume the value of $\beta = 1.5$ 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.

5000 samples of size $n = 75$ and $n = 100$ were drawn independently from the population generated of size $N = 500$ under SRSWOR design. R Software was used for simulation study. The estimates of the above variances have been computed for each sample of size $n = 75$ and $n = 100$. This way 5000 estimates of the variances have been computed for each sample size to examine the performance of the estimators as follows

$$\text{Average Variance} = \frac{1}{5000}\sum_{i=1}^{5000} V_i \tag{34}$$

where V_i is the estimate of variance corresponding to i^{th} sample for different estimators $i = 1,2,3,..........5$. The results of simulation studies are presented in the Table 1.

	Calibration estimator		**Model based estimator**	**Model based calibration estimator**	
Sample size	$\hat{V}(\hat{Y}_{C1})$	$\hat{V}(\hat{Y}_{C2})$	$\hat{V}(\hat{Y}_{MR})$	$\hat{V}(\hat{Y}_{MC1})$	$\hat{V}(\hat{Y}_{MC2})$
n=75	99.52	101.48	1929.49	94.88	117.77
n=100	52.52	53.37	1358.28	50.43	59.75

Table 1: Average estimate of variances of the estimators

Note that $\hat{V}(\hat{Y}_{C1})$ and $\hat{V}(\hat{Y}_{C2})$ are the estimates of variances of regression type calibration estimator and ratio type calibration estimator, $\hat{V}(\hat{Y}_{MR})$ is the estimate of variance of model based estimator and $\hat{V}(\hat{Y}_{MC1})$ and $\hat{V}(\hat{Y}_{MC2})$ are the estimates of variances of model based regression type calibration estimator and model based ratio type calibration estimator.

It can be observed from the results of the Table 1 that the model based regression type calibration estimator has outperformed the other estimators. However, the regression type calibration estimator $(\hat{Y}_{C1})$ has performed better than ratio type calibration estimator $(\hat{Y}_{C2})$, model based estimator $(\hat{Y}_{MR})$ and model based ratio type calibration estimator $(\hat{Y}_{MC2})$.

5. Concluding Remarks

Calibration approach based estimator and model based calibration estimator of finite population total have been developed when study and auxiliary variables are inversely related. It has been found that these estimators are equivalent to model based estimator due to Agrawal and Jain (1989) for arbitrary constant $q_i = x_i$, and $q_i = \frac{1}{\hat{Y}_i}$, respectively. For $q_i = 1$, these calibration estimators become regression type estimator. On the basis of overall results of simulation study it can be concluded that model based regression type calibration estimator is the best performer as compared to other estimators.

Acknowledgement

Authors are very much thankful to the referees for their suggestions and comments which have substantially improved the original manuscript of the paper.

References

1. Agrawal, M.C. and Jain, Nirmal (1989). A new predictive product estimator, Biometrika, 76(4), p.822-823.
2. Deville, J.C. and Särndal, C.E. (1992). Calibration estimators in survey sampling, Journal of the American Statistical Association, 87, p. 376-382.
3. Estevao, V.M. and Sarndal, C.E. (2006). Survey estimates by calibration on complex auxiliary information, International Statistical Review, 74, p.127-147.
4. Farrell, P.J. and Singh, S. (2005). Model-assisted higher order calibration of estimators of variance, Australian and New Zealand Journal of Statistics, 47(3), p.375-383.
5. Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe, Journal of the American Statistical Association, 47, p.663-685.
6. Kim, J.K. and Park, M. (2010). Calibration estimation in survey sampling, International Statistical Review, 78, p.21-39.
7. Kott, P.S. and Day, C.D. (2003). Developing calibration weights and standard-error estimates for a survey of drug-related emergency department visits, Journal of Official Statistics, 30(3), p.521-532.
8. Mourya, K.K., Sisodia, B.V.S. and Chandra, H. (2016). Calibration approach for estimating finite population parameter in two stage sampling, Journal of Statistical Theory and Practice, 10(3), p.550-562.
9. Royall, R.M. (1970). On finite population sampling theory under certain linear regression models, Biometrika, 57(2), p.377-387.
10. Royall, R.M. and Herson, J. (1973). Robust estimation in finite populations, Journal of the American Statistical Association, 68, p.880-889.
11. Sandeep Kumar, Sisodia, B.V.S., Sunil Kumar, Singh, Dhirendra and Pradip Basak (2017). On various approaches for estimating finite population total in survey sampling under ratio super population model, International Journal of Agricultural and Statistical Sciences, 13(2), p.116-123
12. Sud, U.C., Chandra, H. and Gupta, V.K. (2014). Calibration approach based regression type estimator for inverse relationship between study and auxiliary variable, Journal of Statistical Theory and Practice, 8(4), p.707-721.
13. Wu, C. and Sitter, R.R. (2001). A model based calibration approach to using complete auxiliary information from survey data, Journal of the American Statistical Association, 96, p.185-193.