Available online at www.isas.org.in/jisas
Journal of the Indian Society of
Agricultural Statistics 73(2) 2019 99–103

Higher Order Calibration Estimator of Finite Population Total Under Two Stage Sampling Design when Population Level Auxiliary Information is Available at Unit Level

Kaustav Aditya[1], Hukum Chandra[1], Sushil Kumar[1] and Shrila Das[2]
[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi
[2]ICAR- Indian Agricultural Research Institute, New Delhi

Received 23 March 2018; Revised 19 January 2019; Accepted 11 March 2019

SUMMARY

Auxiliary information is often used to improve the precision of estimators of finite population total. Calibration approach (Deville and Sarndal, 1992) is widely used for making efficient use of auxiliary information in survey estimation. Aditya *et al.* (2016) proposed regression type estimators of the population total using the calibration approach under the assumption that the population level auxiliary information is available at secondary stage unit level under two stage sampling design. In this paper we have proposed an improved variance estimator of the regression type estimator proposed by Aditya *et al.* (2016) using higher order calibration approach (Singh *et al.*, 1998). We carried out limited simulation studies to demonstrate the empirical performance of proposed estimators. Our empirical results show that the proposed estimator performs better than the usual estimator of variances of the regression type estimator (Aditya *et al.*, 2016).

Keywords: Auxiliary information, Calibration approach, Regression type estimator, Higher order calibration, Two stage sampling.

1. INTRODUCTION

Calibration approach proposed by Deville *et al.* (1992) is widely used for making efficient use of auxiliary information in survey estimation. Following Deville *et al.* (1992) a lot of work has been carried out in the context of calibration estimation i.e. Singh *et al.* (1998, 1999), Folsom and Singh (2000), Wu and Sitter (2001), Sitter and Wu (2002), Kott (2006), Estevao and Sarndal (2002, 2006), Sud *et al.* (2014). In real life surveys two stage or multistage sampling designs are used. Aditya *et al.* (2016) first proposed a calibration estimator of the finite population total under two stage sampling design for availability of auxiliary information at the primary stage unit (psu) level. Further, Aditya *et al.* (2016) proposed a calibration estimator of the finite population total under two stage sampling design when population level complete auxiliary information is available at the secondary stage unit (ssu) level. Salinas *et al.* (2017) proposed a calibration estimator of the population mean under two stage sampling design. Further, Aditya *et al.* (2017) proposed a calibration estimator under two stage sampling design when population level auxiliary information is available at the ssu level only for the selected psus. In this study they have shown that calibration estimator under two stage sampling design works efficiently in real survey data collected through a pilot project for estimation of seed feed and wastage ratio in India. Several other works was done for development of calibration estimators and evaluating the estimators empirically through simulated data as well as real survey data where as very little work is done in the part of development of improved estimators of theoretical variance of the calibration estimators under two stage sampling design when population level auxiliary information is available at the ssu level. In this study, we have proposed a regression type estimator of variance of the calibration estimator proposed by Aditya *et al.* (2016) when auxiliary information is available at the

Corresponding author: Kaustav Aditya
E-mail address: katu4493@gmail.com

secondary stage unit (ssu) level using higher order calibration approach developed by Singh *et al.* (1998) and Aditya *et al.* (2015). Through limited simulation study we have shown that how our proposed estimator performs better than the estimator of variance of the estimator proposed by Aditya *et al.* (2016).

2. CALIBRATION BASED REGRESSION ESTIMATION UNDER TWO STAGE SAMPLING DESIGN

We consider a simple case where information on only one auxiliary variable is available. Let, the population of elements U={1,…, k,…, N_I } is partitioned into clusters, U_1, U_2,…, U_i,…, U_{N_I} . They are also called the primary stage units (psus) when there are two stages of selection. The size of U_i is denoted as N_i. We have

$$U = \bigcup_{i=1}^{N_I} U_i \text{ and } N = \sum_{i=1}^{N_I} N_i .$$

At stage one, a sample of psus, s_I, is selected from U_I according to the design $p_I(.)$ with the inclusion probabilities π_{Ii} and π_{Iij} at the psu level. The size of s_I is n_I psus. The sampling units at the second stage (ssu) are population elements, labeled $k = 1,\ldots,N$. Given that the psu U_i selected at the first stage a sample s_i of size n_i units is drawn from U_i according to some specified design $p_i(.)$ with inclusion probabilities $\pi_{k/i}$ and $\pi_{kl/i}$. For the second stage sampling we are assuming the invariance and independence property. The whole sample of elements and its size is defined as, $s = \bigcup_{i=1}^{S_I} s_i$ and $n_s = \sum_{i=1}^{n_I} n_i$.

The inclusion probabilities at the first stage is given as,

$$\pi_{Ii} = \Pr(i \in s_I)\,,$$

$$\pi_{Iij} = \begin{cases} \Pr(i \,\&\, j \in s_I), i \text{ and } j \text{ belongs to different psus} \\ \pi_{Ii}, i \text{ and } j \text{ belongs to same psus} \end{cases}$$

The inclusion probabilities for the second stage is given as,

$$\pi_{k/i} = \Pr(k \in s_i \mid i \in s_I) \text{ and}$$

$$\pi_{kl/i} = \begin{cases} \Pr(k \,\&\, l \in s_i | i \in s_I), k \text{ and } l \text{ are different} \\ \pi_{k/i}, k \text{ and } l \text{ are same} \end{cases}$$

Let the study variable be y_k which is observed for $k \in s$. The parameter to estimate is the population total $t_y = \sum_{i=1}^{N} y_k = \sum_{i=1}^{N_I} t_{yi}$ where $t_{yi} = \sum_{i=1}^{N_i} y_k$ = i-th psu total.

It was assumed that the population level complete auxiliary information (x_k) is available at the ssu level i.e. the auxiliary information x_k was known for all elements $k \in U$ and the correlation between the study variable and the auxiliary variable was positive. U was the population of size N. The simple estimator of the population total (Horvitz and Thompson, 1952) in this case will be given as,

$$\hat{t}_{HT} = \sum_{i=1}^{n_I} a_{Ii} \sum_{k=1}^{n_i} a_{k/i} y_k = \sum_{k=1}^{n_s} a_k y_k \qquad (1)$$

The calibration estimator of the population total proposed by Aditya *et al.* (2016) was given as

$$\hat{t}^*_{y\pi u} = \sum_{i=1}^{N_I}\sum_{k=1}^{N_i} \hat{y}_k + \sum_{k=1}^{n_s} w_k e_{ks}$$

$$= \sum_{i=1}^{N_I}\sum_{k=1}^{N_i} \hat{y}_k + \sum_{k=1}^{n_s} a_k e_{ks} + \frac{\sum_{k=1}^{n_s} a_k x_k e_{ks}}{\sum_{k=1}^{n_s} a_k x^2{}_k}\left(\sum_{k=1}^{N} x_k - \sum_{k=1}^{n_s} a_k x_k \right) \qquad (2)$$

where, $\hat{y}_k = \hat{\beta}^* x_k$, $e_{ks} = y_k$ - $\hat{y}_k$ and $\hat{\beta}^* = \dfrac{\sum_{k=1}^{n_s} a_k y_k x_k}{\sum_{k=1}^{n_s} a_k x^2{}_k}$.

The approximate variance of the proposed estimator under this case was obtained by first order Taylor series linearization technique and was given by

$$V(\hat{t}^*_{y\pi u}) = \sum_{i=1}^{N_I}\sum_{j=1}^{N_I} \varDelta_{Iij} \frac{t_{E_i}}{\pi_{Ii}} \frac{t_{E_j}}{\pi_{Ij}} + \sum_{i=1}^{N_I} \frac{1}{\pi_{Ii}} \sum_{k=1}^{N_i}\sum_{l=1}^{N_i} \varDelta_{kl/i} \frac{E_k}{\pi_{k/i}} \frac{E_l}{\pi_{l/i}} \qquad (3)$$

where, $E_k = y_k - \beta x_k$, $t_{E_i} = \sum_{k=1}^{N_i} E_k$,

$\varDelta_{Iij} = (\pi_{Iij} - \pi_{Ii}\pi_{Ij})$, $\varDelta_{kl/i} = \pi_{kl/i} - \pi_{k/i}\pi_{l/i}$ and

$$\beta = \frac{\sum_{k=1}^{N} y_k x_k}{\sum_{k=1}^{N} x^2{}_k}.$$

The Yates–Grundy form of estimator of variance of the proposed calibration estimator was given as,

$$\hat{V}_{YG}(\hat{t}^*_{y\pi u}) = \frac{1}{2}\sum_{i=1}^{n_I}\sum_{j=1}^{n_I} d_{Iij}\left(\frac{\hat{t}_{E_i}}{\pi_{Ii}} - \frac{\hat{t}_{E_j}}{\pi_{Ij}}\right)^2 + \frac{1}{2}\sum_{j=1}^{n_I}\frac{1}{\pi^2_{Ii}}\sum_{k=1}^{n_i}\sum_{l=1}^{n_i} d_{kl/i}(w_k e_{ks} - w_l e_{ls})^2 \tag{4}$$

where, $e_{ks} = y_k - \hat{\beta}^* x_k$, $d_{Iij} = \dfrac{(\pi_{Ii}\pi_{Ij} - \pi_{Iij})}{\pi_{Iij}}$,

$d_{kl/i} = \dfrac{(\pi_{k/i}\pi_{l/i} - \pi_{kl/i})}{\pi_{kl/i}}$ and $\hat{t}_{E_i} = \displaystyle\sum_{k=1}^{n_i}\frac{g_{ks}e_{ks}}{\pi_{k/i}}$.

It is acceptable to use the design weights in the variance estimation but Deville and Sarndal (1992) suggested that using the calibration weight (w_{Ii}) in the variance estimator makes it both design consistent and nearly model-unbiased. It is noteworthy that development reported in Section 2 under two stage sampling design does not make any assumption about the sizes of the psu/ssu.

3. PROPOSED HIGHER ORDER CALIBRATION ESTIMATORS UNDER TWO STAGE SAMPLING DESIGN

We have developed higher order calibration estimator of the variance of the Aditya *et al.* (2016) calibration estimator following the approach of Singh *et al.* (1998) for the case of availability of population level complete auxiliary information (x_k) at the unit level. The estimator of variance of the Aditya *et al.* (2016) calibration estimator of the population total was given by Eq.(4). Following Singh *et al.* (1998), here we minimize the chi-square type distance function $\dfrac{1}{2}\displaystyle\sum_{i=1}^{n_I}\sum_{j=1}^{n_I}\frac{(w_{kl} - d_{kl})^2}{d_{kl}q_{kl}}$ subject to the constraint given by

$$\frac{1}{2}\sum_{i=1}^{n_i}\sum_{j=1}^{n_i} w_{kl}\left(\frac{x_k}{\pi_{k/i}} - \frac{x_l}{\pi_{l/i}}\right)^2 = \frac{1}{2}\sum_{i=1}^{N_i}\sum_{j=1}^{N_i}(\pi_{k/i}\pi_{l/i} - \pi_{kl/i})\left(\frac{x_k}{\pi_{k/i}} - \frac{x_l}{\pi_{l/i}}\right)^2$$

and obtained the higher order calibration weight given as,

$$w_{kl} = d_{kl} + \frac{d_{kl}q_{kl}\left(\dfrac{x_k}{\pi_{k/i}} - \dfrac{x_l}{\pi_{l/i}}\right)^2}{\dfrac{1}{2}\displaystyle\sum_{i=1}^{n_i}\sum_{j=1}^{n_i} d_{kl}q_{kl}\left(\dfrac{x_k}{\pi_{k/i}} - \dfrac{x_l}{\pi_{l/i}}\right)^4}\left[\frac{1}{2}\sum_{i=1}^{N_i}\sum_{j=1}^{N_i}(\pi_{k/i}\pi_{l/i} - \pi_{kl/i})\left(\frac{x_k}{\pi_{k/i}} - \frac{x_l}{\pi_{l/i}}\right)^2 - \frac{1}{2}\sum_{i=1}^{n_i}\sum_{j=1}^{n_i} d_{kl}\left(\frac{x_k}{\pi_{k/i}} - \frac{x_l}{\pi_{l/i}}\right)^2\right]$$

The improved estimator of variance using higher order calibration approach is given as,

$$\hat{V}_{YGHO}(\hat{t}^*_{y\pi u}) = \frac{1}{2}\sum_{i=1}^{n_I}\sum_{j=1}^{n_I} d_{Iij}\left(\frac{\hat{t}_{E_i}}{\pi_{Ii}} - \frac{\hat{t}_{E_j}}{\pi_{Ij}}\right)^2 + \frac{1}{2}\sum_{j=1}^{n_I}\frac{1}{\pi^2_{Ii}}\sum_{k=1}^{n_i}\sum_{l=1}^{n_i} d_{kl}(w_k e_{ks} - w_l e_{ls})^2 + \frac{\displaystyle\sum_{j=1}^{n_I}\frac{1}{\pi^2_{Ii}}\sum_{k=1}^{n_i}\sum_{l=1}^{n_i} d_{kl}(w_k e_{ks} - w_l e_{ls})^2\left(\frac{x_k}{\pi_{k/i}} - \frac{x_l}{\pi_{l/i}}\right)^2}{\displaystyle\sum_{i=1}^{n_I}\sum_{j=1}^{n_I} d_{kl}\left(\frac{x_k}{\pi_{k/i}} - \frac{x_l}{\pi_{l/i}}\right)^4}(V_1'' - \hat{V}_1'')$$

$$= \hat{V}_{YG}(\hat{t}^*_{y\pi u}) + \hat{B}(V_1 - \hat{V}_1), \tag{5}$$

where, $d_{kl} = \dfrac{(\pi_{k/i}\pi_{l/i} - \pi_{kl/i})}{\pi_{kl/i}}$,

$$\hat{V}_1'' = \frac{1}{2}\sum_{i=1}^{n_i}\sum_{j=1}^{n_i} d_{kl}\left(\frac{x_k}{\pi_{k/i}} - \frac{x_l}{\pi_{l/i}}\right)^2,$$

$$V_1'' = \frac{1}{2}\sum_{i=1}^{N_i}\sum_{j=1}^{N_i}(\pi_{k/i}\pi_{l/i} - \pi_{kl/i})\left(\frac{x_k}{\pi_{k/i}} - \frac{x_l}{\pi_{l/i}}\right)^2.$$

4. EMPIRICAL EVALUATION

In this Section, we report the results from simulation studies that aim at assessing the performance of the developed higher order calibration estimator under two stage sampling design. These are described in Table 1.

Table 1. Definition of Various estimators considered in simulation studies

Estimators	Description	Notation
$\hat{t}^*_{y\pi u}$	Calibration estimator under two stage sampling design when population level complete auxiliary information (x_k) was available at the ssu level	$\hat{t}_{cal1\_s}$
$\hat{V}_{YGHO}(\hat{t}^*_{y\pi u})$	Higher order calibration estimator under two stage sampling design when population level complete auxiliary information (x_k) was available at the ssu level	$\hat{V}_{calHO\_s}$

In this study we have considered the case of two stage sampling where sample selection at each stage is governed by equal probability without replacement sampling design (SRSWOR). Here, we also have considered the situation that the size of the psu and the corresponding ssus were fixed. For empirical evaluation, a bi-variate normal population is generated and used for the study where BVN (22, 25, 2, 5, r). For the case of simplicity we have assumed that, $N_I = 50$ and $N_i = 100$ whereas the selected samples are of size $n_I = 15$, $n_i = 30$ and $n_I = 20$, $n_i = 40$ and there is availability of auxiliary information for both psu and ssu level. For the study we have selected a total of 1000 samples from the population using two stage equal probability without replacement sampling design and also considered different levels of correlation between the study variable and the auxiliary variable. We have considered the value of correlation coefficient as r = 0.5, 0.6, 0.7 and 0.8 for simulation study of the higher order calibration estimator. We have compared the developed higher order calibration estimator with the estimator of variance of the Aditya *et al.* (2016) lower level calibration estimator when population level complete auxiliary information (x_k) was available at the ssu level. For the empirical evaluation, SAS macro developed by Aditya *et al.* (2016) was used for selection of the samples under two Stage SRSWOR sampling design. For precision measure we have used the measure of percentage gain in efficiency (%Gain) of the higher order calibration estimator w.r.t. the original estimator of variance of the lower level calibration estimator proposed by Aditya *et al.* (2016) when population level complete auxiliary information (x_k) was available at the ssu level.

5. SUMMARY AND CONCLUSION

A close perusal of Table 2, depicts that the proposed higher order calibration estimator $\hat{V}_{YGHO}(\hat{t}_{y\pi u}^{*})$ performs better than the usual estimator of variance of the Aditya *et al.* (2016) estimator under two stage sampling design when population level complete auxiliary information (x_k) was available at the ssu level. It can be observed that the %Gain in efficiency varies between 6.22 to 7.27 for n_I = 15, n_i = 30 and between 8.2 to 9.42 for n_I = 15, n_i = 30. Further, it can be seen that the %Gain in efficiency increases with the increase of the correlation between the study variable and the auxiliary variable. Also, it was visible that with the increase of psu level sample size from 15 to 20 and ssu level sample size from 30 to 40 with a fixed level of correlation coefficient i.e. r = 0.5, the %Gain in efficiency increased by approximately 33% while for r = 0.8, %Gain in efficiency increased by approximately 29%. So from the simulation study we can conclude that our proposed higher order calibration estimator is better than the estimator of variance of the Aditya *et al.* (2016) estimator under two stage sampling design when population level complete auxiliary information (x_k) was available at the ssu level from the point of view of %Gain in efficiency.

Table 2. %Gain in efficiency of the higher order calibration estimator over the estimator of variance of the of the lower level calibration estimator proposed by Aditya *et al.* (2016) when population level complete auxiliary information (x_k) was available at the ssu level.

Sample size and Correlation	Estimate of variance $\hat{V}_{YG}(\hat{t}_{y\pi u}^{*})$	Estimate of variance using higher order calibration $\hat{V}_{YGHO}(\hat{t}_{y\pi u}^{*})$	%Gain
nI =15, ni =30, r=0.5	2359.98	2185.21	6.22
nI =15, ni =30, r=0.6	2370.33	2202.13	6.79
nI =15, ni =30, r=0.7	2392.59	2215.35	7.18
nI =15, ni =30, r=0.8	2396.03	2209.84	7.27
nI =20, ni =40, r=0.5	1436.33	1388.56	8.21
nI =20, ni =40, r=0.6	1436.20	1387.01	8.65
nI =20, ni =40, r=0.7	1436.17	1383.25	8.91
nI =20, ni =40, r=0.8	1453.81	1391.65	9.42

REFERENCES

Aditya K., Sud U.C., and Chandra H. (2016). Calibration Approach Based Estimation of Finite Population Total Under Two Stage Sampling. *J. Ind. Soc. Agril. Statist.*, **70(3)**, pp. 219-226.

Aditya K, Sud U.C., Chandra H. and Biswas A. (2016). Calibration Based Regression Type Estimator of the Population Total under Two Stage Sampling Design. *J. Ind. Soc. Agril. Statist.*, **70(1)**, pp.19-24.

Aditya K., Biswas A., Gupta A. K. and Chandra H. (2017). District Level Crop Yield Estimation Using Calibration Approach. *Current Science*, **112(9)**, pp:1927-1931.

Aditya K. and Sud, U.C. (2015). Higher order Calibration Estimators under Two Stage Sampling. Book chapter. *Statist. Informatics in Agril. Res.*, 2015, Excel India Publication, New Delhi.

Deville, J.C. and Sarndal, C.E. (1992). Calibration estimators in survey sampling. *J. Amer. Statist. Assoc.*, **87**, 376-382.

Estevao, V.M. and Särndal, C.E. (2002). The ten cases of auxiliary information for calibration in two-phase sampling. *J. Official Statist.*, **18(2)**, 233-255.

Estevao, V.M. and Sarndal, C.E. (2006). Survey Estimates by Calibration on Complex Auxiliary Information, *Int. Statist. Rev.*, **74**, 127–147.

Folsom, R.E., and Singh, A.C. (2000). The generalized exponential model for design weight calibration for extreme values, nonresponse and post stratification. Proceedings, Section on Survey Research Methods, American Statistical Association, 598-603.

Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.*, **47**, 663-685.

Kott, P.S. (2006). Using calibration weighting to adjust for nonresponse and coverage errors. *Survey Methodology*, **32**, 133142.

Salinas V., Stephan A.S. and Singh S. (2017). Calibrated estimators in two stage sampling. *Communication in statistics- Theory and Methods*, **0**, pp.1-21, Published online.

Sarndal, C.E., Swensson, B. and Wretman, J.H. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Singh, Sarjinder, Horn S. and Yu, F. (1998). Estimation of variance of general regression estimator: Higher level calibration approach. *Survey Methodology*, **24(1)**, 41-50.

Singh, S., Horn, S., Choudhury, S. and Yu, F. (1999). Calibration of the estimators of variance, *Australian and New Zealand J. Statist.,* **41(2)**, 199-212.

Sitter, R.R. and Wu, C. (2002). Efficient estimation of quadratic finite population functions. *J. Amer. Statist. Assoc.*, **97**, 535-543.

Sud, U.C., Chandra, H. and Gupta, V.K. (2014). Calibration based product estimator in single and two phase sampling. *J. Statist. Th. Prac.*, **8(1)**, 1-11.

Sud, U.C., Chandra, H. and Gupta, V. K. (2014). Calibration approach based regression type estimator for inverse relationship between study and auxiliary variable. *J. Statist. Th. Prac.*, **8(4)**, 707-721.

Wu, C., and Sitter, R.R. (2001). A model calibration approach to using complete auxiliary information from survey data. *J. Amer. Statist. Assoc.*, **96**, 185-193.