Calibration Estimator Using Different Distance Measures in Stratified Random Sampling

Dr. Nursel Koyuncu, [1] Prof. Dr. Cem Kadılar[2]
[12]*Hacettepe University, Department of Statistics, Beytepe, Ankara, Turkey*

Abstract: *Calibration estimators of finite population mean with different weights are considered. We define some new weights by using various loss functions. The estimators derived are compared by simulation.*

Keywords: *Calibration, Auxiliary information, stratified random sampling.*

I. INTRODUCTION

Calibration estimation adjust the original design weights to incorporate the known population totals of auxiliary variables. The calibration weights are chosen to minimize a given distance measure and these weights satisfy the constraints related auxiliary variable information. In survey sampling many authors, such as Deville and Sarndal (1992), Estevao and Sarndal (2000), Arnab and Singh (2005), Farrell and Singh (2005), Kim and Park (2010) etc., defined some calibration estimators using different constraints. In stratified random sampling, calibration approach is used to get optimum strata weights. Tracy et al. (2003), Kim et al. (2007) and Koyuncu (2012) define some calibration estimators in stratified random sampling. In this study, under the stratified random sampling scheme some new weights for population mean under the different distance measures are proposed.

II. NOTATIONS and CALIBRATION ESTIMATORS

Consider a finite population $U=(u_1,u_2,...,u_N)$ of size N and let Y and X, respectively, be the study and auxiliary variables associated with each unit u_j $(j=1,2,....,N)$ of the population. Let the population of size, N, is stratified into L strata with the h-th stratum containing N_h units, where $h=1,2,...,L$ such that $\sum_{h=1}^{L} N_h = N$. A simple random sample of size n_h is drawn without replacement from the h-th stratum such that $\sum_{h=1}^{L} n_h = n$. Let (y_{hi}, x_{hi}) denote observed values of y and x on the i-th unit of the h-th stratum, where $i=1,2,...,N_h$ and $h=1,2,...,L$. Under this stratified random sampling scheme, the classical unbiased estimator of the population mean is given by ,

$$\bar{y}_{st}=\sum_{h=1}^{L} W_h \bar{y}_h \tag{1.1}$$

where $W_h=\dfrac{N_h}{N}$ is the stratum weight. The calibration estimator for stratified random sampling is defined Tracy et al. (2003) given by

$$\bar{y}_{st}(Tr)=\sum_{h=1}^{L} \Omega_h \bar{y}_h \tag{1.2}$$

where Ω_h are the weights minimize the distance measure. In this study we consider following distance measures

$$L_1=\sum_{h=1}^{L}\frac{(\Omega_h-W_h)^2}{Q_h W_h} \tag{1.3}$$

$$L_2=\sum_{h=1}^{L} 2\frac{\left(\sqrt{\Omega_h}-\sqrt{W_h}\right)^2}{Q_h} \tag{1.4}$$

$$L_3=\sum_{h=1}^{L}\frac{1}{Q_h}\left(\frac{\Omega_h}{W_h}-1\right)^2 \tag{1.5}$$

$$L_4 = \sum_{h=1}^{L} \frac{1}{Q_h} \left(\frac{\sqrt{\Omega_h}}{\sqrt{W_h}} - 1 \right)^2 \tag{1.6}$$

and satisfy the calibration constraint:

$$\sum_{h=1}^{L} \Omega_h \bar{x}_h = \sum_{h=1}^{L} W_h \bar{X}_h \tag{1.7}$$

Case 1: The Lagrange function for the weights Ω_h, which satisfy the calibration equation in (1.7) and minimize the loss function given in equation (1.3) is given by

$$\Delta_1 = \sum_{h=1}^{L} \frac{(\Omega_h - W_h)^2}{Q_h W_h} - 2\lambda \left(\sum_{h=1}^{L} \Omega_h \bar{x}_h - \sum_{h=1}^{L} W_h \bar{X}_h \right) \tag{1.8}$$

Setting $\frac{\partial \Delta_1}{\partial \Omega_h} = 0$, we have

$$\Omega_h = W_h + \lambda \bar{x}_h Q_h W_h \tag{1.9}$$

On substituting weights in (1.7) and solving for lambda we have

$$\lambda = \frac{\sum_{h=1}^{L} W_h \bar{X}_h - \sum_{h=1}^{L} W_h \bar{x}_h}{\sum_{h=1}^{L} Q_h W_h \bar{x}_h^2} \tag{1.10}$$

Substituting (1.10) in (1.9) we get the weights as

$$\Omega_h = W_h + \frac{\sum_{h=1}^{L} W_h \bar{X}_h - \sum_{h=1}^{L} W_h \bar{x}_h}{\sum_{h=1}^{L} Q_h W_h \bar{x}_h^2} \bar{x}_h Q_h W_h \tag{1.11}$$

Substituting (1.11) in (1.2) we have

$$\bar{y}_{st}(Tr1) = \sum_{h=1}^{L} W_h \bar{y}_h + \frac{\sum_{h=1}^{L} W_h \bar{X}_h - \sum_{h=1}^{L} W_h \bar{x}_h}{\sum_{h=1}^{L} Q_h W_h \bar{x}_h^2} \sum_{h=1}^{L} Q_h W_h \bar{x}_h \bar{y}_h \tag{1.12}$$

Case 2: The Lagrange function for the weights Ω_h, which satisfy the calibration equation in (1.7) and minimize the loss function given in equation (1.4) is given by

$$\Delta_2 = \sum_{h=1}^{L} 2 \frac{\left(\sqrt{\Omega_h} - \sqrt{W_h}\right)^2}{Q_h} - 2\lambda \left(\sum_{h=1}^{L} \Omega_h \bar{x}_h - \sum_{h=1}^{L} W_h \bar{X}_h \right) \tag{1.13}$$

Setting $\frac{\partial \Delta_2}{\partial \Omega_h} = 0$, we have

$$\Omega_h = \frac{W_h}{(1 - \lambda \bar{x}_h Q_h)^2} \tag{1.14}$$

Solving (1.14) for lambda, we obtain

$$\lambda = \frac{\bar{x}_h \Omega_h}{\frac{\Omega_h^{3/2} \bar{x}_h^2 Q_h}{\left(\sqrt{\Omega_h} - \sqrt{W_h}\right)}} \tag{1.15}$$

Case 3: The Lagrange function for the weights Ω_h, which satisfy the calibration equation in (1.7) and minimize the loss function given in equation (1.5) is given by

$$\Delta_3 = \sum_{h=1}^{L} \frac{1}{Q_h}\left(\frac{\Omega_h}{W_h} - 1\right)^2 - 2\lambda\left(\sum_{h=1}^{L} \Omega_h \bar{x}_h - \sum_{h=1}^{L} W_h \bar{X}_h\right) \tag{1.16}$$

Setting $\frac{\partial \Delta_3}{\partial \Omega_h} = 0$, we have

$$\Omega_h = W_h + \lambda W_h^2 \bar{x}_h Q_h \tag{1.17}$$

Substituting (1.17) in (1.7) and solving for lambda we have

$$\lambda = \frac{\sum_{h=1}^{L} W_h \bar{X}_h - \sum_{h=1}^{L} W_h \bar{x}_h}{\sum_{h=1}^{L} W_h^2 \bar{x}_h^2 Q_h} \tag{1.18}$$

Substituting (1.18) in (1.17) we get the weights as

$$\Omega_h = W_h + \frac{\sum_{h=1}^{L} W_h \bar{X}_h - \sum_{h=1}^{L} W_h \bar{x}_h}{\sum_{h=1}^{L} W_h^2 \bar{x}_h^2 Q_h} W_h^2 \bar{x}_h Q_h \tag{1.19}$$

Substituting (1.19) in (1.2) we have

$$\bar{y}_{st}(Tr3) = \sum_{h=1}^{L}\left(W_h \bar{y}_h + \frac{\sum_{h=1}^{L} W_h \bar{X}_h - \sum_{h=1}^{L} W_h \bar{x}_h}{\sum_{h=1}^{L} W_h^2 \bar{x}_h^2 Q_h} W_h^2 Q_h \bar{x}_h \bar{y}_h\right) \tag{1.20}$$

Case 4: The Lagrange function for the weights Ω_h, which satisfy the calibration equation in (1.7) and minimize the loss function given in equation (1.6) is given by

$$\Delta_4 = \sum_{h=1}^{L} \frac{1}{Q_h}\left(\frac{\sqrt{\Omega_h}}{\sqrt{W_h}} - 1\right)^2 - 2\lambda\left(\sum_{h=1}^{L} \Omega_h \bar{x}_h - \sum_{h=1}^{L} W_h \bar{X}_h\right) \tag{1.21}$$

Setting $\frac{\partial \Delta_4}{\partial \Omega_h} = 0$ we have

$$\Omega_h = \frac{W_h}{\left(1 - 2\lambda \bar{x}_h Q_h W_h\right)^2} \tag{1.22}$$

Solving (1.22) for lambda, we obtain

$$\lambda = \frac{\Omega_h \bar{x}_h}{\frac{2 Q_h W_h \Omega_h^{3/2} \bar{x}_h^2}{\sqrt{\Omega_h} \pm \sqrt{W_h}}} \tag{1.23}$$

III. SIMULATION STUDY

To study the performance of the proposed estimator, we generated four different artificial populations where x_{hi}^* and y_{hi}^* values are from different distributions as given in Table 1. To get different level of correlations between study and auxiliary variables, we applied some transformations given in Table 2. Every population consists of three strata having 100 units. We selected 5000 times $n_h = 30, 20, 40$ units from each stratum, respectively. The correlation coefficients between study and auxiliary variables for each stratum are taken as $\rho_{xy1} = 0.5$, $\rho_{xy2} = 0.7$, and $\rho_{xy3} = 0.9$, respectively. The quantities $S_{1x} = 4.5$, $S_{2x} = 6.2$, $S_{3x} = 8.4$, and $S_{1y} = S_{2y} = S_{3y} = 4.8$ were fixed in each stratum (see Tracy etal. 2003, Koyuncu (2012)). We calculated empirical mean square error and relative efficiency using following formulas:

$$MSE(\bar{y}_{st}(\alpha)) = \frac{\sum_{k=1}^{\binom{N}{n}} \left[\bar{y}_{st}(\alpha) - \bar{Y}\right]^2}{\binom{N}{n}}, \qquad \alpha = Tr1, Tr3$$

It should be mentioned that in the case of distance function L_2 and L_4, the iterative procedure for finding weights from the Lagrange equations doesn't converge for all selected samples. (See Pumputis (2005)). So we didn't give simulation results for L_2 and L_4.

The simulation study shows that calibration estimator using distance measure L_1 are highly efficient than using distance measure L_3.

IV. CONCLUSION

In this study we derived some new weights using different distance measures theoretically in stratified random sampling. The performance of the weights are compared with a simulation study.

REFERENCES

[1] Arnab, R., Singh, S., 2005, A note on variance estimation for the generalized regression predictor, Australian and New Zealand Journal of Statistics, 47, 2, 231–234.
[2] Deville, J.C., Sarndal, C.E., 1992, Calibration estimators in survey sampling, Journal of the American Statistical Association, 87, 376-382.
[3] Estevao, V.M., Sarndal, C.E., 2000, A functional form approach to calibration, Journal of Official Statistics, 16, 379-399.
[4] Farrell, P.J., Singh, S., 2005, Model-assisted higher order calibration of estimators of variance, Australian and New Zealand Journal of Statistics, 47, 3, 375–383.
[5] Kim, J.M., Sungur, E.A., Heo T.Y., 2007, Calibration approach estimators in stratified sampling, Statistics and Probability Letters, 77, 1, 99-103.
[6] Kim, J.K., Park, M., 2010, Calibration estimation in survey sampling, International Statistical Review, 78, 1, 21-29.
[7] Koyuncu, N., 2012, Application of Calibration Method to Estimators in Sampling Theory, Hacettepe University Department of Statistics, PhD. Thesis.
[8] Pumputis D. Calibrated estimators under different distance measures. Proceedings of the Workshop on Survey Sampling Theory and Methodology, 2005, p. 137–141. ISBN 9955-588-87-X.
[9] Tracy, D.S., Singh, S., Arnab, R., 2003, Note on calibration in stratified and double sampling, Survey Methodology, 29, 99–104.

Table 1: Parameters and distributions of study and auxiliary variables

Parameters and distributions of the study variable	Parameters and distributions of the auxiliary variable
1. Population, $h=1,2,3$	
$f(y_{hi}^*) = \frac{1}{\Gamma(1.5)} y_{hi}^{*1.5-1} e^{-y_{hi}^*}$	$f(x_{hi}^*) = \frac{1}{\Gamma(0.3)} x_{hi}^{*0.3-1} e^{-x_{hi}^*}$
2. Population, $h=1,2,3$	
$f(y_{hi}^*) = \frac{1}{\Gamma(0.3)} y_{hi}^{*0.3-1} e^{-y_{hi}^*}$	$f(x_{hi}^*) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_{hi}^{*2}}{2}}$
3. Population, $h=1,2,3$	
$f(y_{hi}^*) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y_{hi}^{*2}}{2}}$	$f(x_{hi}^*) = \frac{1}{\Gamma(0.3)} x_{hi}^{*0.3-1} e^{-x_{hi}^*}$
4. Population, $h=1,2,3$	
$f(y_{hi}^*) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y_{hi}^{*2}}{2}}$	$f(x_{hi}^*) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_{hi}^{*2}}{2}}$

Table2: Properties of strata

Strata	**Study Variable**	**Auxiliary Variable**
1. Stratum	$y_{1i} = 50 + y_{1i}^{*}$	$x_{1i} = 15 + \sqrt{(1-\rho_{xy1}^{2})}x_{1i}^{*} + \rho_{xy1}\frac{S_{1x}}{S_{1y}}y_{1i}^{*}$
2. Stratum	$y_{2i} = 150 + y_{2i}^{*}$	$x_{2i} = 100 + \sqrt{(1-\rho_{xy2}^{2})}x_{2i}^{*} + \rho_{xy2}\frac{S_{2x}}{S_{2y}}y_{2i}^{*}$
3. Stratum	$y_{3i} = 100 + y_{3i}^{*}$	$x_{3i} = 200 + \sqrt{(1-\rho_{xy3}^{2})}x_{3i}^{*} + \rho_{xy3}\frac{S_{3x}}{S_{3y}}y_{3i}^{*}$

Table3: Mean Square Error of Estimators

	1. Population	**2. Population**	**3. Population**	**4. Population**
$\bar{y}_{st}(Tr1)$	249721461	254676687	246856254	251629447
$\bar{y}_{st}(Tr3)$	328969173	335131390	325711652	331650409