Available online at www.elixirpublishers.com (Elixir International Journal)

Statistics

Elixir Statistics 106 (2017) 46461-46465

ISSN: 2229-712X

A New Ratio Estimator of Mean in Survey Sampling by Calibration Estimation

Etebong P. Clement

Department of Mathematics and Statistics, University of Uyo, Uyo, Nigeria.

ARTICLE INFO

Article history:

Received: 2 January 2016;
Received in revised form:
25 April 2017;
Accepted: 6 May 2017;

Keywords

Calibration estimation,
Efficiency,
Ratio-type estimator,
Stratified sampling.

ABSTRACT

This paper introduces a new improved ratio estimator for population mean in stratified random sampling using calibration estimation theory. Following the results of an empirical study, it is deduced that the proposed estimator is substantially more efficient than existing estimators of its class.

1. Introduction

Calibration estimation is a method that uses auxiliary variables to adjust the original design weights to improve the precision of survey estimates of population parameters. Deville and Sarndal [1] first presented calibration estimators in survey sampling and calibration estimation has been studied by many survey statisticians. A few key references are Wu and Sitter [2], Montanari and Ranalli [3], Farrel and Singh [4], Arnab and Singh [5], Estavao and Sarndal [6], Kott [7], Singh [8, 9], Sarndal [10], Kim and Park [11], Clement *et al* [12], Clement [13, 14] and Clement and Enang [15, 16].

In stratified random sampling, calibration approach is used to obtain optimum strata weights for improving the precision of survey estimates of population parameters. Kim, Sungur and Heo [17] , Koyuncu and Kadilar [18,19] defined some calibration estimators in stratified random sampling for population characteristics and Clement *et al* [20] defined calibration estimators for domain totals in stratified random sampling. Clement and Enang [15] combined some scalars with the mean of the auxiliary variable X and proposed calibration alternative ratio estimator of mean in stratified sampling.

In this paper, based on Vishwakarma and Singh [21] separate ratio-product estimator, a new improved ratio estimator for population mean in stratified random sampling is introduced using the theory of calibration estimation.

2. Notations and review of existing estimator

Consider a finite population U of N elements $U = (U_1, U_2, \dots, U_N)$ which consists of L strata with N_h units in the hth stratum from which a simple random sample of size n_h is taken without replacement. The total population size be $N = \sum_{h=1}^{L} N_h$ and the sample size $n = \sum_{h=1}^{L} n_h$, respectively. Associated with the ith element of the hth stratum are y_{hi} and x_{hi} with $x_{hi} > 0$ being the covariate; where y_{hi} is the y value of the ith element in stratum h, and x_{hi} is the x value of the ith element in stratum h, $h = 1, 2, \dots, L$ and $i = 1, 2, \dots, N_h$ where y and x are the study variable and auxiliary variable respectively. For the hth stratum, let $W_h = N_h/N$ be the stratum weights and $f_h = n_h/N_h$, the sample fraction.

Let the hth stratum means of the study variable y and auxiliary variable x $\left(\bar{y}_h = \sum_{i=1}^{n_h} y_{hi}/n_h \, ; \, \bar{x}_h = \sum_{i=1}^{n_h} x_{hi}/n_h\right)$ be the unbiased estimator of the population mean $\left(\bar{Y}_h = \sum_{i=1}^{N_h} y_{hi}/N_h \, ; \, \bar{X}_h = \sum_{i=1}^{N_h} x_{hi}/N_h\right)$ of y and x respectively, based on n_h observations.

Vishwakarma and Singh [21] proposed a separate ratio-product estimator for population mean in stratified random sampling as:

$$\bar{y}_{st}(\alpha_h) \;= \sum_{h=1}^{L} W_h \bar{y}_h \left\{\alpha_h \frac{\bar{X}_h}{\bar{x}_h} + (1-\alpha_h)\frac{\bar{x}_h}{\bar{X}_h}\right\} \qquad (1)$$

To obtain the first degree approximation, let consider the following definitions:

$$\left.\begin{aligned} \bar{y}_h &= \bar{Y}_h\,(1 + e_{hy}) \\ \bar{x}_h &= \bar{X}_h\,(1 + e_{hx}) \end{aligned}\right\} \qquad (2)$$

Tele:
E-mail address: epclement@yahoo.com

$$\left.\begin{aligned} R_h &= \frac{\bar{x}_h}{\bar{X}_h} \\ R_h^{-1} &= \frac{\bar{X}_h}{\bar{x}_h} \end{aligned}\right\} \tag{3}$$

$$\left.\begin{aligned} E(e_{hx}^2) &= \gamma_h C_{hx}^2 \\ E(e_{hy}^2) &= \gamma_h C_{hy}^2 \\ E(e_{hx}e_{hy}) &= \gamma_h \rho_{hxy} C_{hx} C_{hy} \\ E(e_{hx}) &= E(e_{hy}) = 0 \\ E(e_{hx}e_{lx}) &= E(e_{hx}e_{ly}) = 0 \\ E(e_{hy}e_{lx}) &= E(e_{hy}e_{ly}) = 0 \end{aligned}\right\} \tag{4}$$

$$\left.\begin{aligned} K_h &= (\rho_{hxy} C_{hy}/C_{hx}) \\ C_{hy}^2 &= (S_{hy}^2/\bar{Y}_h^2) \\ C_{hx}^2 &= (S_{hx}^2/\bar{X}_h^2) \\ S_{hxy} &= \rho_{hxy} S_{hx} S_{hy} \end{aligned}\right\} \tag{5}$$

$$\left.\begin{aligned} S_{hx}^2 &= \frac{1}{N_h-1}\sum_{i=1}^{N_h}(x_{hi}-\bar{X}_h)^2 \\ S_{hy}^2 &= \frac{1}{N_h-1}\sum_{i=1}^{N_h}(y_{hi}-\bar{Y}_h)^2 \\ S_{hxy} &= \frac{1}{N_h-1}\sum_{i=1}^{N_h}(x_{hi}-\bar{X}_h)\ (y_{hi}-\bar{Y}_h) \\ \gamma_h &= \left(\frac{1-f_h}{n_h}\right) = \left(\frac{1}{n_h}-\frac{1}{N_h}\right) \end{aligned}\right\} \tag{6}$$

Substituting (2) into (3), gives

$$\left.\begin{aligned} R_h &= (1+e_{hx}) \\ R_h^{-1} &= (1+e_{hx})^{-1} \end{aligned}\right\} \tag{7}$$

Expressing (1) in terms of the e's in (2) with respect to (7), gives

$$\bar{y}_{st}(\alpha_h) = \sum_{h=1}^{L} W_h \bar{Y}_h\,(1+e_{hy})\{\alpha_h(1+e_{hx})^{-1} + (1-\alpha_h)(1+e_{hx})\} \tag{8}$$

It is assumed that $|e_{hx}| < 1$ so that expanding $(1+e_{hx})^{-1}$ as a series in power of e_{hx}, multiplying out and retaining terms of the e 's to the second degree, gives

$$\bar{y}_{st}(\alpha_h) = \sum_{h=1}^{L} W_h \bar{Y}_h\left[1 + e_{hy} + e_{hx} + e_{hx}e_{hy} + \alpha_h(e_{hx}^2 - 2e_{hx} - 2e_{hx}e_{hy})\right]$$

$$\bar{y}_{st}(\alpha_h) - \bar{Y} = \sum_{h=1}^{L} W_h \bar{Y}_h\left[(e_{hy} + e_{hx} + e_{hx}e_{hy}) + \alpha_h(e_{hx}^2 - 2e_{hx} - 2e_{hx}e_{hy})\right] \tag{9}$$

Squaring both sides of (9) and retaining terms to the second degree, gives

$$(\bar{y}_{st}(\alpha_h) - \bar{Y})^2 = \left[\sum_{h=1}^{L} W_h \bar{Y}_h\left[(e_{hy} + e_{hx} + e_{hx}e_{hy}) + \alpha_h(e_{hx}^2 - 2e_{hx} - 2e_{hx}e_{hy})\right]\right]^2$$

$$\begin{aligned}(\bar{y}_{st}(\alpha_h) - \bar{Y})^2 &= \sum_{h=1}^{L} W_h^2 \bar{Y}_h^2\left[e_{hy}^2 + e_{hx}^2 + 4\alpha_h^2 e_{hx}^2 - 4\alpha e_{hx}^2 + 2e_{hx}e_{hy} - 4\alpha e_{hx}e_{hy} + \cdots\right] \\ &+ \sum_{h\neq l=1}^{L} W_h W_l \bar{Y}_h \bar{Y}_l\left[e_{hy}e_{ly} + e_{hx}e_{lx} + 4\alpha_h^2 e_{hx}e_{lx} - 4\alpha e_{hx}e_{lx} + \cdots\right]\end{aligned} \tag{10}$$

Taking expectation of both sides of (10) and using the results in (4), gives the MSE of $\bar{y}_{st}(\alpha_h)$ to the first order of approximation as:

$$MSE(\bar{y}_{st}(\alpha_h)) = \sum_{h=1}^{L} W_h^2 \bar{Y}_h^2 \gamma_h\{C_{hy}^2 + C_{hx}^2[(1-2\alpha_h)^2 + 2k_h(1-2\alpha_h)]\} \tag{11}$$

The MSE of $\bar{y}_{st}(\alpha_h)$ in (11) is minimized when

$$\alpha_h = \left(\frac{1+K_h}{2}\right) \tag{12}$$

Substituting (12) in (11) gives

$$MSE(\bar{y}_{st}(\alpha_h)) = \sum_{h=1}^{L} W_h^{*2} \bar{Y}_h^2 \gamma_h\left[C_{hy}^2 - K_h^2 C_{hx}^2\right]$$

Applying (5) gives

$$\mathbf{MSE}(\bar{y}_{st}(\alpha_h)) = \sum_{h=1}^{L} W_h^2 \gamma_h S_{hy}^2 (1-\rho_{hxy}^2) \quad (13)$$

3. Suggested estimator

Let rewrite (1) as

$$\bar{y}_{st}^*(\alpha_h) = \sum_{h=1}^{L} W_h^* \bar{y}_h \lambda \quad (14)$$

where the coefficient

$\lambda = \left\{\alpha_h \frac{\bar{X}_h}{\bar{x}_h} + (1-\alpha_h)\frac{\bar{x}_h}{\bar{X}_h}\right\}$ and W_h^* is the new weights called the calibration weights and are chosen such that a chi-square-type loss functions of the form:

$$\sum_{h=1}^{L} \left(\frac{W_h^* - W_h}{Q_h W_h}\right)^2 \quad (15)$$

is minimized subject to a calibration constraints of the form:

$$\sum_{h=1}^{L} W_h^* S_{hx}^2 = V(\bar{x}_{st}) \quad (16)$$

Minimizing the loss function (15) subject to the calibration constraints (16) leads to the calibration weights for stratified sampling given by

$$W_h^* = W_h + \left(V(\bar{x}_{st}) - \sum_{h=1}^{L} W_h S_{hx}^2\right) \frac{Q_h W_h S_{hx}^2}{\sum_{h=1}^{H} Q_h W_h (S_{hx}^2)^2} \quad (17)$$

Let

$$W_h^{*2} = \left[W_h + \left(V(\bar{x}_{st}) - \sum_{h=1}^{L} W_h S_{hx}^2\right) \frac{Q_h W_h S_{hx}^2}{\sum_{h=1}^{H} Q_h W_h (S_{hx}^2)^2}\right]^2$$

and setting the tuning parameter $Q_h = S_{hx}^{-2}$, then

$$W_h^{*2} = W_h^2 \left[\frac{V(\bar{x}_{st})}{\hat{V}(\bar{x}_{st})}\right]^2 \quad (18)$$

Where $V(\bar{x}_{st}) = \sum_{h=1}^{L} W_h^2 \gamma_h S_{hx}^2$ and $\hat{V}(\bar{x}_{st}) = \sum_{h=1}^{L} W_h S_{hx}^2$

Substituting (17) into (14) gives the proposed estimator as:

$$\bar{y}_{st}^*(\alpha_h) = \sum_{h=1}^{L} \left[W_h + \left(V(\bar{x}_{st}) - \sum_{h=1}^{L} W_h S_{hx}^2\right) \frac{Q_h W_h S_{hx}^2}{\sum_{h=1}^{H} Q_h W_h (S_{hx}^2)^2}\right] \bar{y}_h \lambda \quad (19)$$

4. Variance estimator

Let

$$\bar{y}_{st}^*(\alpha_h) - \bar{Y} = \sum_{h=1}^{L} W_h^* \bar{y}_h \lambda - \bar{Y} \quad (20)$$

Squaring both sides of (20) and taking expectation gives

$$E(\bar{y}_{st}^*(\alpha_h) - \bar{Y})^2 = E\left(\sum_{h=1}^{L} W_h^* \bar{y}_h \lambda - \bar{Y}\right)^2 \quad (21)$$

$$E(\bar{y}_{st}^*(\alpha_h) - \bar{Y})^2 = E\left(\sum_{h=1}^{L} W_h^* \bar{y}_h \lambda\right)^2 - 2\lambda\bar{Y}E\left(\sum_{h=1}^{L} W_h^* \bar{y}_h\right) + \bar{Y}^2$$

$$E(\bar{y}_{st}^*(\alpha_h) - \bar{Y})^2 = Var\left(\sum_{h=1}^{L} W_h^* \bar{y}_h \lambda\right) + \left(E\left(\sum_{h=1}^{L} W_h^* \bar{y}_h \lambda\right)\right)^2 - 2\lambda\bar{Y}E\left(\sum_{h=1}^{L} W_h^* \bar{y}_h\right) + \bar{Y}^2$$

$$E(\bar{y}_{st}^*(\alpha_h) - \bar{Y})^2 = \lambda^2 Var\left(\sum_{h=1}^{L} W_h^* \bar{y}_h\right) + \lambda^2\left(\sum_{h=1}^{L} W_h^* \bar{Y}_h\right)^2 - 2\lambda\bar{Y}\left(\sum_{h=1}^{L} W_h^* \bar{Y}_h\right) + \bar{Y}^2$$

$$E(\bar{y}_{st}^*(\alpha_h) - \bar{Y})^2 = \lambda^2 \sum_{h=1}^{L} W_h^{*2}\, Var(\bar{y}_h) + \bar{Y}^2(\lambda - 1)^2$$

$$\mathbf{MSE}(\bar{y}_{st}^*(\alpha_h)) = \lambda^2 \sum_{h=1}^{L} W_h^{*2} \gamma_h S_{hy}^2 + \bar{Y}^2(\lambda-1)^2 \quad (22)$$

Substituting (18) into (22)

$$\mathbf{MSE}(\bar{y}_{st}^*(\alpha_h)) = \lambda^2 \left[\frac{V(\bar{x}_{st})}{\hat{V}(\bar{x}_{st})}\right]^2 \sum_{h=1}^{L} W_h^2 \gamma_h S_{hy}^2 + \bar{Y}^2(\lambda-1)^2 \quad (23)$$

Minimizing (23) with respect to λ gives

$$\lambda\left(\left[\frac{V(\bar{x}_{st})}{\hat{V}(\bar{x}_{st})}\right]^2 \sum_{h=1}^{L} W_h^2 \gamma_h S_{hy}^2 + \bar{Y}^2\right) = \bar{Y}^2$$

So that

$$\lambda = \frac{\bar{Y}^2}{\left(\left[\frac{V(\bar{x}_{st})}{\hat{V}(\bar{x}_{st})}\right]^2 \sum_{h=1}^{L} W_h^2 \gamma_h S_{hy}^2 + \bar{Y}^2\right)} \tag{24}$$

where $0 < \lambda < 1$

Substituting (24) into (23) gives

$$MSE(\bar{y}_{st}^*(\alpha_h)) = \frac{\bar{Y}^2\left[\frac{V(\bar{x}_{st})}{\hat{V}(\bar{x}_{st})}\right]^2 \sum_{h=1}^{L} W_h^2 \gamma_h S_{hy}^2}{\left(\left[\frac{V(\bar{x}_{st})}{\hat{V}(\bar{x}_{st})}\right]^2 \sum_{h=1}^{L} W_h^2 \gamma_h S_{hy}^2 + \bar{Y}^2\right)} \tag{25}$$

5. Application

For numerical illustration, the same data set of Vishwakarma and Singh [21] is considered. The values of the parameters related to the study variable (Y) and the auxiliary variable (X) are shown in table 1.

Table 1: Data statistics

Parameter	Stratum 1	Stratum 2	Stratum 3	Total
N_h	6	8	11	$N = 25$
n_h	3	3	4	$n = 10$
$\bar{X}_h$	6.813	10.12	7.967	$\bar{X} = 8.3792$
$\bar{Y}_h$	417.33	503.375	340.00	$\bar{Y} = 410.84$
S_{hx}^2	15.9712	132.66	38.438	$S_x^2 = 59.7368$
S_{hy}^2	74775.467	259113.70	65885.60	$S_y^2 = 1237702$
S_{hxy}^2	1007.0547	5709.1629	1404.71	$S_{xy}^2 = 2524.79$
ρ_{hxy}	0.9215	0.9738	0.8827	$\rho = 0.9285$
γ_h	0.1667	0.2083	0.1591	$R = 49.8610$
w_h^2	0.0576	0.1024	0.1936	$\rho^* = 0.9409$

The percent relative efficiency (PRE) of an estimator θ with respect to the stratified random sampling estimator $(\bar{y}_{st})$ is defined by

$$PRE(\theta, \bar{y}_{st}) = \frac{Var(\bar{y}_{st})}{Var(\theta)} \times 100$$

$$V(\bar{x}_{st}) = \sum_{h=1}^{H} W_h^2 \gamma_h S_{hx}^2 = 4.1669$$

$$\hat{V}(\bar{x}_{st}) = \sum_{h=1}^{H} W_h S_{hx}^2 = 63.1970$$

$$\sum_{h=1}^{H} W_h^2 \gamma_h S_{hy}^2 = 8274.8790$$

$$\left[\frac{V(\bar{x}_{st})}{\hat{V}(\bar{x}_{st})}\right]^2 \sum_{h=1}^{H} W_h^2 \gamma_h S_{hy}^2 = 35.9745$$

Table 2. MSE Values and PREs of Estimators

S/No.	Estimator	MSE	PREs
1	Stratified sampling	8274.8790	100.00
2	Hansen-Hurwitz-Gurney	1159.0469	713.94
3	Singh-Vishwakarma	842.2866	872.13
4	Vishwakarma- Singh	842.2866	982.07
5	Proposed	35.9668	23006.99

In Table 2, the values of MSE and PRE are given. It is observed that the proposed estimator has the minimum MSE and highest gains in efficiency and therefore is the best estimator for the data.

In the same way, when comparing the MSE of proposed estimator with MSE of existing estimators, it is observed that proposed estimator is more efficient than Hansen-Hurwitz-Gurney [22] combined ratio estimator, Singh-Vishwakarma [23] combined ratio-product estimator and Vishwakarma- Singh [21] separate ratio-product estimator respectively.

6. Conclusion

In this study, a new improved ratio estimator is introduced following Vishwakarma-Singh [21] estimator using calibration estimation theory. The relative performances of new estimator are compared with an empirical study. It is found that suggested estimator perform better than existing estimators.

References

[1] Deville, J.C. and Sarndal, C. E. (1992): Calibration estimators in survey sampling. Journal of the American Statistical Association, 87, 376-382.
[2] Wu, C. and Sitter, R.R. (2001): A Model-calibration approach to using complete auxiliary information from survey data, Journal of the American Statistical Association, 96,185-193.
[3] Montanari, G. E. and Ranalli, M.G. (2005): Nonparametric model calibration estimation in survey sampling. Journal of the American Statistical Association, 100, 1429-1442.
[4] Farrell, P.J. and Singh, S (2005). Model-assisted higher order calibration of estimators of variance. Australia and New Zealand Journal of Statistics, 47 (3), 375-383.
[5] Arnab, R. and Singh, S. (2005): A note on variance estimation for the generalized regression predictor. Australia and New Zealand Journal of Statistics.47 (2), 231-234.
[6] Estavao, V. M. and Sarndal, C.E. (2006). Survey estimates by calibration on complex auxiliary information. International Statistical Review, 74, 127-147.
[7] Kott, P.S. (2006): Using calibration weighting to adjust for non-response and coverage errors Survey Methodology, 32, 133-142.
[8] Singh, S. (2006). Survey statistician celebrate golden jubilee year- 2003 of the linear regression estimator. Metrika. 1-18.
[9] Singh, S. (2006). Calibrated empirical likelihood estimation using a displacement function: Sir R. A. Fisher's Honest Balance, Presented at INTERFACE 2006.Pasadena, CA,USA.
[10] Sarndal, C-E. (2007):The calibration approach in survey theory and practice. Survey Methodology, 33, 99-119.
[11] Kim, J. K. and Park, M. (2010): Calibration estimation in survey sampling. International Statistical Review, 78 (1), 21-29.
[12] Clement, E. P., Udofia, G.A. and Enang, E. I. (2014). Estimation for domains in stratified random sampling in the presence of non-response. American Journal of Mathematics and Statistics, 4 (2), 65-71.
[13] Clement, E. P. (2015). Calibration Approach Separate Ratio Estimator for Population Mean in Stratified Sampling. International Journal of Modern Mathematical Sciences, 13(4):377-384
[14] Clement, E. P. (2017). Efficient Exponential Estimators of Population Mean in Survey Sampling. International Journal of Mathematics and Computation 28 (3):94-106
[15] Clement, E. P. and Enang, E. I. (2015): Calibration approach alternative ratio estimator for population mean in stratified sampling. International Journal of Statistics and Economics, 16(1), 83-93.
[16] Clement, E. P. and Enang, E. I. (2017). On the Efficiency of Ratio Estimator over the Regression Estimator, Communication in Statistics: Theory and Methods, 46(11):5357-5367
[17] Kim, J.M. Sungur, E.A. and Heo T.Y. (2007): Calibration approach estimators in stratified sampling. Statistics and Probability Letters, 77(1), 99-103.
[18] Koyuncu, N. and Kadilar, C. (2013). Calibration estimators using different measures in stratified random sampling. International Journal of Modern Engineering Research, 3 (1), 415-419.
[19] Koyuncu, N. and Kadilar, C. (2014): A new calibration estimators in stratified double sampling. Hacettepe Journal of Mathematics and Statistics 43(2), 337-346.
[20] Clement, E. P., Udofia, G.A. and Enang, E. I. (2014). Sample design for domain calibration estimators. International Journal of Probability and Statistics, 3 (1), 8-14.
[21] Vishwakarma, G. K. and Singh, H. P. (2011): Separate ratio-product estimator for estimating population mean using auxiliary information. Journal of Statistical Theory and Applications, 10 (4), 653-664.
[22] Hansen, M.H., Hurwitz, W.N. and Gurney, M. (1946): The problems and methods of the Sample survey of business. Journal of the American Statistical Association, 41,173-189.
[23] Singh H.P. and Vishwakarma, G. K. (2006): Combined ratio-product estimator of finite population mean in stratified sampling. Metodologia de Encuestas. 7 (1), 35-44