Asian Journal of Probability and Statistics

10(1): 25-35, 2020; Article no.AJPAS.61000

ISSN: 2582-0230

On the Estimation of Variance of Calibration Regression Estimators with Multiple Auxiliary Information

Etebong P. Clement[1*]

[1]Department of Statistics, University of Uyo, Uyo, Nigeria.

Author's contribution

The sole author designed, analysed, interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/AJPAS/2020/v10i130238
Editor(s):
(1) Dr. Jiteng Jia, Xidian University, China.
Reviewers:
(1) Andri Donal, University of Pasir Pengaraian, Indonesia.
(2) Faizan Danish, New York University, United States.
Complete Peer review History: http://www.sdiarticle4.com/review-history/61000

Original Research Article

Received: 15 July 2020
Accepted: 20 September 2020
Published: 12 December 2020

Abstract

This paper introduces the concept of calibration estimators to Statistical Regression Estimation and proposes a multivariate calibration regression *(M-REG)* estimator of population mean in stratified random sampling. It develops a new approach to variance estimation that is more efficient in estimating populations with multiple auxiliary variables using the principle of analysis of variance (*ANOVA*). The relative performance of the new variance estimation method with respect to the estimation of variance of the proposed *M-REG* estimator is compared empirically with a corresponding global variance estimation method. Analysis and evaluation presented, proved the dominance of the suggested new approach to variance estimation.

Keywords: Analysis of variance; calibration estimation; efficiency; optimality conditions; stratified random sampling; variance estimation.

Mathematics Subject Classification: 62D05; 62G05; 62H12

1 Introduction

The concept of calibration estimator was introduced by [1] in survey sampling. Calibration estimation is a method that uses auxiliary variable(s) to adjust the original design weights to improve the precision of

**Corresponding author: E-mail: epclement@yahoo.com, etebongclemnt@uniuyo.edu.org;*

survey estimates of population or subpopulation parameters. The calibration weights are chosen to minimize a given distance measure (or loss function) and these weights satisfy the constraints related auxiliary variable information. Calibration estimation has been studied by many survey Statisticians. A few key references are [2-21].

The large sample approximation (*LASAP*) method has been the most dominant approach to variance estimation in survey sampling. However, it has been observed that this approach always depends on certain optimality conditions that need to be satisfied to guarantee a better and efficient estimator. Again, it is well established in sample surveys that incorporate auxiliary information, that the precision of survey estimates is always improved when multiple auxiliary information are available. Keeping this in view, this paper introduces calibration weightings to statistical regression estimation, develops a new variance estimation method that is more efficient for estimating populations with multiple auxiliary variables using the principle of analysis of variance (*ANOVA)* and proposes a multivariate calibration regression estimator of population mean in stratified random sampling. The efficiency of the proposed variance estimation method is compared with the conventional large sample approximation (*LASAP*) method.

2 The Suggested Multivariate Calibration Regression (*M-REG*) Estimator

The calibration estimator for the stratified random sampling is defined by [22] as given by:

$$\bar{y}_{st}(Tr) = \sum_{h=1}^{H} W_h^* \bar{y}_h \tag{1}$$

where W_h^* is the calibration weights which minimizes given calibration constraints.

Motivated by [22], this paper introduces a multivariate calibration regression (*M-REG*) estimator in stratified random sampling as given by

$$\bar{y}_{st}^*(MREG) = \sum_{h=1}^{H} \varphi_h^* \bar{y}_h \tag{2}$$

with the new weights φ_h^* called the *multivariate calibration weights*. The multivariate calibration weights φ_h^* are chosen such that a chi-square-type loss functions of the form:

$$L(\varphi_h^*, W_h) = \sum_{h=1}^{H} \frac{(\varphi_h^* - W_h)^2}{W_h Q_h} \tag{3}$$

is minimized while satisfying the calibration constraints

$$\sum_{h=1}^{H} \varphi_h^* \bar{X}_{1h} = \sum_{h=1}^{H} W_h \bar{X}_1 \tag{4}$$

$$\sum_{h=1}^{H} \varphi_h^* \bar{X}_{2h} = \sum_{h=1}^{H} W_h \bar{X}_2 \tag{5}$$

$$\sum_{h=1}^{H} \varphi_h^* \bar{X}_{3h} = \sum_{h=1}^{H} W_h \bar{X}_3 \tag{6}$$

So that

$$\Delta = \sum_{h=1}^{H} \frac{(\varphi_h^* - W_h)^2}{W_h Q_h} - 2\lambda_1 \left(\sum_{h=1}^{H} \varphi_h^* \bar{X}_{1h} - \sum_{h=1}^{H} W_h \bar{X}_1 \right) - 2\lambda_2 \left(\sum_{h=1}^{H} \varphi_h^* \bar{X}_{2h} - \sum_{h=1}^{H} W_h \bar{X}_2 \right)$$
$$-2\lambda_3 \left(\sum_{h=1}^{H} \varphi_h^* \bar{X}_{3h} - \sum_{h=1}^{H} W_h \bar{X}_3 \right) \tag{7}$$

Minimizing the chi-square-type loss functions (3) subject to the calibration constraints [(4),(5), (6)] gives the multivariate calibration weights for stratified random sampling as:

$$\varphi_h^* = W_h + W_h Q_h (\lambda_1 \bar{X}_{1h} + \lambda_2 \bar{X}_{2h} + \lambda_3 \bar{X}_{3h}) \tag{8}$$

Substituting (8) into [(4),(5), (6)] respectively gives the following system of equations as:

$$\begin{bmatrix} \mu_{11} & \mu_{12} & \mu_{13} \\ \mu_{12} & \mu_{22} & \mu_{23} \\ \mu_{13} & \mu_{23} & \mu_{33} \end{bmatrix} \begin{bmatrix} \lambda \\ \lambda \\ \lambda \end{bmatrix} = \begin{bmatrix} \mu_{10} \\ \mu_{20} \\ \mu_{30} \end{bmatrix} \tag{9}$$

Where

$$\mu_{11} = \textstyle\sum_{h=1}^{H} W_h Q_h \bar{X}_{1h}^2 \qquad \mu_{22} = \sum_{h=1}^{H} W_h Q_h \bar{X}_{2h}^2 \qquad \mu_{33} = \sum_{h=1}^{H} W_h Q_h \bar{X}_{3h}^2$$

$$\mu_{12} = \textstyle\sum_{h=1}^{H} W_h Q_h \bar{X}_{1h} \bar{X}_{2h} \qquad \mu_{13} = \sum_{h=1}^{H} W_h Q_h \bar{X}_{1h} \bar{X}_{3h} \qquad \mu_{23} = \sum_{h=1}^{H} W_h Q_h \bar{X}_{2h} \bar{X}_{3h}$$

$$\mu_{10} = \textstyle\sum_{h=1}^{H} W_h (\bar{X}_1 - \bar{X}_{1h}) \quad \mu_{20} = \sum_{h=1}^{H} W_h (\bar{X}_2 - \bar{X}_{2h}) \quad \mu_{30} = \sum_{h=1}^{H} W_h (\bar{X}_3 - \bar{X}_{3h})$$

Solving the system of equations in (9) for λs gives

$$\lambda_1 = \frac{(\mu_{13}\mu_{23} - \mu_{12}\mu_{33})(\mu_{12}\mu_{20} - \mu_{22}\mu_{10}) - (\mu_{13}\mu_{22} - \mu_{12}\mu_{23})(\mu_{12}\mu_{30} - \mu_{23}\mu_{10})}{(\mu_{12}^2 - \mu_{11}\mu_{22})(\mu_{13}\mu_{23} - \mu_{12}\mu_{33}) - (\mu_{13}\mu_{22} - \mu_{12}\mu_{23})(\mu_{12}\mu_{13} - \mu_{11}\mu_{23})}$$

$$\lambda_2 = \frac{(\mu_{13}\mu_{23} - \mu_{12}\mu_{33})(\mu_{12}\mu_{10} - \mu_{11}\mu_{20}) - (\mu_{12}\mu_{13} - \mu_{11}\mu_{23})(\mu_{13}\mu_{20} - \mu_{12}\mu_{30})}{(\mu_{12}^2 - \mu_{11}\mu_{22})(\mu_{13}\mu_{23} - \mu_{12}\mu_{33}) - (\mu_{13}\mu_{22} - \mu_{12}\mu_{23})(\mu_{12}\mu_{13} - \mu_{11}\mu_{23})}$$

$$\lambda_3 = \frac{(\mu_{12}^2 - \mu_{11}\mu_{22})(\mu_{13}\mu_{20} - \mu_{12}\mu_{30}) - (\mu_{13}\mu_{22} - \mu_{12}\mu_{23})(\mu_{12}\mu_{10} - \mu_{11}\mu_{20})}{(\mu_{12}^2 - \mu_{11}\mu_{22})(\mu_{13}\mu_{23} - \mu_{12}\mu_{33}) - (\mu_{13}\mu_{22} - \mu_{12}\mu_{23})(\mu_{12}\mu_{13} - \mu_{11}\mu_{23})}$$

Substituting (8) in (2) gives

$$\bar{y}_{st}^*(MREG) = \sum_{h=1}^{H} W_h \bar{y}_h + \sum_{h=1}^{H} W_h Q_h \left[\lambda_1 \bar{X}_{1h} + \lambda_2 \bar{X}_{2h} + \lambda_3 \bar{X}_{3h}\right] \bar{y}_h \tag{10}$$

By substituting the λs in (10) and setting $Q_h = 1$ gives the proposed multivariate calibration regression (*M-REG*) estimator of population mean in stratified random sampling as given by:

$$\bar{y}_{st}^*(MREG) = \bar{y}_{st} + \beta_1 \mu_{10} + \beta_2 \mu_{20} + \beta_3 \mu_{30} \tag{11}$$

Where

$$\beta_1 = \frac{b_{12}[b_{14}(b_{22}b_{33} - b_{23}^2) + b_{24}(b_{13}b_{23} - b_{12}b_{23}) + b_{34}(b_{12}b_{23} - b_{13}b_{22})]}{(b_{12}^2 - b_{11}b_{22})(b_{13}b_{23} - b_{12}^2) - (b_{13}b_{22} - b_{12}b_{23})(b_{12}b_{13} - b_{11}b_{23})}$$

$$\beta_2 = \frac{b_{12}[b_{14}(b_{13}b_{23} - b_{12}b_{33}) + b_{24}(b_{11}b_{33} - b_{13}^2) + b_{34}(b_{12}b_{13} - b_{11}b_{23})]}{(b_{12}^2 - b_{11}b_{22})(b_{13}b_{23} - b_{12}^2) - (b_{13}b_{22} - b_{12}b_{23})(b_{12}b_{13} - b_{11}b_{23})}$$

$$\beta_3 = \frac{b_{12}[b_{14}(b_{13}b_{22} - b_{12}b_{23}) + b_{24}(b_{12}b_{13} - b_{11}b_{23}) + b_{34}(b_{11}b_{22} - b_{12}^2)]}{(b_{12}^2 - b_{11}b_{22})(b_{13}b_{23} - b_{12}^2) - (b_{13}b_{22} - b_{12}b_{23})(b_{12}b_{13} - b_{11}b_{23})}$$

Where

$$b_{11} = \sum_{h=1}^{H} W_h \bar{X}_{1h}^2 \qquad b_{22} = \sum_{h=1}^{H} W_h \bar{X}_{2h}^2 \qquad b_{33} = \sum_{h=1}^{H} W_h \bar{X}_{3h}^2$$

$$b_{12} = \sum_{h=1}^{H} W_h \bar{X}_{1h}\bar{X}_{2h} \qquad b_{13} = \sum_{h=1}^{H} W_h \bar{X}_{1h}\bar{X}_{3h} \qquad b_{23} = \sum_{h=1}^{H} W_h \bar{X}_{2h}\bar{X}_{3h}$$

$$b_{14} = \sum_{h=1}^{H} W_h \bar{X}_{1h}\bar{y}_h \qquad b_{24} = \sum_{h=1}^{H} W_h \bar{X}_{2h}\bar{y}_h \qquad b_{34} = \sum_{h=1}^{H} W_h \bar{X}_{3h}\bar{y}_h$$

3 Estimation of Variance for the Proposed Estimator

This section attempts to derive the estimator of variance of the proposed multivariate calibration regression (*M-REG*) estimator under the large sample approximation (*LASAP*) method and the suggested analysis of variance (*ANOVA*) method as discuss in sections 4.1 and 4.2 respectively.

3.1 Large sample approximation (LASAP) method

Let define the following equations

$$e_{hy} = \left(\frac{\bar{y}_h - \bar{Y}_h}{\bar{Y}_h}\right), e_{hx1} = \left(\frac{\bar{X}_1 - \bar{X}_{1h}}{\bar{X}_1}\right), e_{hx2} = \left(\frac{\bar{X}_2 - \bar{X}_{2h}}{\bar{X}_2}\right), e_{hx3} = \left(\frac{\bar{X}_3 - \bar{X}_{3h}}{\bar{X}_3}\right)$$

where $\bar{y}_h$ and $\bar{Y}_h$ denote respectively, the sample stratum mean and population stratum mean of the study variable Y while $\bar{X}_{ih}$ and $\bar{X}_i$ denote respectively, the population stratum mean and population mean of the ith auxiliary variable X_i.

So that

$$\bar{y}_h = \bar{Y}_h(1 + e_{hy}), \bar{X}_{1h} = \bar{X}_1(1 - e_{hx1}), \bar{X}_{2h} = \bar{X}_2(1 - e_{hx2}), \bar{X}_{3h} = \bar{X}_3(1 - e_{hx3})$$

$$E(e_{hy}^2) = \gamma_h C_{hy}^2, E(e_{hx1}^2) = \gamma_h C_{hx1}^2, E(e_{hx2}^2) = \gamma_h C_{hx2}^2, E(e_{hx3}^2) = \gamma_h C_{hx3}^2$$

$$E(e_{hy}e_{hx1}) = \gamma_h C_{hy}C_{hx1}, E(e_{hy}e_{hx2}) = \gamma_h C_{hy}C_{hx2}, E(e_{hy}e_{hx3}) = \gamma_h C_{hy}C_{hx3}$$

$$E(e_{hx1}e_{hx2}) = \gamma_h C_{hx1}C_{hx2}, E(e_{hx1}e_{h3}) = \gamma_h C_{hx1}C_{hx3},$$

$$E(e_{hx2}e_{hx3}) = \gamma_h C_{hx2}C_{hx3}$$

Expressing (11) in terms of the e's gives

$$\bar{y}_{st}^*(MREG) - \bar{Y} = \sum_{h=1}^{H} W_h[\bar{Y}_h e_{hy} + \beta_1\bar{X}_1 e_{hx1} + \beta_2\bar{X}_2 e_{hx2} + \beta_3\bar{X}_3 e_{hx3}] \qquad (12)$$

Squaring both sides of (12) gives

$$[\bar{y}_{st}^{*}(MREG)-\bar{Y}]^{2}=\sum_{h=1}^{H}W_h^2\,[\bar{Y}_h^2e_h^2+\beta_{1h}^2\bar{X}_1^2e_{hx1}^2+\beta_{2h}^2\bar{X}_2^2e_{hx2}^2+\beta_{3h}^2\bar{X}_3^2e_{hx3}^2$$
$$+2\bar{Y}_h\beta_{1h}\bar{X}_1e_{hy}e_{hx1}+2\bar{Y}_h\beta_{2h}\bar{X}_2e_{hy}e_{hx2}+2\bar{Y}_h\beta_{3h}\bar{X}_3e_{hy}e_{hx3}$$
$$+2\beta_{1h}\beta_{2h}\bar{X}_1\bar{X}_2e_{hx1}e_{hx2}+2\beta_{1h}\beta_{3h}\bar{X}_1\bar{X}_3e_{hx1}e_{hx3}$$
$$+2\beta_{2h}\beta_{3h}\bar{X}_2\bar{X}_3e_{hx2}e_{hx3}] \tag{13}$$

Taking expectation of both sides of (13) gives

$$V[\bar{y}_{st}^{*}(MREG)]=\sum_{h=1}^{H}W_h^2\gamma_h\,[\bar{Y}_h^2C_{hy}^2+\beta_{1h}^2\bar{X}_1^2C_{hx1}^2+\beta_{2h}^2\bar{X}_2^2C_{hx2}^2$$
$$+2\bar{Y}_h\beta_{1h}\bar{X}_1\rho_{hyx1}C_{hy}C_{hx1}+2\bar{Y}_h\beta_{2h}\bar{X}_2\rho_{hyx2}C_{hy}C_{hx2}+2\bar{Y}_h\beta_{3h}\bar{X}_3\rho_{hyx3}C_{hy}C_{hx3}$$
$$+2\beta_{1h}\beta_{2h}\bar{X}_1\bar{X}_2\rho_{hx1x2}C_{hx1}C_{hx2}+2\beta_{1h}\beta_{3h}\bar{X}_1\bar{X}_3\rho_{hx1x3}C_{hx1}C_{hx3}$$
$$+2\beta_{2h}\beta_{3h}\bar{X}_2\bar{X}_3\rho_{hx2x3}C_{hx2}C_{hx3}] \tag{14}$$

3.1.1 Optimality condition

This section deduced the optimality conditions that would guarantee optimum performance of the estimator on satisfaction. Setting

$$\frac{\partial\hat{V}[\bar{y}_{st}^{*}(MREG)]}{\partial\beta_1}=0,\qquad \frac{\partial\hat{V}[\bar{y}_{st}^{2}(MREG)]}{\partial\beta_2}=0\ \ and\ \ \frac{\partial\hat{V}[\bar{y}_{st}^{2}(MREG)]}{\partial\beta_3}=0$$

respectively, gives the following system of equations

$$\begin{bmatrix}\bar{X}_1^2C_{hx1}^2 & \bar{X}_1\bar{X}_2\rho_{hx1x2}C_{hx1}C_{hx2} & \bar{X}_1\bar{X}_3\rho_{hx1x3}C_{hx1}C_{hx3}\\ \bar{X}_1\bar{X}_2\rho_{hx1x2}C_{hx1}C_{hx2} & \bar{X}_2^2C_{hx2}^2 & \bar{X}_2\bar{X}_3\rho_{hx2x3}C_{hx2}C_{hx3}\\ \bar{X}_1\bar{X}_3\rho_{hx1x3}C_{hx1}C_{hx3} & \bar{X}_2\bar{X}_3\rho_{hx2x3}C_{hx2}C_{hx3} & \bar{X}_3^2C_{hx3}^2\end{bmatrix}\begin{bmatrix}\beta_{1h}\\ \beta_{2h}\\ \beta_{3h}\end{bmatrix}$$
$$=-\begin{bmatrix}\bar{Y}_h\bar{X}_1\rho_{hyx1}C_{hy}C_{hx1}\\ \bar{Y}_h\bar{X}_2\rho_{hyx2}C_{hy}C_{hx2}\\ \bar{Y}_h\bar{X}_3\rho_{hyx3}C_{hy}C_{hx3}\end{bmatrix} \tag{15}$$

Solving the system of equations in (15) gives

$$\beta_{1h}=\frac{\bar{Y}_hC_{hy}\big[\rho_{hx1x3}+\rho_{hyx1}\rho_{hx2x3}+\rho_{hx1x2}\rho_{hyx2}-\rho_{hx1x3}\rho_{hx2x3}\rho_{hyx3}-\rho_{hx1x2}\rho_{hyx3}-\rho_{hyx1}\big]}{\bar{X}_1C_{hx1}[1+2\rho_{hx1x2}\rho_{hx1x3}\rho_{hx2x3}-\rho_{hx1x2}^2-\rho_{hx1x3}^2-\rho_{hx2x3}^2]} \tag{16}$$

$$\beta_{2h}=\frac{\bar{Y}_hC_{hy}\big[\rho_{hx1x3}^2\rho_{hyx2}+\rho_{hx2x3}\rho_{hyx3}+\rho_{hx1x2}\rho_{hyx1}-\rho_{hyx2}-\rho_{hx1x2}\rho_{hx1x3}\rho_{hyx3}-\rho_{hx1x3}\rho_{hx2x3}\rho_{hyx1}\big]}{\bar{X}_2C_{hx2}[1+2\rho_{hx1x2}\rho_{hx1x3}\rho_{hx2x3}-\rho_{hx1x2}^2-\rho_{hx1x3}^2-\rho_{hx2x3}^2]} \tag{17}$$

$$\beta_{3h}=\frac{\bar{Y}_hC_{hy}\big[\rho_{hx1x2}^2\rho_{hyx3}+\rho_{hx1x3}\rho_{hyx3}+\rho_{hx2x3}\rho_{hyx2}-\rho_{hx1x3}\rho_{hyx2}-\rho_{hx2x3}\rho_{hyx1}-\rho_{hyx3}\big]}{\bar{X}_3C_{hx3}[1+2\rho_{hx1x2}\rho_{hx1x3}\rho_{hx2x3}-\rho_{hx1x2}^2-\rho_{hx1x3}^2-\rho_{hx2x3}^2]} \tag{18}$$

Substituting [(16), (17), (18)] in (15), the resulting (optimum) estimator of variance of $\bar{y}_{st}^{*}(MREG)$ is given by:

$$\hat{V}_{opt}[\bar{y}^*_{st}(MREG)] = \sum_{h=1}^{H} w_h^2 \gamma_h \bar{Y}_h^2 C_{hy}^2 \left[1 + B_{1h}^2 + B_{2h}^2 + B_{3h}^2 + 2\rho_{hyx1}B_{1h} + 2\rho_{hyx2}B_{2h} + 2\rho_{hyx3}B_{3h} + 2\rho_{hx1x2}B_{1h}B_{2h} + 2\rho_{hx1x3}B_{1h}B_{3h} \; 2\rho_{hx2x3}B_{2h}B_{3h}\right] \quad (19)$$

Where

$$B_{1h} = \frac{[\rho_{hx1x3} + \rho_{hyx1}\rho_{hx2x3} + \rho_{hx1x2}\rho_{hyx2} - \rho_{hx1x3}\rho_{hx2x3}\rho_{hyx3} - \rho_{hx1x2}\rho_{hyx3} - \rho_{hyx1}]}{[1 + 2\rho_{hx1x2}\rho_{hx1x3}\rho_{hx2x3} - \rho_{hx1x2}^2 - \rho_{hx1x3}^2 - \rho_{hx2x3}^2]}$$

$$B_{2h} = \frac{[\rho_{hx1x3}^2\rho_{hyx2} + \rho_{hx2x3}\rho_{hyx3} + \rho_{hx1x2}\rho_{hyx1} - \rho_{hyx2} - \rho_{hx1x2}\rho_{hx1x3}\rho_{hyx3} - \rho_{hx1x3}\rho_{hx2x3}\rho_{hyx1}]}{[1 + 2\rho_{hx1x2}\rho_{hx1x3}\rho_{hx2x3} - \rho_{hx1x2}^2 - \rho_{hx1x3}^2 - \rho_{hx2x3}^2]}$$

$$B_{3h} = \frac{[\rho_{hx1x2}^2\rho_{hyx3} + \rho_{hx1x3}\rho_{hyx3} + \rho_{hx2x3}\rho_{hyx2} - \rho_{hx1x3}\rho_{hyx2} - \rho_{hx2x3}\rho_{hyx1} - \rho_{hyx3}]}{[1 + 2\rho_{hx1x2}\rho_{hx1x3}\rho_{hx2x3} - \rho_{hx1x2}^2 - \rho_{hx1x3}^2 - \rho_{hx2x3}^2]}$$

3.2 Analysis of variance (ANOVA) method

In this section, the estimator of variance of the proposed multivariate calibration regression (*M-REG*) estimator is derived by the analysis of variance (*ANOVA*) approach.

Let the corrected sum of products be defined by

$$S_{ij} = \sum_{h=1}^{H} (\bar{X}_{ih} - \bar{X}_i)(\bar{X}_{jh} - \bar{X}_j), \qquad i, j = 1,2, \dots m$$

So that

$$S_{0i} = \sum_{h=1}^{H} (\bar{y}_h - \bar{Y})(\bar{X}_{ih} - \bar{X}_i), \; i = 1,2, \dots m$$

where m denotes the number of auxiliary variables

Let the sum of square regression (*SSR)* be defined by

$$S_{0R} = \sum_{i=1}^{m} \hat{\beta}_i S_{0i} \quad (20)$$

where $\hat{\beta}_i = C_{ij}S_{0i}$ and C_{ij} is the reciprocal of S_{ij}.

Let the sum of square residual be defined by:

$$S_{00} = \sum_{h=1}^{H} w_h^{*2}(\bar{y}_h - \bar{Y})^2 \quad (21)$$

where w_h^* are the calibration weights minimizing the Chi-square loss function

$$L(W_h^*, W_h) = \sum_{h=1}^{H} \frac{(W_h^* - W_h)^2}{W_h Q_h} \quad (22)$$

Subject to calibration constraint defined by

$$\sum_{h=1}^{H}\sum_{i=1}^{m} W_h^* S_{0i} = \sum_{h=1}^{H}\sum_{i=1}^{m} W_h \bar{X}_{ih}^2 \tag{23}$$

So that

$$\Delta_1 = \sum_{h=1}^{H} \frac{(W_h^* - W_h)^2}{W_h W_h} - 2\vartheta\left(\sum_{h=1}^{H}\sum_{i=1}^{m} W_h^* S_{0i} - \sum_{h=1}^{H}\sum_{i=1}^{m} W_h \bar{X}_{ih}^2\right)$$

Setting

$$\frac{\partial \Delta_1}{\partial w_h^*} = 0$$
$$W_h^* = W_h + \vartheta W_h Q_h S_{0i} \tag{24}$$

Substituting (24) in (23) gives

$$\vartheta = \frac{\sum_{h=1}^{H}\sum_{i=1}^{m} W_h(\bar{X}_{ih}^2 - S_{0i})}{\sum_{h=1}^{H}\sum_{i=1}^{m} W_h Q_h S_{0i}^2} \tag{25}$$

Substituting (25) in (24) gives

$$W_h^* = W_h + \sum_{h=1}^{H}\sum_{i=1}^{m} W_h(\bar{X}_{ih}^2 - S_{0i}) \frac{W_h Q_h S_{0i}}{\sum_{h=1}^{H}\sum_{i=1}^{m} W_h Q_h S_{0i}^2} \tag{26}$$

Setting the calibration tuning parameter $Q_h = S_{0i}^{-1}$ and substituting in (21) gives

$$S_{00} = \sum_{h=1}^{H} W_h^2(\bar{y}_h - \bar{Y})^2 \left(\sum_{h=1}^{H}\sum_{i=1}^{m} W_h \bar{X}_{ih}^2\right)^2 \left(\sum_{h=1}^{H}\sum_{i=1}^{m} W_h S_{01}\right)^{-2} \tag{27}$$

The variance of $\bar{y}_{st}^*(MREG)$ given X_1, X_2, X_3 is estimated by:

$$\hat{V}[\bar{y}_{st}^*(MREG)] = \frac{S_{00}}{n-(m+1)} \tag{28}$$

where S_{00} is the sum of squares of error, $n-(m+1)$ is the degree of freedom associated with the error, n is the set of measurements and m is the number of auxiliary variables.

Table 1. ANOVA Table

Source	**DF**	**SS**	**MS**	**F-Ratio**
Factor	m	$\sum_{i=1}^{m} \hat{\beta}_i S_{0i}$	$\frac{\sum_{i=1}^{m} \hat{\beta}_i S_{0i}}{m}$	$\frac{\sum_{i=1}^{m} \hat{\beta}_i S_{0i}[n-(m+1)]}{S_{00}(m)}$
Error	$n-m-1$	S_{00}	$\frac{S_{00}}{n-m-1}$	$-$
Total	$n-1$	$S_{00} + \sum_{i=1}^{m} \hat{\beta}_i S_{0i}$	$-$	$-$

4 Empirical Study

To judge the relative performances of the proposed *ANOVA* approach to variance estimation [with respect to the estimation of variance of the proposed (*M-REG)* estimator] over the *LASAP* approach, the data set in table 2 was considered. The MSE of the proposed multivariate calibration regression (*M-REG*) estimator under the *LASAP* Method and the *ANOVA* Method of variance estimation are given in Table 4.

4.1 Isaki (1983) multivariate regression estimator

The multivariate regression estimator in stratified random sampling established by [23] is given by

$$\hat{S}^2_{MR} = S^2_{yh} + \sum_{h=1}^{H} \mathrm{B}_{ih}\left(S^2_{x_{ih}} - s^2_{x_{ih}}\right) \tag{29}$$

with variance estimator given by:

$$\hat{V}\left(\hat{S}^2_{MR}\right) = V\left(S^2_{yh}\right) + \sum_{h=1}^{H} \mathrm{B}^2_{ih}\, V\left(S^2_{xih}\right) - 2\sum_{h=1}^{H} \mathrm{B}_{ih} Cov\left(S^2_{yh}, S^2_{xih}\right) + \sum_{i \neq j} \mathrm{B}_{ih}\mathrm{B}_{jh} Cov\left(S^2_{xih}, S^2_{xjh}\right) \tag{30}$$

4.2 The percent relative efficiency (PRE)

The percent relative efficiency (PRE) of an estimator θ with respect to the [23] multivariate regression estimator $\left(\hat{S}^2_{MR}\right)$ in stratified sampling is defined by

$$PRE\left(\theta, \hat{S}^2_{MR}\right) = \frac{V\left(\hat{S}^2_{MR}\right)}{V(\theta)} \times 100 \tag{31}$$

Table 2. Data statistics

Parameter	Stratum 1	Stratum 2	Stratum 3	Stratum 4
N_h	10	9	26	7
n_h	3	2	5	2
w_h	0.300	0.2222	0.1923	0.2857
$\bar{X}_{1h}$	11.90	10.38	12.120	11.98
$\bar{X}_{2h}$	9.880	8.120	9.860	9.740
$\bar{X}_{3h}$	10.75	10.80	9.780	10.84
$\bar{Y}_h$	15.72	14.84	13.46	16.32
C_{hy}	1.062	0.986	1.208	1.023
C_{hx1}	1.234	1.306	1.032	0.926
C_{hx2}	1.165	0.946	1.010	1.062
C_{hx3}	1.246	0.864	1.026	0.926
ρ_{hyx1}	0.940	0.900	0.840	0.890
ρ_{hyx2}	0.820	0.860	0.920	0.780
ρ_{hyx3}	0.923	0.968	0.842	0.956
ρ_{hx1x2}	0.860	0.800	0.760	0.840
ρ_{hx1x3}	0.910	0.942	0.864	0.760
ρ_{hx2x3}	0.840	0.860	0.780	0.920
$Mean$	$\bar{X}_1 = 11.04$	$\bar{X}_2 = 9.04$	$\bar{X}_3 = 10.20$	$\bar{Y} = 12.62$

$$S_{ij} = \begin{pmatrix} 3.2252 & 2.8732 & 0.225 \\ 2.8732 & 2.7144 & 0.0136 \\ 0.225 & 0.0136 & 1.2485 \end{pmatrix}, C_{ij} = \begin{pmatrix} 6.7394 & -7.1279 & -1.1369 \\ -7.1279 & 7.9074 & 1.1984 \\ -1.1369 & 1.1984 & 0.9928 \end{pmatrix}$$

$$S_{0i} = \begin{pmatrix} 5.586 \\ 3.8404 \\ 5.0522 \end{pmatrix}, \quad \hat{\beta}_{ih} = \begin{pmatrix} 4.5285 \\ -3.3943 \\ 3.2674 \end{pmatrix}$$

$$\hat{V}_{opt}[\bar{y}^*_{st}(MREG)] = 513.0802$$

$$\hat{V}(\hat{S}^2_{MR}) = 826.4326$$

Table 3. ANOVA Table

Source	DF	SS	MS	F-Ratio
Factor	3	28.7683	9.5894	0.0626
Error	8	1,226.3428	153.2928	–
Total	11	–	–	–

Table 4. *MSE* and *PREs* of the proposed estimator under the *LASAP* and *ANOVA* methods

S/No	Estimator	MSE	PREs
1.	$\hat{S}^2_{MR}$	826.4326	100
2.	$\bar{y}^*_{st}(MREG)$ ***ANOVA Method***	153.2928	539.1202
3.	$\bar{y}^*_{st}(MREG)$ ***LASAP Method***	513.0802	161.0728

6 Discussion of Results

Numerical results for the percent relative efficiency (*PREs*) in table 4 reveals that the proposed *ANOVA* approach to variance estimation with respect to the estimator of variance of the proposed *M-REG* estimator has 439 percent gains in efficiency while the *LASAP* approach to variance estimation with respect to the estimator of variance of the proposed *M-REG* estimator has 61 percent gains in efficiency; this shows that the proposed *ANOVA* approach to variance estimation is 378 percent more efficient than the conventional *LASAP* method of variance estimation. This means that in using our proposed *ANOVA* approach to variance estimation with respect to the estimator of variance of the proposed *M-REG* estimator, one will have 378 percent efficiency gain over the conventional *LASAP* method of variance estimation. Also the proposed multivariate calibration regression (*M-REG*) estimator is more efficient than the [23] multivariate regression estimator ($\hat{S}^2_{MR}$).

7 Conclusion

Sequel to the discussion of results above, it is concluded that the proposed *ANOVA* approach to variance estimation fares better than the conventional *LASAP* method of variance estimation. The new approach to variance estimation is very efficient in estimating populations with multiple auxiliary variables.

Therefore, the new approach to variance estimation is very attractive to survey researchers as it gives consistent and more precise estimates of the population parameters.

Competing Interests

Author has declared that no competing interests exist.

References

[1] Deville JC, Sarndal CE. Calibration estimators in survey sampling. Journal of the American Statistical Association. 1992;87:376-382.

[2] Wu C, Sitter RR. A model-calibration approach to using complete auxiliary information from survey data. Journal of the American Statistical Association. 2001;96:185-193.

[3] Montanari GE, Ranalli MG. Nonparametric model calibration estimation in survey sampling. Journal of the American Statistical Association. 2005;100:1429-1442.

[4] Farrell PJ, Singh S. Model-assisted higher order calibration of estimators of variance. Australia and New Zealand Journal of Statistics. 2005;47(3):375-383.

[5] Arnab R, Singh S. A note on variance estimation for the generalized regression predictor. Australia and New Zealand Journal of Statistics. 2005;47(2):231-234.

[6] Estavao VM, Sarndal CE. Survey estimates by calibration on complex auxiliary information. International Statistical Review. 2006;74:127-147.

[7] Kott PS. Using calibration weighting to adjust for non-response and coverage errors Survey Methodology. 2006;32:133-142.

[8] Kim JM, Sungur EA, Heo TY. Calibration approach estimators in stratified sampling. Statistics and Probability Letters. 2007;77(1):99-103.

[9] Sarndal CE. The calibration approach in survey theory and practice. Survey Methodology. 2007;33: 99-119.

[10] Kim JK, Park M. Calibration estimation in survey sampling. International Statistical Review. 2010; 78(1):21-29.

[11] Rao D, Khan MGM, Khan S. Mathematical programming on multivariate calibration estimation in stratified sampling. World Academy of Science. Engineering and Technology. 2012;72:12-27.

[12] Koyuncu N, Kadilar C. Calibration estimators using different measures in stratified random sampling. International Journal of Modern Engineering Research. 2013;3(1):415-419.

[13] Clement EP. Calibration approach separate ratio estimator for population mean in stratified sampling. International Journal of Modern Mathematical Sciences. 2015;13(4):377-384.

[14] Clement EP. A new ratio estimator of mean in survey sampling by calibration estimation elixir statistics. 2017;106:46461-46465.

[15] Clement EP. Estimation of population mean in calibration ratio-type estimator under systematic sampling elixir statistics. 2017;106:46480-46486.

[16] Clement EP, Udofia GA, Enang EI. Sample design for domain calibration estimators. International Journal of Probability and Statistics. 2014;3(1):8-14.

[17] Clement EP, Udofia GA, Enang EI. Estimation for domains in stratified sampling design in the presence of non-response. Aligarch Journal of Statistics. 2015;35:117-132.

[18] Clement EP, Enang EI. Multivariate calibration estimation for domain in stratified random sampling. International Journal of Modern Mathematical Sciences. 2015;13(2):187-197.

[19] Clement EP, Enang EI. Calibration approach alternative ratio estimator for population mean in stratified sampling. International Journal of Statistics and Economics. 2015;16(1):83-93.

[20] Clement EP, Enang EI. On the efficiency of ratio estimator over the regression estimator. Communications in Statistics - Theory and Methods. 2017;46(11):5357-5367. DOI: 10.1080/03610926.2015.1100741

[21] Enang EI, Clement EP. An efficient class of calibration ratio estimators of domain mean in survey sampling. Communication in Mathematics and Statistics; 2019. DOI: 10.1007/s40304-018-00174-z

[22] Tracy DS, Singh S, Arnab R. Note on calibration in stratified and double sampling. Survey Methodology. 2003;29:99-104.

[23] Isaki CT. Variance estimation using auxiliary information. Journal of the American Statistical Association. 1983;78:117-123.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://www.sdiarticle4.com/review-history/61000