Calibration of Stratified Random Sampling with Combined Ratio Estimators

JAMIU OLASUNKANMI MUILI*, RAN VIJAY KUMAR SINGH†, GERALD IKECHUKWU ONWUKA‡, and AHMED AUDU§

†Department of Mathematics, Kebbi State University of Science and Technology Aliero, Nigeria. ‡Department of Statistics, Usmanu Danfodiyo University, Sokoto, Nigeria.

Abstract
This study considered modification of combined ratio type calibration estimators in stratified random sampling using calibration estimation approaches. The estimators of population mean in stratified random sampling depends on the strata estimated sample means. However, the means are sensitive to the extreme values or outliers in the sample observations of the study variables and strata sizes respectively. A new sets of calibration weights and property of the suggested combined calibration estimators of population mean in stratified sampling were derived. Empirical study through simulation was conducted to investigate the efficiency of the modified combined ratio-type calibration estimators of population mean obtained and the results revealed that the suggested estimators of population mean performed better than some existing estimators considered in the study.

Introduction
Proper utilization of auxiliary information to obtain the efficiency of estimates of the population mean has increased in the theory of sample surveys. Many researchers have used auxiliary information in product, ratio and regression type estimators to obtain more efficient estimator under different sampling scheme. Calibration resolution is used in stratified random sampling in order to achieve optimum strata weights for precision improvement of estimates of parameters. In calibration estimation, new stratum weights are calculated to minimize a certain distance measure from the original design weights while meeting a set of auxiliary information restrictions. Deville and Sarndal1 established the approach of estimate by calibration in survey sampling in 1992. The concept is to employ auxiliary data (auxiliary information) to improve a population...
statistic estimate. Following Deville and Sarndal, Singh et al. was the first to extend a method of calibrating to a stratified sampling design. Many other researchers have investigated calibration estimates in survey sampling design utilizing various calibration constraints. These researchers include Singh, Tracy et al., Kim et al., Clement and Enang, Koyuncu and Kadilar. In stratified sampling, Rao et al. suggested a multivariate calibration estimator for the population mean based on different distance measures and two auxiliary variables. In the previous studied, none have considered calibration estimation in combined ratio estimators. In this study, calibration approaches have been adopted in combined ratio estimator with aim to obtain highly efficient estimators of population mean in stratified random sampling. The presence of extreme values in the observation of the study variable have no or little effect on the other estimates.

Take a look at a finite population of elements, \( T = \{ T_1, T_2, T_3, \ldots, T_N \} \) consists of \( L \) strata with \( N_h \) units in the \( h \)th stratum from which a simple random of size \( n \) can be generated from the population using SRSWOR. Total Population size \( N \), sample size \( n \), where \( y_{ih} \), \( i=1,2,\ldots, N_h \) and \( x_{ih} \), \( i=1,2,\ldots, N_h \) of auxiliary variable \( x \) and study variable \( y \). Let \( W_h = N_h / N \) be the weights of the strata, \( \bar{y}_i = \sum_{i=1}^{N_h} y_{ih} / N_h \) population mean for the study variable.

**Literature Review**

According to Cochran, with stratified sampling, the classic estimator of population mean is given as:

\[
\bar{y}_u = \frac{1}{n} \sum_{i=1}^{n} W_h \bar{y}_h \quad \text{...(2.1)}
\]

\[
V(\bar{y}_u) = \sum_{h=1}^{L} W_h^2 \left( \frac{1 - \bar{y}_h}{n_h} \right) s_y^2 \quad \text{...(2.2)}
\]

where \( s_y^2 = \sum_{h=1}^{L} \sum_{i=1}^{N_h} (y_{ih} - \bar{y}_h)^2 \)

Hansen et al. suggested a combined ratio estimator as:

\[
\bar{y}_{CR} = \frac{1}{n} \sum_{i=1}^{N_h} W_h \bar{x}_h - \bar{x} \quad \text{...(2.3)}
\]

The combined ratio estimator's variance is given as follows:

\[
V(\bar{y}_{CR}) = \sum_{h=1}^{L} W_h^2 \bar{y}_h \left( s_y^2 + R^2 s_x^2 - 2RS_{xy} \right) \quad \text{...(2.4)}
\]

where \( R = \frac{\bar{y}}{\bar{x}} \)

Singh et al. presented the calibration approach for the combined general regression estimator (GREG) for the population mean given by

\[
\bar{y}_{GREG} = \sum_{h=1}^{L} \Omega_h^{SFF} \bar{y}_h \quad \text{...(2.5)}
\]

By minimizing the Chi-Square distance measure, Singh et al. were able to get new calibration weights.

\[
Z_i = \left( \sum_{h=1}^{L} \Omega_h^{SFF} - W_h \right) ^2 / Q_h W_h \quad \text{...(2.6)}
\]

subject to the constraint

\[
\sum_{h=1}^{L} \Omega_h^{SFF} \frac{\bar{x}_h}{\bar{y}_h} = \bar{x} \quad \text{...(2.7)}
\]

The calibrated weights and the estimator are obtained as show in (2.6) and (2.7) respectively.

\[
\Omega_h^{SFF} = W_h + \left( \sum_{i=1}^{n_h} W_i \bar{x}_i / \sum_{i=1}^{n_h} W_i \bar{x}_i ^2 \right) \left( \bar{x} - \sum_{i=1}^{n_h} W_i \bar{x}_i / n_h \right) \quad \text{...(2.8)}
\]

As a result, the modified calibrated estimator of Singh et al. is

\[
\bar{y}_{MC} = \sum_{h=1}^{L} W_h \bar{y}_h + \left( \sum_{i=1}^{n_h} W_i \bar{x}_i / \sum_{i=1}^{n_h} W_i \bar{x}_i ^2 \right) \left( \bar{x} - \sum_{i=1}^{n_h} W_i \bar{x}_i / n_h \right) \quad \text{...(2.9)}
\]

The calibrated estimator's \( \bar{y}_{MC} \) estimated variance is given by

\[
V(\bar{y}_{MC}) = \sum_{h=1}^{L} \Omega_h^{SFF} s_y^2 \quad \text{...(2.10)}
\]

where \( s_y^2 = \sum_{h=1}^{L} \sum_{i=1}^{N_h} (y_{ih} - \bar{y}_h)^2 \)

is the \( h \)th strata sample mean square and \( e_{ih} = (y_{ih} - \bar{y}_h) - b(x_{ih} - \bar{x}_h) \)
with \( b = \sum_{i=1}^{L} w_i \bar{y}_i \)  

Singh\(^1\) introduced new calibration equations to a population mean calibration estimator in stratified sampling. Under stratified sampling, the Singh\(^1\) calibration estimator of the population mean \( \bar{Y} \) is given by

\[
\bar{Y}_{cal} = \frac{\sum_{h=1}^{H} \Omega_h \bar{y}_h}{\sum_{h=1}^{H} \Omega_h} \quad (2.11)
\]

where \( \Omega_h \) is calibrated weight which is chosen so that the Chi-Square distance mean sum is

\[
Z_s = \sum_{h=1}^{H} \left( \Omega_h - w_h \right)^2 / Q_h \quad (2.12)
\]

subject to the following constraints as a minimum

\[
\sum_{h=1}^{H} \Omega_h = 1 \quad (2.13)
\]

\[
\sum_{h=1}^{H} \Omega_h \bar{y}_h = \bar{X} \quad (2.14)
\]

The calibrated weight and the estimator (2.11) are obtained as show in (2.15) and (2.16) respectively

\[
\Omega_h = w_h + \frac{1}{Q_h} \left( \sum_{h=1}^{H} w_i \bar{y}_i - \left( \sum_{i=1}^{L} w_i \bar{y}_i \right) \right) \left( \sum_{h=1}^{H} w_h \bar{y}_h - \left( \sum_{h=1}^{H} w_h \bar{y}_h \right) \right)^{-1} \left( \sum_{h=1}^{H} w_h \bar{y}_h - \left( \sum_{h=1}^{H} w_h \bar{y}_h \right) \right) \quad (2.15)
\]

\[
\bar{Y}_{cal} = \frac{\sum_{h=1}^{H} w_h \bar{y}_h + \hat{\beta}(\bar{X} - \sum_{h=1}^{H} w_h \bar{y}_h)}{\sum_{h=1}^{H} w_h} \quad (2.16)
\]

where

\[
\hat{\beta}(\bar{X} - \sum_{h=1}^{H} w_h \bar{y}_h) = \frac{1}{Q_h} \left( \sum_{h=1}^{H} w_h \bar{y}_h - \left( \sum_{h=1}^{H} w_h \bar{y}_h \right) \right) \left( \sum_{h=1}^{H} w_h \bar{y}_h - \left( \sum_{h=1}^{H} w_h \bar{y}_h \right) \right)^{-1} \left( \sum_{h=1}^{H} w_h \bar{y}_h - \left( \sum_{h=1}^{H} w_h \bar{y}_h \right) \right) \quad (2.21)
\]

The estimated variance of Singh\(^1\) calibrated estimator is given by

\[
\text{Var} \left( \bar{y}_{cal} \right) = \sum_{h=1}^{H} \Omega_h s_h^2 \quad (2.17)
\]

where \( s_h^2 = (n_h - 1)^{-1} \sum_{i=1}^{n_h} e_i^2 \) is the mean square of the hth stratum sample while

\[
\hat{\beta}_d = \frac{1}{Q_h} \left( \sum_{h=1}^{H} w_h \bar{y}_h - \left( \sum_{h=1}^{H} w_h \bar{y}_h \right) \right) \left( \sum_{h=1}^{H} w_h \bar{y}_h - \left( \sum_{h=1}^{H} w_h \bar{y}_h \right) \right)^{-1} \left( \sum_{h=1}^{H} w_h \bar{y}_h - \left( \sum_{h=1}^{H} w_h \bar{y}_h \right) \right) \quad (2.23)
\]

And the calibrated estimator is

\[
\bar{y}_{cal} = \sum_{h=1}^{H} w_h \bar{y}_h - \hat{\beta}_d \left( \bar{X} - \sum_{h=1}^{H} w_h \bar{y}_h \right) \quad (2.22)
\]

The estimated variance of Alam et al.\(^1\) calibration estimator is given by

\[
\text{Var} \left( \bar{y}_{cal} \right) = \sum_{h=1}^{H} \Omega_h s_h^2 \quad (2.24)
\]
Proposed Estimator

In stratified random sampling, the conventional Combined Ratio Estimator is given in (2.3) can be written as

\[ \bar{y}^{RC}_{st} = \frac{\sum h=1^{L} W^*_h \bar{y}_h}{\sum h=1^{L} W^*_h \bar{x}_h} \] \hspace{1cm} \text{(2.25)}

where \( W^*_h = \frac{W_h x}{\sum h=1^{L} W_v x_h} \)

Motivated by Alam et al., calibrated combined ratio estimator denoted by \( \bar{y}^{RC}_{st(M)} \) is modified as

\[ \bar{y}^{RC}_{st(M)} = \sum h=1^{L} \Omega_h \bar{y}_h \] \hspace{1cm} \text{(2.26)}

where \( \Omega_h \) denotes the new calibrated weights that minimize the Chi-square distance.

\[ Z^* = \frac{1}{2} \sum h=1^{L} \left( \Omega^*_h - W^*_h \right)^2 + \frac{1}{2} \sum h=1^{L} \sum_{v=1}^{V} \left( \Omega^*_v - W^*_v \right) \left( \Omega^*_v - W^*_v \right) \] \hspace{1cm} \text{(2.27)}

Subject to calibration constraint given by

\[ \sum h=1^{L} \Omega_h \bar{x}_h = \sum h=1^{L} W^*_h \bar{x}_h \] \hspace{1cm} \text{(2.28)}

The Lagrange multipliers technique is employed to compute new calibrated weights(\( \Omega_h \)), and the following results are obtained

\[ \hat{\lambda} = \frac{\sum h=1^{L} W^*_h \bar{y}_h - \sum h=1^{L} W^*_h \bar{x}_h}{\sum h=1^{L} W^*_h Q_h \bar{x}_h} \] \hspace{1cm} \text{(2.29)}

Differentiating (2.29) partially in relation to \( \Omega_h \), and \( \hat{\lambda} \), equal to zero

\[ \Omega_h = W^*_h + \hat{\lambda} \left( \frac{W^*_h Q_h \bar{x}_h}{1 - W^*_h Q_h} \right) \] \hspace{1cm} \text{(2.30)}

\[ \sum h=1^{L} \Omega_h \bar{x}_h - \sum h=1^{L} W^*_h \bar{x}_h = 0 \] \hspace{1cm} \text{(2.31)}

Substitute (2.30) in (2.31), the results are obtained as

\[ \hat{\lambda} \sum h=1^{L} \left( \frac{W^*_h Q_h \bar{x}_h^2}{1 - W^*_h Q_h} \right) = \sum h=1^{L} W^*_h \bar{y}_h - \sum h=1^{L} W^*_h \bar{x}_h \] \hspace{1cm} \text{(2.32)}

Make \( \hat{\lambda} \) the subject of formula, obtained as:

\[ \hat{\lambda} = \frac{\frac{\sum h=1^{L} W^*_h \bar{y}_h - \sum h=1^{L} W^*_h \bar{x}_h}{\sum h=1^{L} W^*_h Q_h \bar{x}_h}}{\sum h=1^{L} \left( \frac{W^*_h Q_h \bar{x}_h^2}{1 - W^*_h Q_h} \right)} \] \hspace{1cm} \text{(2.33)}

On substituting (2.33) in (2.30) the calibrated weights can be written as

\[ \Omega_h = W^*_h + \beta \sum h=1^{L} W^*_h \left( \bar{x}_h - \bar{x}_h \right) \] \hspace{1cm} \text{(2.34)}

Substituting \( W^*_h = W_h x / \sum h=1^{L} W_v x_h \) in (2.35), gives

\[ \bar{y}^{R(C)}_{st(M)} = \sum h=1^{L} W^*_h \bar{y}_h + \beta \sum h=1^{L} W^*_h \left( \bar{x}_h - \bar{x}_h \right) \] \hspace{1cm} \text{(2.36)}

where

\[ \hat{\beta} = \frac{\sum h=1^{L} \left( \frac{W^*_h Q_h \bar{x}_h}{1 - W^*_h Q_h} \right)}{\sum h=1^{L} \left( \frac{W^*_h Q_h \bar{x}_h^2}{1 - W^*_h Q_h} \right)} \] \hspace{1cm} \text{(2.37)}

Setting \( Q_h = 1 \), and \( Q_h = \bar{x}_h^{-1} \) we have the following new set of calibration combined ratio estimators respectively

\[ \bar{y}^{RC}_{st(M)} = \sum h=1^{L} W^*_h \bar{y}_h + \hat{\beta} \sum h=1^{L} W^*_h \left( \bar{x}_h - \bar{x}_h \right) \] \hspace{1cm} \text{(2.38)}

where

\[ \hat{\beta} = \frac{\sum h=1^{L} \left( \frac{W^*_h Q_h \bar{x}_h}{1 - W^*_h Q_h} \right)}{\sum h=1^{L} \left( \frac{W^*_h Q_h \bar{x}_h^2}{1 - W^*_h Q_h} \right)} \] \hspace{1cm} \text{(2.39)}

and

\[ \hat{\beta} = \frac{\sum h=1^{L} \left( \frac{W^*_h Q_h \bar{x}_h^2}{1 - W^*_h Q_h} \right)}{\sum h=1^{L} \left( \frac{W^*_h Q_h \bar{x}_h}{1 - W^*_h Q_h} \right)} \] \hspace{1cm} \text{(2.40)}
The suggested estimator’s estimated variance

\[ \text{Var} (\bar{y}_{st(M)}) = \sum_{i=1}^{L} \frac{\Omega_i^2}{n_{hi}^2} s_{eh}^2 \]  

...(2.39)

where \( s_{eh}^2 = (n_{hi} - 1)^{-1} \sum_{i=1}^{n_{hi}} \left( x_{ih} - \bar{x}_{ih} \right) \) is the hth strata sample mean square and \( \bar{x}_{ih} = \frac{1}{n_{hi}} \sum_{j=1}^{n_{hi}} x_{ij} \).

Empirical Study Using Simulation

In this section, a simulation research was carried out to see if the proposed estimators were better than the other estimators evaluated in the study.

For this investigation, 1000-unit data was generated. Using the function defined in Table 1, populations were stratified into three non-overlapping heterogeneous groups of 200, 300, and 500. Method SRSWOR was used to randomly choose samples of sizes 20, 30, and 50 from each stratum 10,000 times. The precision (PRE) of the estimators under consideration was calculated using (2.40)

\[ \text{PRE} (\hat{\theta}) = \left( \frac{\text{var} (\bar{y}_{st})}{\text{var} (\hat{\theta})} \right) 100 \]  

...(2.40)

where \( \text{var} (\hat{\theta}) = \frac{1}{10000} \sum_{j=1}^{10000} (\hat{\theta}_j - \bar{\theta})^2 \), \( \bar{\theta} = \frac{1}{n_{hi}} \sum_{j=1}^{n_{hi}} \bar{y}_{ij} \), \( \bar{y}_{ij} = \bar{y}_{st} + \bar{y}_{sh} \), \( \bar{y}_{sh} = \bar{y}_{st} - \bar{y}_{st(M)} \).

| Populations | Auxiliary variable \( x \) | Study variable \( y \) |
|-------------|-----------------------------|-------------------------|
| \( I \)     | \( x_h \approx \text{chisq}(\theta_1, \theta_2, \theta_3) \), \( \theta_1 = 5, \theta_2 = 6, \theta_3 = 4, n_{hi} = 1, 2, 3 \) | \( y_{hi} = \alpha_{hi} x_{hi}^2 + \xi_{hi}, \alpha_{hi} = E(x_{hi}) \), \( \alpha = 0.5, \xi_h \approx N(0,1), n_{hi} = 1, 2, 3 \) |
| \( II \)    | \( x_h \approx \text{gamma}(\theta_1, \eta_1), \theta_1 = 3, \eta_1 = 2 \), \( \theta_2 = 3, \eta_2 = 1, \theta_3 = 3, \eta_3 = 3 \) | \( y_{hi} = \alpha_{hi} x_{hi}^3 + \xi_{hi} \), \( \alpha_{hi} = E(x_{hi}) \), \( \alpha = 0.5, \xi_h \approx N(0,1), n_{hi} = 1, 2, 3 \) |

Table 1: Populations Involved in the Empirical Research

| Estimator | \( y_{hi} = \alpha_{hi} x_{hi}^2 + \xi_{hi} \) | \( Q_h = \frac{Q_{\bar{y}_{hi}}}{Q_{\bar{y}_{sh}}} \) |
|-----------|------------------------------------|-----------------|
| \( \bar{y}_{st} \) | 100 | 100 |
| Combined ratio \( \bar{y}_{st}^{RC} \) | 276.4634 | 276.4634 |
| Alam et al. et al. \( \bar{y}_{st}^{A} \) | 350.0407 | 361.6427 |

\[ \text{Suggested Estimator} \bar{y}_{st(M)}^{RC} = 483.9979 \]  

\[ \text{Suggested Estimator} \bar{y}_{st(M)}^{RC} = 459.3958 \]

Table 2: PREs of Some Existing and Suggested Estimators Using Population I

| Estimator | \( y_{hi} = \alpha_{hi} x_{hi}^3 + \xi_{hi} \) | \( Q_h = \frac{Q_{\bar{y}_{hi}}}{Q_{\bar{y}_{sh}}} \) |
|-----------|------------------------------------|-----------------|
| \( \bar{y}_{st} \) | 100 | 100 |
| Combined ratio \( \bar{y}_{st}^{RC} \) | 163.2217 | 163.2217 |
| Estimator          | $\bar{y}_{st}$ | $\bar{y}_{st}^{RC}$ |
|-------------------|---------------|---------------------|
| Alam et al.       | 184.4712      | 185.8651            |
| **Suggested Estimator** | **247.3609** | **246.3516**        |
| Estimator         | $\bar{y}_{st}$ |                     |
|                   | 100           | 100                 |
| Combined ratio    | $\bar{y}_{st}^{RC}$ |          |
|                   | 131.9602      | 131.9602            |
| Alam et al.       | 141.1144      | 140.7026            |
| **Suggested Estimator** | **165.7013** | **165.0921**        |

Table 3: PREs of Some Existing and Suggested Estimators Using Population II

| Estimator          | $\bar{y}_{st}$ | $\bar{y}_{st}^{RC}$ |
|-------------------|---------------|---------------------|
| $\bar{y}_{st}$    | 100           | 100                 |
| Combined ratio    | $\bar{y}_{st}^{RC}$ |          |
|                   | 197.4018      | 197.4018            |
| Alam et al.       | 272.0550      | 266.2928            |
| **Suggested Estimator** | **355.2338** | **345.9467**        |
| Estimator         | $\bar{y}_{st}$ |                     |
|                   | 100           | 100                 |
| Combined ratio    | $\bar{y}_{st}^{RC}$ |          |
|                   | 143.1531      | 143.1531            |
| Alam et al.       | 182.1500      | 177.3475            |
| **Suggested Estimator** | **233.6707** | **229.5379**        |
| Estimator         | $\bar{y}_{st}$ |                     |
|                   | 100           | 100                 |
| Combined ratio    | $\bar{y}_{st}^{RC}$ |          |
|                   | 124.2426      | 124.2426            |
| Alam et al.       | 144.5459      | 140.7251            |
| **Suggested Estimator** | **166.9500** | **163.7467**        |
Tables 2 and 3 shown the PREs of some existing and suggested estimators considered in this study using data generated by populations I and II respectively. From the results obtained, it revealed that the suggested estimators outperformed the existing estimators considered in the study.

Conclusion
From the results obtained, the empirical study revealed that on the efficiency of the suggested calibration estimators versus the study's current related estimators, the suggested estimators have higher PREs compared to some existing calibration estimators in the numerical analysis. The suggested estimators outperformed other calibration estimators because the suggested estimators demonstrated high level of efficiency over other estimators. Hence, the suggested estimators are closers to the true values of the population mean compared to other existing calibration estimators in which the suggested estimators have more chances of producing estimates that are closer to the population mean's true value.

Acknowledgements
The authors acknowledge the Associate Editor and anonymous reviewers for their valuable work and recommendations.

Funding
The authors received no financial support for the research, authorship and publication of this article.

Conflict of Interest
The authors declare no conflict of interest.

References

1. Alam S., Singh S. and Shabbir, J., New Methodology of Calibration in Stratified Random Sampling. *Survey Research Methods Section*, 90 – 95, (2019). doi:10.1080/00949655.2019.1688813
2. Clement E. P. and Enang E. I., Calibration Approach Alternative Ratio Estimator for Population Mean in Stratified Sampling. *International Journal of Statistics and Economics*, 16(1), 83-93, (2015).
3. Cochran W. G., Sampling Techniques. Third Edition. John Wiley and Sons Limited. 1977.
4. Deville J. C. and Särndal C. E., Calibration Estimators in Survey Sampling. *J. Amer. Statist. Assoc.*, 87, 376–382, (1992).
5. Hansen M. H., Hurwitz W. N. and Gurney M., The Problems and Methods of the Sample survey of business, *Journal of the American Statistical Association*, 41,173-189, (1946).
6. Kim J. M. Sungur E. A. and Heo T. Y., Calibration Approach Estimators in Stratified Sampling, *Statistics and Probability Letters*, 77(1), 99-103, (2007). doi:10.1016/j.spl.2006.05.015
7. Koyuncu N. and Kadilar C., Calibration Weighting in Stratified Random Sampling. *Communications in Statistics-Simulation and Computation*, 45, 2267-2275, (2016).
8. Rao D., Khan M. G. M., and Khan S., Mathematical Programming on Multivariate Calibration Estimation in Stratified Sampling. World Academy of Science, Engineering and Technology *International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering*, 6 (12), 1623-1627, (2012).
9. Singh S., Horn S., and Yu F., Estimation Variance of General Regression Estimator: Higher Level Calibration Approach. *Survey Methodology*, 48, 41-50, (1998).
10. Singh S., Advanced Sampling Theory with Applications. Dordrecht: Kluwer Academic Publishers, 2003.
11. Tracy D. S., Singh S., and Arnab R., Note on Calibration in Stratified and Double Sampling, *Survey Methodology*, 29, 99–104, (2003).