Iseh, Matthew Joshua; Enang, Ekaette Inyang

Article

A calibrated synthetic estimator for small area estimation

Statistics in Transition New Series

Provided in Cooperation with:
Polish Statistical Association

Suggested Citation: Iseh, Matthew Joshua; Enang, Ekaette Inyang (2021) : A calibrated synthetic estimator for small area estimation, Statistics in Transition New Series, ISSN 2450-0291, Exeley, New York, Vol. 22, Iss. 3, pp. 15-30, https://doi.org/10.21307/stattrans-2021-025

This Version is available at:
https://hdl.handle.net/10419/266269

STATISTICS IN TRANSITION new series, September 2021
Vol. 22, No. 3, pp. 15–30, DOI 10.21307/stattrans-2021-025
Received – 21.07.2020; accepted – 12.01.2021

A calibrated synthetic estimator for small area estimation

Matthew Joshua Iseh[1], **Ekaette Inyang Enang**[2]

ABSTRACT

Synthetic estimators are known to produce estimates of population mean in areas where no sampled data are available, but such estimates are usually highly biased with invalid confidence statements. This paper presents a calibrated synthetic estimator of the population mean which addresses these problematic issues. Two known special cases of this estimator were obtained in the form of combined ratio and combined regression synthetic estimators, using selected tuning parameters under stratified sampling. In result, their biases and variance estimators were derived. The empirical demonstration of the usage involving the proposed calibrated estimators shows that they provide better estimates of the population mean than the existing estimators discussed in this study. In particular, the estimators were examined through simulation under three distributional assumptions, namely the normal, gamma and exponential distributions. The results show that they provide estimates of the mean displaying less relative bias and greater efficiency. Moreover, they prove more consistent than the existing classical synthetic estimator. The further evaluation carried out using the coefficient of variation provides additional confirmation of the calibrated estimator's advantage over the existing ones in relation to small area estimation.

Key words: auxiliary variable, calibration estimation, simulation, synthetic estimation.

1. Introduction

The theory of small area estimation (SAE) revolves around the use of statistical modelling techniques to produce required estimates for several geographic sub-populations and socio demographic groups when the available survey data are not enough to calculate reliable direct estimates. The inherent challenges facing SAE revolve around finding the best statistical model to be fitted on the available data when a survey is designed for national purposes but preferably used for inferences about small areas to increase the accuracy of sub-national estimates and selecting the best

[1] Department of Statistics, Akwa Ibom State University, Mkpat Enin, Nigeria. E-mail: eeseaglechild@gmail.com. ORCID: https://orcid.org/0000-0003-2696-7319.
[2] Department of Statistics, University of Calabar, Calabar, Nigeria. E-mail: edikanenang@gmail.com. ORCID: https://orcid.org/0000-0002-1273-0436.

estimation method having known that SAE is likely to be used in the survey (Pandey and Tikkiwal 2007).

Authors like Gonzales (1973) and Sarndal (1981, 1984) have made useful contributions on the use of synthetic estimators in domains with zero/small sample sizes. Although the synthetic estimators have been shown to produce estimates for domains without sample units with an attractive property of small mean square error, it has also been noted that these estimators are sometimes characterized with large bias, hence, researchers are advised to apply caution in using this method (Sarndal, Swensson and Wretman 1992; Rao 2003; Rao and Choudhry 1995; Marker 1999).

In progression for improving the performance of small area estimators, a number of estimators have been constructed using weighted linear combination of different statistical principles like: the *empirical Bayes approach* by Fay and Herriot, (1979), *sample dependent (composite) method* by Drew Singh and Choudhry (1982), *Error Prediction approach* by Battesse and Fuller (1984). However, these techniques are identified with non-negligible bias and large MSE in areas with a small to modest sample size, which might constitute an invalid confidence interval.

In a bid to improve on the efficiency of small area estimators in the last two decades, several authors have proposed various types of estimators through the use of the calibration approach. In particular, among them are: Lundstrom and Sarndal (2001), Chambers (2006), Sarndal and Lundstrom (2007), Pfefferman (2013), Hidiroglou and Estavao (2014), Rao and Molina (2015), Clement and Enang (2017). Nevertheless, none of these works has considered improving on the post stratified synthetic estimator whose strength in producing estimates even in areas of no unit is based on the assumption of structural similarities. By keeping in mind the proposition of testability of the assumption of structural similarity of characteristics and a careful choice of auxiliary variable proposed by Rao (2003), this paper seeks to formulate synthetic estimators through calibration techniques in stratified sampling to reduce the bias and improve upon the precision of the synthetic estimators for small area.

2. Notations

Consider a finite population consisting of N units which is divided into D non-overlapping domains $U_d,\ d = 1,2,\ldots,D$ with N_d units such that $\sum_d^D N_d = N$. Let the population be further partitioned into G non-overlapping groups (considered to be strata) which are considered to be larger than the domains $U_{.g},\ g = 1,2,\ldots,G$ with $N_{.g}$ units such that $\sum_g^G N_{.g} = N$ so that the G groups cuts across the D domains to form a grid of DG cells denoted by U_{dg} with N_{dg} units such that $U = \bigcup_{d=1}^{D} U_d = \bigcup_{g=1}^{G} U_{.g} = \bigcup_{d=1}^{D}\bigcup_{g=1}^{G} U_{dg}$ and $N = \sum_d^D N_d = \sum_g^G N_{.g} = \sum_d^D \sum_g^G N_{dg}$. The sample s is analogously partitioned into domain subsamples s_d, group subsamples $s_{.g}$ and cells subsamples s_{dg}

with corresponding sample sizes n, n_d, $n_{.g}$ and n_{dg} as $s = \cup_{d=1}^{D} s_d = \cup_{g=1}^{G} s_{.g} = \cup_{d=1}^{D} \cup_{g=1}^{G} s_{dg}$ and $n = \sum_d^D n_d = \sum_g^G n_{.g} = \sum_d^D \sum_g^G n_{dg}$. The cells subsamples n_{dg} are assumed to be random. Ordinarily, n_d and $n_{.g}$ are also random but $n_{.g}$would be fixed if the g^{th} group is a stratum from which a fixed number of elements is drawn. Let Y be the study variable whose values y_{dgk} are known for just the element of a sample s, where $k = 1,2,\ldots,N_{dg}$ (the number of population units in the $(dg)^{th}$ cell) and X be the auxiliary variable whose values $x_{dgk} > 0$ may or may not be known *a priori* for all units in U.

Let the population mean $\bar{Y}_d$ for the domain be defined as $\bar{Y}_d = \sum_{g=1}^{G} W_{dg} \bar{Y}_{dg}$, where $\bar{Y}_{dg} = \sum_{k=1}^{N_{dg}} \frac{Y_{dgk}}{N_{dg}}$ is the population mean per $(dg)^{th}$ cell for the small area. A Horvitz-Thompson (1952)-type direct unbiased estimator of the population mean $\bar{Y}_d$ for the domain under stratified sampling is given as:

$$\hat{\bar{y}}_d = \sum_g^G W_{dg} \bar{y}_{dg} \tag{1}$$

where W_{dg} is the stratum weight given as $W_{dg} = N_d^{-1} N_{dg}$ and $\bar{y}_{dg} = \sum_{k=1}^{n_{dg}} \frac{y_{dgk}}{n_{dg}}$ is the sample mean per $(dg)^{th}$ cell for the small area. Equation (1) will perform at its best when n_d is sufficiently large as well as n_{dg}. However, under SAE, even if n_d is large, there is the likelihood that n_{dg} might turn out to be zero for some cells. Consequently, the direct estimator might be very unstable with large variance as well as lead to underestimation in areas with small sample sizes and also impossible to compute where there is no sample observation in the domain of interest. Under the aforementioned conditions, assuming that the groups g's $(g = 1,2,\ldots,G)$ are similar for small area d's $(d = 1,2,\ldots,D)$, Marker (1999) suggested the synthetic estimator of the average of characteristic Y for small area d, as:

$$\hat{\bar{y}}_d^s = \sum_g^G W_{dg} \bar{y}_{.g} \tag{2}$$

And the bias of Eq.2 given as $B(\hat{\bar{y}}_d^s) = \sum_g^G W_{dg} \left(\bar{Y}_{.g} - \bar{Y}_{dg}\right)$ with mean square error (MSE) as:

$$MSE(\hat{\bar{y}}_d^s) = \sum_g^G W_{dg}^2 \left(\frac{1}{n_{.g}} - \frac{1}{N_{.g}}\right) S_g^2 + \left[\sum_g^G W_{dg} \left(\bar{Y}_{.g} - \bar{Y}_{dg}\right)\right]^2,$$

where $\bar{y}_{.g} = \sum_d^D \sum_k^{N_{dg}} \frac{y_{dgk}}{n_{.g}}$ is the sample average of $\bar{Y}_{.g}$ (the population mean for the g^{th} subgroup) across all domains, $\bar{Y}_{dg}$ is the population mean for the d^{th} domain within the g^{th} subgroup and $S_g^2 = \left(N_{.g} - 1\right)^{-1} \sum_d^D \sum_k^{N_{dg}} \left(Y_{dgk} - \bar{Y}_{.g}\right)^2$.

It is assumed that if small areas are similar across the groups, $\bar{Y}_{.g} = \bar{Y}_{dg}$, then the synthetic estimator is almost unbiased. However, in practical terms this assumption of structural similarity of the characteristics within the groups might not hold, hence equation (2) becomes heavily biased.

To further enhance the efficiency of the direct estimator for domain estimation, Clement and Enang (2017) calibrated on equation (1) and obtained combined ratio and regression estimators for domain means under stratified sampling as follows:

$$\hat{\bar{y}}_{dc} = \sum_g^G W_{dg}\bar{y}_{dg} + \frac{\sum_g^G W_{dg}q_{dg}\bar{x}_{dg}\bar{y}_{dg}}{\sum_g^G W_{dg}q_{dg}\bar{x}_{dg}^2}\left(\bar{X} - \sum_g^G W_{dg}\bar{x}_{dg}\right). \quad (3)$$

By extension, setting $q_{dg} = \bar{x}_{dg}^{-1}$ and $q_{dg} = 1$ the authors obtained

$$\hat{\bar{y}}_{dcr} = \frac{\sum_g^G W_{dg}\bar{y}_{dg}}{\sum_g^G W_{dg}\bar{x}_{dg}}\bar{X} \quad (4)$$

as the calibration approach combined ratio estimator and

$$\hat{\bar{y}}_{dcreg} = \hat{\bar{y}}_d + b(\bar{X} - \hat{\bar{x}}_d). \quad (5)$$

as the calibration approach combined regression estimator respectively, where $\hat{\bar{x}}_d = \sum_g^G W_{dg}\bar{x}_{dg}$ is analogous to equation (1) is the domain direct estimator for the auxiliary variable, $\bar{y}_{dg}$ and $\bar{x}_{dg}$ are the cell means for the interest and auxiliary variables respectively and $b = \frac{\sum_g^G W_{dg}\bar{x}_{dg}\bar{y}_{dg}}{\sum_g^G W_{dg}\bar{x}_{dg}^2}$ is the regression coefficient.

Note: Although the estimators in equations (4) and (5) exhibited some level of improvements over that in equation (1), they perform poorly in areas with a small sample size and are impossible to compute in the domains of interest with no sample observation, hence the need for a modified synthetic estimator.

3. Calibrated synthetic estimators.

Consider the Marker (1999) synthetic estimator in equation (2) for a domain of interest with small or no sample observation. If small areas have similar characteristics as large areas (groups), it suffices to borrow strength cross-sectionally (i.e. from larger areas having similar region for the small areas). The idea here is that "the groups (strata) are a powerful factor in explaining the variance of the variables whereas the domains are not". For example, as illustrated in Section 4.1 (Real-life Data Based Evaluation), 'sex' as used in the groupings will often explain a good part of individual variations but beyond that the States (Domains) may be a weak explanatory factor, see Sarndal, Swensson and Wretman 1992. In addition, the idea of calibration allows us to borrow strength from auxiliary variable, hence, a calibrated synthetic estimator of the population mean $\bar{Y}_d$, is obtained as follows:

$$\hat{\bar{y}}_{dc}^{s*} = \sum_g^G W_{dg}^c\bar{y}_{.g} \quad (6)$$

where W_{dg}^c is the new calibration weight chosen such that the distance measure given by:

$$\Phi_1(W^c, W) = \frac{1}{2}\sum_g^G \frac{\left(W_{dg}^c - W_{dg}\right)^2}{W_{dg}q_{dg}} \quad (7)$$

is minimized subject to the calibration constraints;

$$\sum_g^G W_{dg}^c \bar{x}_{.g} = \bar{X}_d \tag{8}$$

and q_{dg} are known positive weights unrelated to W_{dg}^c called the tuning parameter. Minimizing the loss function (7) subject to the calibration constraint (8) yields the calibration weights for small area under stratified sampling as

$$W_{dg}^c = W_{dg} + \frac{(\bar{X}_d - \sum_g^G W_{dg}\bar{x}_{.g})}{\sum_g^G W_{dg} q_{dg} \bar{x}_{.g}^2} W_{dg} q_{dg} \bar{x}_{.g} \tag{9}$$

Substituting (9) into (6) gives

$$\hat{\bar{y}}_{dc}^{s*} = \sum_g^G W_{dg}\bar{y}_{.g} + \frac{\sum_g^G W_{dg} q_{dg} \bar{x}_{.g}\bar{y}_{.g}}{\sum_g^G W_{dg} q_{dg} \bar{x}_{.g}^2}\left(\bar{X}_d - \sum_g^G W_{dg}\bar{x}_{.g}\right). \tag{10}$$

Equation (10) can also be written in the form of GREG estimator:

$$\hat{\bar{y}}_{dc}^{s*} = \hat{\bar{y}}_d^s + (\bar{X}_d - \hat{\bar{x}}_d^s)\hat{B}_d^s \tag{11}$$

where $\hat{\bar{y}}_d^s$ is as defined in Eq.2 and $\hat{\bar{x}}_d^s$ analogously defined as Eq.2 for the auxiliary variable, and

$$\hat{B}_d^s = \left(\sum_g^G W_{dg} q_{dg} \bar{x}_{.g}^2\right)^{-1} \sum_g^G W_{dg} q_{dg} \bar{x}_{.g}\bar{y}_{.g}$$

3.1. Estimator of the variance of $\hat{\bar{y}}_{dc}^{s*}$

*Lemma***:** The variance of the estimator in Eq.10 with one constraint is given as

$$\hat{V}_H(\hat{\bar{y}}_{dc}^{s*}) = \sum_g^G \frac{D_{.g}(W_{dg}^c)^2}{W_{dg}^2} s_{\hat{e}.g}^2 + \frac{\sum_g^G D_{.g}(W_{dg}^c)^2 Q_{dg} s_{.gx}^2}{\sum_g^G D_{.g} W_{dg}^2 Q_{dg} (s_{.gx}^2)^2} s_{\hat{e}.g}^2 \left[V_{st}(\bar{x}_d^s) - \hat{V}_{st}(\bar{x}_d^s)\right].$$

Proof: Consider the estimator of variance of the combined regression estimator under stratified sampling by Sarndal (1996) given as

$$\hat{V}_c(\hat{\bar{y}}_d) = \sum_g^G \frac{D_{.g}(W_{dg}^c)^2}{W_{dg}^2} s_{\hat{e}.g}^2 \tag{12}$$

where $s_{\hat{e}.g}^2 = \left(n_{.g} - 1\right)^{-1} \sum_k^{n_{dg}} \hat{e}_{dgk}^2$ is the g^{th} group (stratum) sample variance, $\hat{e}_{dgk} = y_{dgk} - \bar{y}_{.g} - \hat{B}_d^s\left(x_{dgk} - \bar{x}_{.g}\right)$, W_{dg}^c as given in Eq.9 and $D_{.g} = W_{dg}^2 \gamma_{.g}$ is the initial weight of Eq.12 and $\gamma_{.g} = \frac{1}{n_{.g}} - \frac{1}{N_{.g}}$. Following the procedure by Singh and Arnab (2014), the estimate of variance of the estimator $\hat{\bar{y}}_{dc}^{s*}$ obtained by calibrating on Eq.12 is given as

$$\hat{V}_H(\hat{\bar{y}}_{dc}^{s*}) = \sum_g^G \frac{\Omega_{.g}(W_{dg}^c)^2}{W_{dg}^2} s_{\hat{e}.g}^2 \tag{13}$$

where $\Omega_{.g}$ is the new weights chosen such that the chi-square distance function

$$\Phi_2 = \sum_g^G \frac{(\Omega_{.g} - D_{.g})^2}{D_{.g} Q_{dg}} \tag{14}$$

is minimized subject to the calibration constraint

$$\sum_g^G \Omega_{.g} s_{.gx}^2 = V_{st}(\bar{x}_d^s) \tag{15}$$

where Q_{dg} is the tuning parameter unrelated with $\Omega_{.g}$. Hence, the calibration weight is obtained as

$$\Omega_{.g} = D_{.g} + \frac{D_{.g} Q_{dg} s_{.gx}^2}{\sum_g^G D_{.g} Q_{dg} (s_{.gx}^2)^2} \left[V_{st}(\bar{x}_d^s) - \sum_g^G D_{.g} s_{.gx}^2 \right]. \tag{16}$$

Substituting (16) in (13) gives

$$\hat{V}_H(\hat{\bar{y}}_{dc}^{s*}) = \sum_g^G \frac{D_{.g} (W_{dg}^c)^2}{W_{dg}^2} s_{\hat{e}.g}^2 + \frac{\sum_g^G D_{.g} (W_{dg}^c)^2 Q_{dg} s_{.gx}^2}{\sum_g^G D_{.g} W_{dg}^2 Q_{dg} (s_{.gx}^2)^2} s_{\hat{e}.g}^2 \left[V_{st}(\bar{x}_d^s) - \hat{V}_{st}(\bar{x}_d^s) \right] \tag{17}$$

where $V_{st}(\bar{x}_d^s) = \sum_g^G D_{.g} S_{.gx}^2$ is assumed to be known variance of $\bar{X}_d$ and $s_{.gx}^2 = \frac{1}{n_{.g}-1} \sum_d^D \sum_k^{n_{dg}} \left(x_{dgk} - \bar{x}_{.g} \right)^2$ is an unbiased of $S_{.gx}^2 = \frac{1}{N_{.g}-1} \sum_d^D \sum_k^{n_{dg}} \left(X_{dgk} - \bar{X}_{.g} \right)^2$.

Eq (17) can further be written as

$$\hat{V}_H(\hat{\bar{y}}_{dc}^{s*}) = \hat{V}_c(\hat{\bar{y}}_d) + \hat{B}_{dc} \left[V_{st}(\bar{x}_d^s) - \hat{V}_{st}(\bar{x}_d^s) \right] \tag{18}$$

where $\hat{V}_c(\hat{\bar{y}}_d) = \sum_g^G \frac{D_{.g} (W_{dg}^c)^2}{W_{dg}^2} s_{\hat{e}.g}^2$ as earlier defined in Eq.12 and $\hat{B}_{dc} = \frac{\sum_g^G D_{.g} (W_{dg}^c)^2 Q_{dg} s_{.gx}^2}{\sum_g^G D_{.g} W_{dg}^2 Q_{dg} (s_{.gx}^2)^2} S_{\hat{e}.g}^2$.

3.2. Combined ratio synthetic estimator

Here, we consider special cases of the estimator in Eq.10.

Case 1: Suppose we set the tuning parameter $q_{dg} = \bar{x}_{.g}^{-1}$ in Eq.10, then

$$\hat{\bar{y}}_{dcr}^{s*} = \frac{\sum_g^G W_{dg} \bar{y}_{.g}}{\sum_g^G W_{dg} \bar{x}_{.g}} \bar{X}_d. \tag{19}$$

The approximate form of the bias of Eq.19 is obtained through Taylor's series approximation method. Equation (19) can be written as

$$\hat{\bar{y}}_{dcr}^{s*} = \frac{\hat{\bar{y}}_d^s}{\hat{\bar{x}}_d^s} \bar{X}_d = \hat{R}_d^s \bar{X}_d \tag{20}$$

$$\hat{\bar{y}}_{dcr}^{s*} - \bar{Y}_d = \bar{X}_d \left(\hat{R}_d^s - R_d \right), \text{ where } R_d = \frac{\bar{Y}_d}{\bar{X}_d}.$$

The bias $B(\hat{\bar{y}}_{dcr}^{s*}) = E(\hat{\bar{y}}_{dcr}^{s*} - \bar{Y}_d) = E\left[\bar{X}_d(\hat{R}_d^s - R_d)\right]$

$$B(\hat{\bar{y}}_{dcr}^{s*}) = \bar{X}_d E\left[\frac{1}{\bar{x}_d^s}(\bar{y}_d^s - R_d\bar{x}_d^s)\right]. \tag{21}$$

But $\frac{1}{\bar{x}_d^s} = \frac{1}{\bar{X}_d}\left[1 + \frac{\bar{x}_d^s - \bar{X}_d}{\bar{X}_d}\right]^{-1}$, such that Taylor's series expansion to the first order approximation gives $\frac{1}{\bar{x}_d^s} = \frac{1}{\bar{X}_d}\left[1 - \frac{\bar{x}_d^s - \bar{X}_d}{\bar{X}_d}\right]$, then Equation (21) becomes $B(\hat{\bar{y}}_{dcr}^{s*}) = \bar{X}_d E\left[\frac{1}{\bar{X}_d}\left(1 - \frac{\bar{x}_d^s - \bar{X}_d}{\bar{X}_d}\right)(\bar{y}_d^s - R_d\bar{x}_d^s)\right]$

$$B(\hat{\bar{y}}_{dcr}^{s*}) = \frac{1}{\bar{X}_d}\sum_g^G W_{dg}^2\gamma_{.g}\left(R_d S_{.gx}^2 - S_{.gxy}\right) \tag{22}$$

where, $S_{.gxy} = \frac{1}{N_{.g}-1}\sum_d^D\sum_k^{N_{dg}}\left(X_{dgk} - \bar{X}_{.g}\right)\left(Y_{dgk} - \bar{Y}_{.g}\right)$ is estimated by

$$s_{.gxy} = \frac{1}{n_{.g}-1}\sum_d^D\sum_k^{n_{dg}}\left(x_{dgk} - \bar{x}_{.g}\right)\left(y_{dgk} - \bar{y}_{.g}\right)$$

and the bias estimator of $\hat{\bar{y}}_{dcr}^{s*}$ is given as;

$$\hat{B}(\hat{\bar{y}}_{dcr}^{s}) = \frac{1}{\bar{x}_d}\sum_g^G W_{dg}^2\gamma_{.g}\left(\hat{R}_d s_{.gx}^2 - s_{.gxy}\right). \tag{23}$$

To obtain the estimator of the variance for Eq.19, we set $q_{dg} = \bar{x}_{.g}^{-1}$, $Q_{dg} = s_{.gx}^{-2}$ and replaced $s_{\hat{e}.g}^2$ by $s_{\hat{e}.gr}^2$ in Eq.17 as;

$$\hat{V}_{st}(\hat{\bar{y}}_{dcr}^{s*}) = \left(\frac{\bar{X}_d}{\bar{x}_d^s}\right)^2\left[\frac{V_{st}(\bar{x}_d^s)}{\hat{V}_{st}(\bar{x}_d^s)}\right]\sum_g^G W_{dg}^2\gamma_{.g}s_{\hat{e}.gr}^2 \tag{24}$$

Equation (24) is in the form of the ratio-type estimator proposed by Wu & Deng (1983) for estimating variance of the combined ratio estimator. The difference here is that it makes use of extra knowledge of known variance of the auxiliary variable at the estimation stage, where

$$s_{\hat{e}.gr}^2 = s_{.gy}^2 + \hat{R}_d^2 s_{.gx}^2 - 2\hat{R}_d s_{.gxy}$$

3.3. Combined regression synthetic estimator

Case 2: Again, we set the tuning parameter $q_{dg} = 1$ in Eq.10, then the combined regression-synthetic estimator in stratified sampling is given as

$$\hat{\bar{y}}_{dcREG}^{s*} = \bar{y}_d^s + \hat{b}_{.g}^*(\bar{X}_d - \bar{x}_d^s) \tag{25}$$

where $\hat{b}_{.g}^* = \frac{\sum_g^G W_{dg}\bar{x}_{.g}\bar{y}_{.g}}{\sum_g^G W_{dg}\bar{x}_{.g}^2}$ is the synthetic regression coefficient of the domain. The estimator in Eq.25 is in the form of Hansen, Hurwitz and Madow (1953) combined regression estimator. The bias of Eq.25 is obtained by replacing R_d by $\hat{b}_{.g}^*$ in Eq.22 such that

$$B(\hat{\bar{y}}_{dcREG}^{s*}) = \frac{1}{\bar{X}_d}\sum_g^G W_{dg}^2\gamma_{.g}\left(\hat{b}_{.g}^* S_{.gx}^2 - S_{.gxy}\right) \tag{26}$$

and its estimator is

$$\hat{B}(\hat{\bar{y}}_{dcREG}^{S*}) = \frac{1}{\bar{X}_d}\sum_g^G W_{dg}^2 \gamma_{.g}\left(\hat{b}_{.g}^{*} s_{.gx}^2 - s_{.gxy}\right). \quad (27)$$

An estimator of variance of the calibration approach combined regression synthetic estimator $\hat{\bar{y}}_{dcREG}^{S*}$ is obtained by setting $q_{dg} = \bar{x}_{.g}^{-1}$ and $Q_{dg} = s_{.gx}^{-2}$ in Eq.17 and replacing $s_{\hat{e}.g}^2$ with $s_{\hat{e}.greg}^2$ as

$$\hat{V}_{st}\left(\hat{\bar{y}}_{dcreg}^{S}\right) = \left(\frac{\bar{X}_d}{\bar{x}_d^s}\right)^2 \left[\frac{V_{st}(\bar{x}_d^s)}{\bar{V}_{st}(\bar{x}_d^s)}\right] \sum_g^G W_{dg}^2 \gamma_{.g} s_{\hat{e}.greg}^2 \quad (28)$$

where

$$s_{\hat{e}.greg}^2 = s_{.gy}^2 + \hat{b}_d^2 s_{.gx}^2 - 2\hat{b}_d s_{.gxy} \quad (29)$$

4. Data and methods for empirical evaluation

In this section, data and methods for empirical evaluation of the proposed and existing estimators are discussed. Real-life and simulated data are used. When domains of interest have no sample observation, the existing calibration estimator becomes impossible to compute for the area of interest. This was illustrated using real-life data as shown in Table 1. This will help to validate the theoretical claims. However, on a general note, simulation analysis was done using R-software to ascertain the level of performance of both existing and suggested estimators.

4.1. Real-life data based evaluation

The real-life data used in this analysis were obtained from the population of the household finances and consumptions survey (HFCS) conducted in 2017 by the Statistics Department of the Central Bank of Nigeria. The population comprised of 2986 male and female heads of household. The population was partitioned into two strata of male and female household heads with subpopulation sizes of 1625 and 1361, respectively, across the 37 domains (States).

To illustrate an ideal situation of small area estimation, the study variable y is considered as the household expenditure and the auxiliary variable x as the household income. The object is to estimate the mean of y for all the 37 domains. To compute the estimates for all the domains, the proportional allocation procedure was applied and a sample s of size 10% was drawn from each stratum (group) using simple random sample without replacement (SRSWOR) with the cells averages on bold points as shown in Appendix A. The results are shown in Table 1.

Table 1. Estimators of mean expenditures for domains using existing and proposed estimators

DOMAIN	$\hat{\bar{y}}_{dcr}$	$\hat{\bar{y}}_{dcreg}$	$\hat{\bar{y}}_{d}^{s}$	$\hat{\bar{y}}_{dcr}^{s*}$	$\hat{\bar{y}}_{dcREG}^{s*}$	$\bar{Y}_d$
ABIA	21393.61	21392.2	24061.07	23131.81	23143.91	21366.1
ADAMAWA	21861.68	21856.98	25267.49	36151.67	36035.95	34303.16
AKWA IBOM	19173.4	19176.94	24444.22	34921.18	34791.54	26161.88
ANAMBRA	18401.87	18392.88	24112.91	25195.98	25181.96	23548.92
BAUCHI	12556.76	12557.61	25003.59	15856.92	15959.68	15773.49
BAYELSA	6634.7	6634.7	24343.82	29799.31	29730.82	26115.66
BENUE	31350.08	31345.96	24298.10	37359.16	37194.12	30794.01
BORNO	22672.94	22753.52	25484.83	21287.21	21329.65	20671.39
CROSS R	12685.94	12682.54	23830.67	17343.67	17430.28	17085.43
DELTA	44291.46	44272.31	24021.61	46528.64	46234.19	41404.02
EBONYI	20677.67	20677.67	24343.82	23069.98	23085.97	22713.72
EDO	25419.3	25412.72	24063.92	34683.85	34545.6	30517.07
EKITI	38276.63	38289.31	24821.58	30122.35	30060.72	30346.33
ENUGU	27117.48	27132.43	23628.31	22174.25	22194.02	25256.79
FCT	7508.05	7508.05	25636.96	26374.9	26367.72	24184.29
GOMBE	33280.32	33364.8	24763.67	27232.24	27203.23	29613.17
IMO	39268.29	39190.17	23934.60	30677.41	30588.34	31019.32
JIGAWA	21597.66	21597.22	24824.90	25092.69	25089.58	24196.47
KADUNA	38844.85	38844.88	25463.42	28631.39	28599.19	27648.58
KANO	19557.45	19636.3	25313.68	27801.67	27775.49	30336.33
KATSINA	21793.15	21797.54	24647.22	27130.71	27100.95	28684.67
KEBBI	35688.25	35696.3	24393.48	35026.58	34894.02	33024.15
KOGI	22144.18	22147.94	25059.34	18667.27	18738.29	20948.49
KWARA	27003.13	27003.29	24763.67	31499.15	31420.01	34756.05
LAGOS	31918.62	31948.07	25136.32	43977.99	43771.93	52055.29
NASARA	56889.76	56929.3	23574.10	25923.61	25891.53	28630.6
NIGER	24252.95	24241.96	25095.65	35049.99	34940.2	33504.88
OGUN	47334.04	47321.89	24544.62	35066.09	34937.89	34929.4
ONDO	NA	NA	25239.92	**33167.93**	**33083.13**	**33375.22**
OSUN	14038.37	14038.32	24005.90	14977.6	15095.92	16472.88
OYO	26364.26	26364.43	24120.36	29933.23	29858.07	29499.42
PLATEAU	25735.76	25735.58	24456.47	24168.14	24171.7	24785.42
RIVERS	27626.89	27633.98	25313.11	33789.51	33700.3	35138.01
SOKOTO	10216.44	10216.44	27422.72	16539.78	16591.33	18813.12
TARABA	9306.65	9306.65	25036.58	43455.52	43249.94	46880.2
YOBE	14446.75	14446.75	27251.67	26217.78	26223.21	27139.53
ZAMFARA	20969.1	20968.26	24586.90	19218.79	19283.76	21479.42
AVERAGE	24952.73	24284.21	24765.17	**28574.22**	**28526.87**	**28464.13**

4.2. Simulation Study

Here, the procedure of population generation and sample selection by Hidiroglou and Estevao (2014) was adopted. Bivariate observations (x_{ij}, y_{ij}) were generated to comprise finite population of size 4950 units. The population U considered was created by generating data for three separate subsets of the populations termed *groups* (strata) with different intercepts and slopes. Each group was split into ten domains that are mutually exclusive and exhaustive, as follows: *Group 1;* $U_{11}, U_{21}, \ldots, U_{101}$, *Group 2;* $U_{12}, U_{22}, \ldots, U_{102}$, and *Group 3;* $U_{13}, U_{23} \ldots, U_{103}$. The number of units in each cell N_{dg} was sequentially allocated in a monotonic manner: cell U_{11} with 20 units; cell U_{21} with 30 units; and cell U_{103} with 310 units. The values of x in each group were generated from three different distributions, *Gamma* $(\alpha = 5, \beta = 10)$, *Norm* (5,1) and *Exp* (1.5) distributions. The simulation for the variable of interest y was obtained using the model $y_{dk} = \beta_{og} + \beta_{1g} x_{dk} + v_d + e_{dk}$, where $d = 1,2,\ldots,30$; $k = 1,2,\ldots,N$ and $g = 1,2,3$; $e_{dk} \sim N(0, C_{dk}^2 \sigma_e^2)$, $v_d \sim N(0, \sigma_v^2)$. It is assumed that $\sigma_e^2 = \sigma_v^2 = 20^2 = 400$ for the gamma distribution, $\sigma_e^2 = \sigma_v^2 = 1^2 = 1$ for normal and exponential distributions. $c_{dk} = x_{dk}$ is set to reflect the heterogeneity of the model errors for the synthetic and calibration estimators.

4.2.1. Simulation results

The summary of the representation of units in each group across the domains is presented in Table 2 and 3. Table 2 shows how the population was split into the three groups with the respective values of intercepts and slopes for the Gamma, Normal and Exponential distributions. Table 3 illustrates the population under study divided into domains and further partitioned into groups that are larger than the domains and cut across the domains to form grids that are mutually exclusive and exhaustive. The result of the simulation study using R software for selection of independent samples of sizes $n = 248(5\%)$, $n = 495(10\%)$, $n = 744(15\%)$, $n = 990(20\%)$, $n = 1239(25\%)$ drawn using SRSWOR from U and the computation of various estimates is presented in Table 4.

Summary statistics of the simulated data will be done using Average Percent Absolute Relative Bias, Average Percent Relative Efficiency and Average Percent Coefficient of Variation $\%\overline{ARB}$, $\%\overline{RE}$ and $\%\overline{CV}$ respectively, and are obtained as

$$\%\overline{ARB}(\hat{\bar{y}}_{dP}) = \left[\frac{1}{D}\sum_{d=1}^{D} ARB(\hat{\bar{y}}_{dP})\right] \times 100,$$

where $ARB(\hat{\bar{y}}_{dP}) = \left|\frac{1}{R}\sum_{r=1}^{R}\left(\frac{\hat{\bar{y}}_{dP}^{(r)}}{\bar{Y}_d} - 1\right)\right|$

$$\%\overline{RE}(\hat{\bar{y}}_{dP}) = \left[\frac{\overline{MSE}(\hat{\bar{y}}_{dE})}{\overline{MSE}(\hat{\bar{y}}_{dP})}\right] \times 100,$$

where $\overline{MSE}(\hat{\bar{y}}_{dP}) = \frac{1}{D}\sum_{d=1}^{D} MSE(\hat{\bar{y}}_{dP})$ and

$$MSE(\hat{\bar{y}}_{dP}) = \frac{1}{R}\sum_{r=1}^{R}\left(\hat{\bar{y}}_{dp}^{(r)} - \bar{Y}_d\right)^2$$

$$\%\overline{CV}(\hat{\bar{y}}_{dP}) = \left[\frac{1}{D}\sum_{d=1}^{D} CV(\hat{\bar{y}}_{dP})\right] \times 100, \text{ where } CV(\hat{\bar{y}}_{dP}) = \frac{\sqrt{MSE(\hat{\bar{y}}_{dP})}}{\bar{Y}_d}$$

where $\hat{\bar{y}}_{dP}^{(r)}$ and $\hat{\bar{y}}_{dE}^{(r)}$ denote, say, the proposed and existing estimators respectively, produced for the r^{th} sample, $r = 1,2,\ldots,R$, and for each small area $d = 1,2,\ldots,D$. For each selected sample in each simulation run $= 1,2,\ldots,R$ $(R = 100{,}000)$, we shall compute estimates of $\overline{Y}_d$ for the estimators.

Note: In small area estimation, Molina and Rao (2010) suggested a benchmark value for $\%\overline{CV}(\hat{\bar{y}}_{dP})$ at 20-25% as being reliable. As a result, a high value of $\%\overline{CV}(\hat{\bar{y}}_{dP})$ above 25% is considered as unreliable estimates while estimators with values of $\%\overline{CV}(\hat{\bar{y}}_{dP})$ below 25% are considered reliable and suitable for SAE.

Table 2. Partitioning the Population into Groups with their Respective Slopes and Intercepts under Gamma, Normal and Exponential Distributions

Distributions		Gamma		Normal and Exponential	
Group (g)	Cells in groups	β_{0g}	β_{1g}	β_{0g}	β_{1g}
1	U_{d1} for $k = 1,2,\ldots,10$	200	30	5	1.5
2	U_{d2} $for\ k = 11,\ldots,20$	300	20	10	1.0
3	U_{d3} $for\ k = 22,\ldots,30$	400	10	15	0.5

Table 3. Summary of Splitting the Population into Cells, Groups and Domains

Domain number (d)	Group Number (g)			Domains (U_d)
	1	**2**	**3**	
1	U_{11}	U_{12}	U_{13}	U_1
2	U_{21}	U_{22}	U_{23}	U_2
3	U_{31}	U_{32}	U_{33}	U_3
4	U_{41}	U_{42}	U_{43}	U_4
5	U_{51}	U_{52}	U_{53}	U_5
6	U_{61}	U_{62}	U_{63}	U_6
7	U_{71}	U_{72}	U_{73}	U_7
8	U_{81}	U_{82}	U_{83}	U_8
9	U_{91}	U_{92}	U_{93}	U_9
10	U_{101}	U_{102}	U_{103}	U_{10}
Groups $(\boldsymbol{U_{.g}})$	$U_{.1}$	$U_{.2}$	$U_{.3}$	U

Table 4. Result of Simulation Evaluation for Gamma (5,10), Norm (5,1) and Exp (1.5)

Sample size	Distribution	$\%\overline{ARB}$			$\%\overline{RE}$			$\%\overline{CV}$		
		$\hat{\bar{y}}_d^s$	$\hat{\bar{y}}_{dcr}^{s*}$	$\hat{\bar{y}}_{dcREG}^{s*}$	$\hat{\bar{y}}_d^s$	$\hat{\bar{y}}_{dcr}^{s*}$	$\hat{\bar{y}}_{dcREG}^{s*}$	$\hat{\bar{y}}_d^s$	$\hat{\bar{y}}_{dcr}^{s*}$	$\hat{\bar{y}}_{dcREG}^{s*}$
5%	Gamma	65.7	13.4	13.7	100	2350	2360	65.7	13.4	13.7
	Normal	73.2	5.5	5.5	100	12832.7	12845.9	73.2	5.5	5.5
	Exponential	74.2	14.2	**61.7**	100	2451.9	**145.5**	74.2	15.3	**61.8**
10%	Gamma	65.7	13.3	13.4	100	4150	4110	65.7	13.3	13.4
	Normal	71.9	5.5	5.5	100	12786.6	12846.4	71.9	5.5	5.5
	Exponential	74.2	13.6	**61.7**	100	2383.3	**145.4**	74.2	14.6	**61.8**
15%	Gamma	65.7	13.3	13.4	100	4590	4520	65.7	13.3	13.4
	Normal	71.9	5.5	5.5	100	12770.3	12844.3	71.9	5.5	5.5
	Exponential	74.2	13.7	**61.8**	100	2360.8	**145.3**	74.2	14.4	**61.8**
20%	Gamma	65.7	13.2	13.7	100	4840	4740	65.7	13.2	13.7
	Normal	71.9	5.5	5.5	100	12776.9	12859.7	71.9	5.5	5.5
	Exponential	74.2	13.7	**61.8**	100	2358.7	**145.3**	74.2	14.8	**61.8**
25%	Gamma	65.7	13.2	13.8	100	4750	4750	65.7	13.2	13.8
	Normal	71.9	5.5	5.5	100	12767.3	12853.1	71.9	5.5	5.5
	Exponential	74.2	13.7	**61.8**	100	2353.9	**145.3**	74.2	14.1	**61.8**

5. Discussion of results

From Table 1, it can be seen that the domain called ONDO has no estimate under the existing calibration estimators $\hat{\bar{y}}_{dcr}$ and $\hat{\bar{y}}_{dcreg}$ because there was no sampled unit selected for that domain and the estimates of the population mean could not be computed. This agrees with Purcell and Kish (1979) and Rao (2003), that in areas without sample observations, the direct estimator could not be computed. However, the synthetic estimators (both existing and proposed) $\hat{\bar{y}}_d^S$, $\hat{\bar{y}}_{dcr}^{s*}$ and $\hat{\bar{y}}_{dcREG}^{S*}$ produced estimates (of 25239.92, **33167.93 and 33083.13 naira** respectively) for the average population expenditure (of **33375.22 naira)** for ONDO. This agrees with Rao (2003) proposition on testability of the assumption of structural similarities of characteristics and a careful choice of auxiliary variable in the use of synthetic estimators. It could also be seen that, although the existing synthetic estimator $\hat{\bar{y}}_d^S$ produced estimate for ONDO where there are no sample units, the value obtained underestimated the population mean expenditure of the domain compared to the proposed synthetic estimators that made use of additional supplementary information. Furthermore, on average, the mean expenditures of all the domains produced by the proposed synthetic estimators $\hat{\bar{y}}_{dcr}^{s*}$

and $\hat{\bar{y}}_{dcREG}^{S*}$ as **28574.22** and **28526.87 naira** respectively, are almost the same as the population mean expenditure (**28464.13 naira)** compared to that obtained from the existing synthetic and calibrated estimators. This agrees with the appealing property of the domain estimator (that the sum of the domains estimates will equal the population parameter) as suggested in the literature by Lundstrom and Sarndal (2001). These results confirm the need to borrow strength cross-sectionally in addition to a highly correlated auxiliary variable. In this case, the proposed synthetic estimator gained dominance over the existing synthetic and direct estimators in domains of interest where there are no sample observations.

Results of analysis in Table 4 showed values of $\%\overline{ARB}$ between 5.5% to 14.2% and 5.5% to 61.8% for $\hat{\bar{y}}_{dcr}^{S*}$ and $\hat{\bar{y}}_{dcREG}^{S*}$ respectively, while that of the existing synthetic estimator $\hat{\bar{y}}_{d}^{S}$ was 65.7% to 71.9%. From the result, the proposed synthetic estimators have been found to exhibit a remarkably smaller $\%\overline{ARB}$ than the existing synthetic estimator for all the probability distributions and in all sample sizes under study.

It was further observed that under normal distribution, the proposed estimators $\hat{\bar{y}}_{dcr}^{S*}$ and $\hat{\bar{y}}_{dcREG}^{S*}$ have a constant $\%\overline{ARB}$ of 5.5%, which is regarded as the least for all sample sizes. Under gamma distribution, they recorded between 13.2% to 13.8%. However, under exponential distribution, the $\%\overline{ARB}$ values of $\hat{\bar{y}}_{dcr}^{S*}$ lies between 13.6% to 14.4% while that of $\hat{\bar{y}}_{dcREG}^{S*}$ was seen to be highly biased with 61.7% to 61.8% as indicated in column 5 of Table 4 with bold points. This result conforms to an established fact in the literature by Clement and Enang (2017) that under domain estimation, the calibration approach combined ratio estimator outperforms the combined regression estimator. In addition, this result suggests that for real life data that follow exponential distribution, the proposed combined ratio is more preferred to the combined regression estimator.

From Table 4, the proposed synthetic estimators $\hat{\bar{y}}_{dcr}^{S*}$ and $\hat{\bar{y}}_{dcREG}^{S*}$ were observed to have higher gains in efficiency than the existing estimator $\hat{\bar{y}}_{d}^{S}$ in all sample sizes for all the three probability distributions considered. Contrary to popular claims that the existing synthetic estimator produced estimates for domains without sample units with a very small mean square error, the proposed synthetic estimators have been shown to be more efficient and superior to the existing estimator. The $\%\overline{RE}$ of the proposed synthetic estimators for the three distributions were observed to be between 12770.3% to 12894.3% for normal distribution followed by gamma distribution with 2350% to 4840% and the exponential distribution between 145.5% to 2451.9%. Again, as expected, $\hat{\bar{y}}_{dcr}^{S*}$ was clearly more efficient than $\hat{\bar{y}}_{dcREG}^{S*}$ in all sample sizes under exponential distribution, which supports the results in the literature by Clement and Enang (2017). Furthermore, the gain in efficiency of the proposed $\hat{\bar{y}}_{dcr}^{S*}$ for the three distributions and in all sample sizes considered is an improvement in SAE contrary to

the results in the literature by Rao and Choudhry (1996) on the instability of ratio synthetic MSE.

Again, the result in Table 4 showed that $\hat{\bar{y}}_{dcr}^{S*}$ has the smallest percent average coefficient of variation of 5.5% to 15.3% while $\hat{\bar{y}}_{dcREG}^{S*}$ has 5.5% to 61.8% against $\hat{\bar{y}}_{d}^{S}$ with $\%\overline{CV}$ of 65.7% to 71.9% for all the sample sizes considered in this work. This suggest that the proposed synthetic estimators are highly preferred for small area estimation having met the required benchmark of falling below 25% as suggested by Molina and Rao (2010) but the same could not be said about the existing synthetic estimator. It could be said that synthesizing on the approaches of borrowing strength cross-sectionally and through calibration has been profitable in this study. It was further observed that, for exponential distribution, the combined regression synthetic estimator $\hat{\bar{y}}_{dcREG}^{S*}$ has a constant $\%\overline{CV}$ as high as 61.8% for all sample sizes. This is an indication that $\hat{\bar{y}}_{dcREG}^{S*}$ is not suitable for any real-life data that follow exponential distribution for small domains under stratified sampling. It is convenient to say that the proposed combined ratio has an edge over the combined regression synthetic estimator under exponential distribution in small area estimation. The normal distribution produced a constant value of $\%\overline{CV}$ of 5.5% and is seen as the smallest for all the sample sizes followed by Gamma with 13.2% to 13.8% and exponential with 14.4% to 61.8% for the proposed synthetic estimators. This points to the desired qualities of the Normal distribution in small area estimation under stratified sampling design.

6. Conclusion

The synthetic estimation technique has been shown to be the only remedy if no sampled units are available in some domains of interest as shown by the result of this study. Therefore, it can be concluded that the proposed small area estimators (which borrowed strength cross-sectionally and with auxiliary variable) are an improvement over the Marker (1999) synthetic estimator (that only borrowed strength cross-sectionally) and the calibration approach direct estimators for the estimation of population mean in areas that are characterized by small/no sample sizes. Also, the proposed combined ratio synthetic estimator has shown dominance over the combined regression synthetic estimator suggesting that the latter is not suitable for any real-life data that follow exponential distribution for small domains under stratified sampling.

Acknowledgement

The author is grateful to the anonymous Reviewers who painstakingly read this manuscript and made useful contributions to see the paper through this stage.

References

Battese, G. E., Fuller, W. A., (1984). An Error Components Model for Predictions of County Crop Areas using Survey and Satellite Data. Survey section, Statistics Laboratory Iowa State University, Ames.

Chambers, R. L., (2006). Small Area Estimation for Business Surveys. Proceedings of the American Statistical Association, Section on Survey Research Methods, pp. 2803–2809.

Clement, E. P., Enang, E. I., (2017). On the Efficiency of Ratio Estimators over Regression Estimators, *Communication in Statistics - Theory and Methods*, Vol. 46, pp. 5357–5367.

Deville, J. C., Sarndal, C. E., (1992). Calibration Estimators in Survey Sampling, *Journal of the American Statistical Association*, Vol. 87, pp. 376–382.

Drew, J. D., Singh, M. P., Choudhry, G. H., (1982). Evaluation of Small Area Techniques for the Canadian Labor Force Survey, *Survey Methodology*, Vol. 8, pp. 17–47.

Fay, R. E., Herriot, R. A., (1979). Estimates of Income for Small Places. An Application of James-Stein Procedures to Census Data, *Journal of the American Statistical Association*, Vol. 74, pp. 269–277.

Gonzales, M. E., (1973). Use and Evaluation of Synthetic Estimator, Proceedings of the Section on Social statistics, American statistical Association, pp. 33–36.

Hansen, M. H., Hurwitz, W. N., Madow, W. G., (1953). Sample Survey Methods and Theory, John Wiley and Sons.

Hidiroglou, M. A., Estevao, V. M., (2014). A Comparison of Small Area and Calibration Estimators Via Simulation, *Statistics in Transition*, New Series, Vol. 17, pp. 133–154.

Holt, D., Smith, T. M. F., Tomberlin, T. J., (1979). A Model Based Approach to Estimation for Small subgroups of a Population, *Journal of the American Statistical Association*, Vol. 74, pp. 405–410.

Horvitz, D. G., Thompson, D. J., (1952). A Generalization of Sampling without Replacement from a Finite Universe, *Journal of the American Statistical Association,* Vol. 47, pp. 663–687.

Lundstrom, S., Sarndal, C. E., (2001). Estimation in the presence of Nonresponse and Frame Imperfections, Statistics Sweden.

Marker, D. S., (1999). Organization of Small Area Estimation. Using Generalized Linear Regression Framework, *Journal of Official Statistics*, Vol. 15, pp. 1–24.

Molina, I., Rao, J. N. K., (2010). Small Area Estimation Poverty Indicators, *Canadian Journal of statistics,* Vol. 38, pp. 369–385.

Pfeffermann, D., (2013). New Important Developments in Small Area Estimation, *Statistical Sciences*, Vol. 28, pp. 40–68.

Purcell, N. J., Kish, L., (1979). Estimation for Small Domains, *Biometrics*, Vol. 35, pp. 365–354.

Rao, J. N. K., (2003). Small Area Estimation, Wiley, New York.

Rao, J. N. K., Molina, I. (2015). Small Area, Second Edition. John Wiley, New York.

Sarndal, C. E., Swensson. B., Wretman, J., (1992). Model-Assisted Surveys, New York: Springer-Verlag.

Sarndal, C. E., (1981). When Robust Estimation is not an Obvious Answer: The case of the Synthetic Estimator versus Alternatives for Small Areas, Proceedings of the American Statistical Association, Survey Research Section, pp. 53–59.

Sarndal, C. E., (1984). Design-Consistent Versus Model Dependent Estimation for Domains, Journal *of the American Statistical Association*, Vol. 79, pp. 624–637.

Sarndal, C. E., (1996). Efficient Estimators with Simple Variance in Unequal Probability Sampling, *Journal of the American Statistical Association*, Vol. 91, pp. 1289–1300.

Sarndal, C. E. (2007). The Calibration Approach in Survey Theory and Practice. Survey Methodology, Vol. 33, pp. 99–119.

Singh, S., Arnab, R., (2014). On Calibration of design weights, *International Journal of Statistics*. Vol. 69, pp. 185–205.

Tikkiwal, G. C., Pandey, K. K., (2007). On Synthetic and Composite Estimators for Small Area Estimation under Lahiri-Midzuno Sampling Scheme, *Statistics in Transition-new Series, Poland*, Vol. 8, pp. 111–123.

Wu, C. F., Deng, L. Y., (1983). Estimation of Variance of the Ratio Estimator; An Empirical Study, In: G. E. P. Box et al. (ed.), Scientific Inference, Data Analysis and Robustness. New York, Academic Press.