

Research Article

Small Area Estimation of Household Consumption-expenditure Pattern in Nigeria During COVID-19 Pandemic

Udofia Blessing-Oxford Udeme, Iseh Matthew Joshua^{*} ,
Bassey Mbuotidem Okon

Department of Statistics, Akwa Ibom State University, Mkpato Enin, Nigeria

Abstract

The problem of Small Area Estimation is the non-availability of sample data in areas of interest. The idea behind this study is to adopt and modify some calibration estimators that could produce reliable estimates with minimum mean square error in small areas to determine the pattern of household consumption-expenditure in Nigeria before, during and after COVID-19 pandemic. A combined direct and synthetic ratio/regression estimators are used in the formulation of the longitudinal estimators. The bias and mean square errors of the estimators are derived using Taylor's series approximation techniques different from the existing estimators. It is observed that the calibrated estimators have provided more reliable estimates against the instability of the existing synthetic estimators and the higher variance of the existing direct estimators. Consequently, the gains made on the performance of the modified estimators cannot be overemphasized. From the empirical results, the performance of the suggested estimators are outstanding using the average mean square error, average relative bias and average coefficient of variation across the survey periods (WAVES) of 2019, 2020 and 2021. This indicates that the use of auxiliary variable (income) into the existing estimators by calibration technique has yielded the desirable result which agrees with the literature. Again, this result is validated since the modified calibrated estimators provide estimates within the acceptable region of 25% benchmark of the average coefficient of variation in the area of interest. In addition, the performance of the estimators in predicting the estimates of the population mean expenditure are also carried out. The pattern of household consumption-expenditure signifies that households in Nigeria consumed more during COVID-19 period while at home and the consumption burden lessens after the pandemic. This study has established the use of auxiliary variable that is strongly correlated with the study variable in domain estimation where there is small/no sample data in areas of interest.

Keywords

Auxiliary Variable, Calibration, COVID-19, Domain Estimation, Longitudinal Survey

1. Introduction

Long time, sample surveys have been a useful tool for gathering data about the population of interest for appropriate policy formulation and implementation. Sarndal et al. and Rao

claim that sample surveys became popular in 1930 because they were economical and time-efficient while still yielding accurate data on the demographic characteristics of interest

^{*}Corresponding author: blessingoxfordu@gmail.com (Udofia Blessing-Oxford Udeme)

Received: 14 June 2025; **Accepted:** 30 June 2025; **Published:** 23 July 2025



Copyright: © The Author(s), 2025. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

through statistical inference [1, 2] respectively. The main essence of survey sampling is to obtain statistical information about a finite population by selecting a sample from the population, measuring the required information about the units in this sample and estimating its finite population parameters such as means, totals, variances, standard deviations, proportions, ratios etc. These data are often used for analytic studies or analyses of a survey.

A variety of methods are available to obtain data from individuals like face to face interview, data collection through telephone surveys, mail surveys and computer aided data collection methods, etc. Data collection can also be made through email, internet or interactive voice recognition. Besides these methods, mixed mode methods are also used for data collection purpose in which more than one method of data collection is used. Large-Scale sample surveys are usually designed to produce reliable estimates of various characteristic of interest for large geographic areas. However, for effective planning of health, social and other services, and for apportioning government funds, there is a growing demand to produce similar estimates for smaller geographic areas and sub-populations, called small areas, for which adequate samples are not available [3, 4].

The Central Bank of Nigeria (CBN) conducts a number of surveys on a weekly, monthly, quarterly, and annual basis, including the Business Expectation Surveys (BES), Consumer Expectation Surveys (CES), Domestic Consumer Price Surveys (DCPS), Integrated Household Surveys (IHS), Household Finances and Consumption Surveys (HFCS), and Credit Condition Surveys (CCS). These surveys are used to collect data on a wide range of topics in many aspects of life in order to derive reliable estimators of means and totals for sizable regions or domains.

Surveys dealing with multiple observations spread over time are defined as longitudinal surveys. And when one is concerned about estimates of population sub-groups in longitudinal surveys, particularly the pattern of changes at individual level as well as aggregate level over time, the techniques of small areas estimation (SAE) can possibly be employed to achieve such goal. Small domain or area refers to a population for which reliable statistics of interest cannot be produced due to certain limitations of the available data. Examples of domains include a geographical region (e.g. a state, country, census enumeration area, etc.), a demographic group (e.g. age, gender) within a geographic region etc.

Small area estimation (SAE) is categorized into two types of estimators: direct and indirect estimators. A direct estimator is one that uses values of the variable of interest, only from the sample units in the domain of interest. However, a major disadvantage of such estimators is that for a small area, the usual direct survey estimators are likely to yield unacceptably large standard errors due to the unduly small size of the sample in the domain. An indirect estimator uses values of the variable of interest from a domain and/or time period other than the domain and time period of interest. There is a wide

variety of indirect estimators available, and a good summary is provided in [5]. We will confine ourselves to just a few of them that include the synthetic estimator in [6, 7].

Hidiroglou and Patak compared a number of the direct estimators [8]. One of their conclusions was that the direct estimators would be best if the domains of interest coincided as closely as possible with the design strata. The synthetic estimator uses reliable information of a direct estimator for a large area that spans several small areas, and this information is used to obtain an indirect estimator for a small area. It is assumed that the small areas have the same characteristics as the large area. Gonzalez provides a good account of how these estimators were obtained and used to obtain unemployment statistics at levels lower than those planned in the survey design [6].

Singh and Sisodia proposed direct estimators in domain estimation under longitudinal survey [9]. But direct estimators are characterized by large variances and relatively low precision which sometimes becomes impossible to compute in areas of small/no sample size. The shortfall in their work was that it does not contain auxiliary variables which could increase the precision of the estimators. More so, synthetic estimators which could produce estimates in smaller domains even with small/no sample size with minimum mean square error usually seems to be unstable. However, [10] calibrated on the Singh and Sisodia direct and synthetic estimators to obtain a reliable result.

On that basis, this study attempts to propose a methodology for estimation of population parameters in longitudinal surveys with special reference to study the temporal patterns of changes in population parameters with respect to small area.

2. Theoretical Framework

Rao, Ghosh and Rao, Lahiri and Meza, Pfeffermann, Hidiroglou and Patak among others, gave comprehensive reviews of small area estimation [2, 11-14] respectively. 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 estimation method having known that SAE is likely to be used in the survey [15, 16].

Sarndal *et al.*, Rao, Molina and Rao and Khare *et al.* compared direct estimators with synthetic estimators when reliable units are not directly accessible in the studied domain [1, 2, 17, 18] respectively. In addition, while [18], suggested a class of direct estimators for domain mean with the use of a single auxiliary character, [19] used calibration technique to obtain direct domain estimator with one auxiliary variable. The authors concluded that the proposed estimators performed better than the direct ratio estimator for domain mean using a single auxiliary character.

Many studies dealing with the small area estimation problem have been discussed by various authors [5, 7, 9, 11,

20], etc. However, [7] argues that though it will never be possible to anticipate all survey uses, or to allocate sufficient sample sizes to all domains of interest, so that indirect estimators will always be needed, it is possible to make design choices that will greatly improve the ability of national surveys to support direct estimation for many small areas. Authors like [6, 10, 21-23] etc. 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 [1]. However, [10] proposed a calibrated synthetic estimator of the population mean which addresses these problematic issues of estimates that are highly biased with invalid confidence statements. The estimators were examined through simulation studies under three distributional assumptions, namely the normal, gamma and exponential distributions. The results showed that they provide estimates of the mean displaying less relative bias and greater efficiency. Moreover, they proved more consistent than the existing classical synthetic estimator. Calibration estimation has been found to be very fruitful in minimizing survey errors and thereby obtaining a reliable estimators. [24].

This study would seek to modify the [9, 10] estimators in a longitudinal survey for a reliable estimation of the population mean of the consumption-expenditure pattern of households in Nigeria during before, during and after the COVID-19 pandemic.

3. Methodology

3.1. Data Source

In order to illustrate the performance of the estimators, data from the Statistics Department of the Central Bank of Nigeria, is used in this study. This real-life data was obtained from the Integrated Household Finance & Consumption Survey (IH-FCS) conducted between 2019 and 2021. This survey is an orchestration of two distinct longitudinal household-based surveys, namely: Household Finance and Consumption Survey (HFCS), and the Integrated Household (IHS). This longitudinal household survey (IH-FCS) remains the most comprehensive survey for gathering information on the assets and liabilities of households and members of the households aged 18 years and above on a quarterly basis popularly called WAVE. The IH-FCS sampling procedure involves a multi-stage probability sampling for the households and their respective enumeration areas. The survey is planned for every quarter of the year (4 waves in a year) as a routine data generation activity of the Statistics Department and the reporting domains are the 36 states of the Federation and the Federal

Capital Territory (FCT).

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 population mean of y for all the 37 domains and obtain the average population mean expenditure of households for both rural and urban areas of Nigeria during the COVID-19 pandemic era.

3.2. Sample Design

For a typical domain estimation, consider a finite population consisting of N units which is divided into D non-overlapping domains U_d , $d=1,2,\dots, D$ with N_d units such that $\sum_d^D N_d = N$. Let the population be further partitioned in G non-overlapping groups (considered to be strata) which are considered to be larger than the domains U_g , $g=1,2,\dots, G$ 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 $U = \sum_{d=1}^D U_d = \sum_{g=1}^G U_g = \sum_{d=1}^D \sum_{g=1}^G U_{dg}$ 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 sub-samples s_d , group sub-samples s_g and cells sub-samples s_{dg} with corresponding sample size n, n_d, n_g and n_{dg} as $s = \sum_{d=1}^D s_d = \sum_{g=1}^G s_g = \sum_{d=1}^D \sum_{g=1}^G s_{dg}$ and $n = \sum_d^D n_d = \sum_g^G n_g = \sum_d^D \sum_g^G n_{dg}$. The cell sub-sample 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 and X be the characteristic of interest and auxiliary variable respectively and object is to estimate the population mean, $\bar{Y}_d^{(t)}$, for a specific point of time t ; $t=1,2,\dots,T$ and $d=1,2,\dots,D$ such that: $\bar{Y}_d^{(t)} = \sum_{g=1}^G \sum_{k=1}^{n_{dg}} \frac{Y_{dgk}}{n_{dg}}$.

3.3. Some Existing Estimators

3.3.1. Singh and Sisodia Estimators [9]

Direct Estimator for longitudinal studies

$$\bar{y}_d^{(t)} = \sum_{g=1}^G W_{dg} \bar{y}_{dg}^{(t)} \quad (1)$$

where $\bar{y}_{dg}^{(t)} = \sum_k^{n_{dg}} \frac{y_{dgk}^{(t)}}{n_{dg}}$, and $W_{dg} = \frac{N_{dg}}{N_d}$

with variance $V(\bar{y}_d^{(t)}) = \sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_{dg}} - \frac{1}{N_d} \right) S_{dg}^{2(t)}$

The Synthetic Estimator for longitudinal studies

$$\bar{y}_s^{(t)} = \sum_{g=1}^G W_{dg} \bar{y}_g^{(t)} \quad (2)$$

where $\bar{y}_g^{(t)} = \sum_d^D \sum_k^{n_{dg}} \frac{y_{dgk}^{(t)}}{n_g}$, $n_g = \sum_{d=1}^D n_{dg}$

with bias $B(\bar{y}_s^{(t)}) = \sum_{g=1}^G W_{dg} (\bar{Y}_{.g}^{(t)} - \bar{Y}_{dg}^{(t)})$, and mean square error given as

$$MSE(\bar{y}_s^{(t)}) = \left[\sum_{g=1}^G W_{dg} (\bar{Y}_{.g}^{(t)} - \bar{Y}_{dg}^{(t)}) \right]^2 + \sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_{dg}} - \frac{1}{N_{dg}} \right) S_{dg}^2(t)$$

3.3.2. Iseh and Enang Synthetic Estimator [10]

Iseh and Enang proposed calibrated synthetic estimators as follows;

$$\hat{\bar{y}}_s = \sum_{g=1}^G W_{dg} \bar{y}_{.g} + \frac{(\sum_{g=1}^G W_{dg} q_{dg} \bar{x}_{.g} \bar{y}_{.g})}{\sum_{g=1}^G W_{dg} q_{dg} \bar{x}_{.g}^2} (\bar{X}_d - \sum_{g=1}^G W_{dg} \bar{x}_{.g}) \quad (3)$$

with variance

$$\hat{V}_H(\hat{\bar{y}}_s) = \sum_{g=1}^G \frac{D_g(W_{dg}^2)}{W_{dg}^2} S_{dg}^2 + \frac{\sum_{g=1}^G D_g(W_{dg}^2) q_{dg} S_{dg}^2}{\sum_{g=1}^G D_g W_{dg}^2 q_{dg} (S_{dg}^2)^2} S_{dg}^2 [\hat{V}_{st}(\bar{x}_d^s) - V_{st}(\bar{x}_d^s)]$$

Two special cases were considered

CASE 1: when $q_{dg} = \frac{1}{\bar{x}_{.g}}$, Eq. 3, becomes

$$\hat{\bar{y}}_{sr} = \frac{\sum_{g=1}^G W_{dg} \bar{y}_{.g}}{\sum_{g=1}^G W_{dg} \bar{x}_{.g}} \bar{X}_d \quad (4)$$

with

$$\text{Bias}(\hat{\bar{y}}_{sr}) = \frac{1}{\bar{X}_d} \sum_{g=1}^G W_{dg}^2 \gamma_{.g} (\hat{R}_d S_{dg}^2 - S_{gxy})$$

and

$$MSE(\hat{\bar{y}}_{sr}) = \left(\frac{\bar{X}_d}{\bar{x}_d^s} \right)^2 \left[\frac{V_{st}(\bar{x}_d^s)}{\bar{V}_{st}(\bar{x}_d^s)} \right] \sum_{g=1}^G W_{dg}^2 \gamma_{.g} S_{dg}^2$$

where $S_{dg}^2 = S_{gy}^2 + \hat{R}_d S_{gx}^2 - 2\hat{R}_d S_{gxy}$, and $R_d = \frac{\bar{Y}_d}{\bar{X}_d}$

Case 2: when $q_{dg} = 1$, equation (3) becomes

$$\hat{\bar{y}}_{sreg} = \sum_{g=1}^G W_{dg} \bar{y}_{.g} + \beta_1 (\bar{X}_d - \sum_{g=1}^G W_{dg} \bar{x}_{.g}) \quad (5)$$

Equation (5) is the combined synthetic regression estimator.

with $\text{Bias}(\hat{\bar{y}}_{sreg}) = \frac{1}{\bar{X}_d} \sum_{g=1}^G W_{dg}^2 \gamma_{.g} (b_{.g} S_{dg}^2 - S_{gxy})$

and $MSE(\hat{\bar{y}}_{sreg}) = \left(\frac{\bar{X}_d}{\bar{x}_d^s} \right)^2 \left[\frac{V_{st}(\bar{x}_d^s)}{\bar{V}_{st}(\bar{x}_d^s)} \right] \sum_{g=1}^G W_{dg}^2 \gamma_{.g} S_{dg}^2$

where $S_{dg}^2 = S_{gy}^2 + b_d^2 S_{gx}^2 - 2b_d S_{gxy}$ and $V_{st}(\bar{x}_d^s) = \sum_{g=1}^G D_g S_{gx}^2$ is the population variance of the auxiliary variable with estimate

$$\hat{V}_{st}(\bar{x}_d^s) = \sum_{g=1}^G D_g S_{gx}^2$$

3.4. Suggested Estimators

Modified Synthetic Estimator

i) *Proposition 1a*: A combined synthetic ratio estimator for longitudinal survey is bias and has mean square error as follows

$$\text{Bias}(\hat{\bar{y}}_{sr}^{(t)}) = \sum_{g=1}^G W_{dg} \bar{Y}_{.g}^{(t)} \left[\left(\frac{1}{n_{dg}} - \frac{1}{N_{dg}} \right) C_{x_{.g}}^2 - \left(\frac{1}{n_{dg}} - \frac{1}{N_{dg}} \right) \rho_{.g} C_{y_{.g}}^{(t)} C_{x_{.g}}^{(t)} \right]$$

$$MSE(\hat{\bar{y}}_{sr}^{(t)}) = \sum_{g=1}^G W_{dg}^2 \bar{Y}_{.g}^{(t)2} \left[\left(\frac{1}{n_{dg}} - \frac{1}{N_{dg}} \right) \{ C_{x_{.g}}^2 - 2\rho_{.g} C_{y_{.g}}^{(t)} C_{x_{.g}}^{(t)} + C_{x_{.g}}^2 \} \right]$$

Bias of $\hat{\bar{y}}_{sr}^{(t)}$

Proof:

Let

$$\begin{cases} \bar{y}_{dg}^{(t)} = \bar{Y}_{dg}^{(t)} (1 + e_0) \\ \bar{y}_{.g}^{(t)} = \bar{Y}_{.g}^{(t)} (1 + e_1) \\ \bar{x}_{.g}^{(t)} = \bar{X}_{.g}^{(t)} (1 + e_2) \end{cases} \quad (6)$$

Such that

$$E(e_0) = E(e_1) = E(e_2) = 0$$

$$E(e_0^2) = f_{dg} C_{ydg}^2, E(e_1^2) = f_{.g} C_{y.g}^2, E(e_2^2) = f_{.g} C_{x.g}^2$$

where $f_{dg} = \left(\frac{1}{n_{dg}} - \frac{1}{N_{dg}} \right)$, $f_{.g} = \left(\frac{1}{n_{.g}} - \frac{1}{N_{.g}} \right)$, $C_{ydg}^2 = \frac{S_{ydg}^2}{\bar{Y}_{dg}^2}$,

$$C_{x.g}^2 = \frac{S_{x.g}^2}{\bar{X}_{.g}^2}$$

$$C_{y.g}^2 = \frac{S_{y.g}^2}{\bar{Y}_{.g}^2}, C_{xdg}^2 = \frac{S_{xdg}^2}{\bar{X}_{dg}^2}$$

$$E(e_0 e_1) = f_{dg} \rho_{dg} C_{ydg} C_{y.g}, E(e_1 e_2) = f_{.g} \rho_{.g} C_{y.g} C_{x.g}$$

$$E(e_0 e_2) = f_{dg} \rho_{dg} C_{ydg} C_{x.g}$$

Eq. 4 could be written as

$$\hat{\bar{y}}_{sr}^{(t)} = \frac{\sum_{g=1}^G W_{dg} \bar{Y}_{.g}^{(t)} (1 + e_1)}{\sum_{g=1}^G W_{dg} \bar{X}_{.g}^{(t)} (1 + e_2)} \bar{X}_d \quad (7)$$

$$= \sum_{g=1}^G W_{dg} \bar{Y}_{.g}^{(t)} (1 + e_1) (1 + e_2)^{-1}$$

By using Taylor's series approximation, we obtained

$$\hat{\bar{y}}_{sr}^{(t)} - \bar{Y}_{.g}^{(t)} = \sum_{g=1}^G W_{dg} \bar{Y}_{.g}^{(t)} (e_2 + e_1 + e_2^2 - e_1 e_2)$$

Taking expectation, $E(\hat{\bar{y}}_{sr}^{(t)} - \bar{Y}_{.g}^{(t)}) = B(\hat{\bar{y}}_{sr}^{(t)})$

$$= \sum_{g=1}^G W_{dg} \bar{Y}_{.g}^{(t)} E(e_2 + e_1 + e_2^2 - e_1 e_2) \quad (8)$$

$$= \sum_{g=1}^G W_{dg} \bar{Y}_{.g}^{(t)} \left[\left(\frac{1}{n_{dg}} - \frac{1}{N_{dg}} \right) \{ C_{x_{.g}}^2 - \rho_{.g} C_{y_{.g}}^{(t)} C_{x_{.g}}^{(t)} \} \right] \quad (9)$$

which proves that the combined synthetic ratio estimator under longitudinal survey is biased.

MSE of $\hat{y}_{sr}^{(t)}$

Unlike [10] that made use of calibration approach in the derivation of the MSE, here, the Taylor's approximation

method is adopted to obtain the MSE as follows:

By squaring both sides of Eq. 8 to the first order of approximation, we have

$$E\left(\hat{y}_{sr}^{(t)} - \bar{Y}_g^{(t)}\right)^2 = E\left[\sum_{g=1}^G W_{dg} \bar{Y}_g^{(t)} (e_2 + e_1 + e_2^2 - e_1 e_2)\right]^2$$

$$MSE(\hat{y}_{sr}^{(t)}) = \sum_{g=1}^G W_{dg}^2 \bar{Y}_g^{(t)2} \left[\left(\frac{1}{n_g} - \frac{1}{N_g}\right) \{C_{y_g}^2 - 2\rho C_{y_g} C_{x_g} + C_{x_g}^2\}\right] \quad (10)$$

which shows the MSE of the combined synthetic ratio estimator under longitudinal survey.

ii) *Proposition 1b*: A combined synthetic regression estimator for longitudinal survey is unbiased and has variance given as follows:

$$V(\hat{y}_{sreg}^{(t)}) = \sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_g} - \frac{1}{N_g}\right) S_{y_g}^2 + B_s^2 \sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_g} - \frac{1}{N_g}\right) S_{x_g}^2 - 2B_s \sum_{g=1}^G W_{dg} \left(\frac{1}{n_g} - \frac{1}{N_g}\right) \rho_g S_{y_g} S_{x_g}$$

Unbiasedness of $\hat{y}_{sreg}^{(t)}$

Proof

To show that $\hat{y}_{sreg}^{(t)}$ is asymptotically unbiased with $E(\hat{y}_{sreg}^{(t)}) \approx \bar{Y}_d^{(t)}$

We rewrite Eq. 5 using the large sample approximation for $\bar{y}_g^{(t)}$ and $\bar{x}_g^{(t)}$,

$$\hat{y}_{sreg}^{(t)} = \sum_{g=1}^G W_{dg} \bar{Y}_g^{(t)} (1 + e_1) + B_s (\bar{X}_d^{(t)} - \sum_{g=1}^G W_{dg} \bar{X}_g^{(t)} (1 + e_2)) \quad (11)$$

$$= \sum_{g=1}^G W_{dg} \bar{Y}_g^{(t)} + \sum_{g=1}^G W_{dg} \bar{Y}_g^{(t)} e_1 + B_s (\bar{X}_d^{(t)} - \sum_{g=1}^G W_{dg} \bar{X}_g^{(t)} - \sum_{g=1}^G W_{dg} \bar{X}_g^{(t)} e_2)$$

Taking expectation of both sides

$$\begin{aligned} E(\hat{y}_{sreg}^{(t)}) &= \sum_{g=1}^G W_{dg} \bar{Y}_g^{(t)} + \sum_{g=1}^G W_{dg} \bar{Y}_g^{(t)} E(e_1) + \\ &B_s (\bar{X}_d^{(t)} - \sum_{g=1}^G W_{dg} \bar{X}_g^{(t)} - \sum_{g=1}^G W_{dg} \bar{X}_g^{(t)} E(e_2)) \\ &= \sum_{g=1}^G W_{dg} \bar{Y}_g^{(t)} + B_s (\bar{X}_d^{(t)} - \sum_{g=1}^G W_{dg} \bar{X}_g^{(t)}) \\ &= \bar{Y}_d^{(t)} + B_s (\bar{X}_d^{(t)} - \bar{X}_d^{(t)}) \\ &\approx \bar{Y}_d^{(t)} \end{aligned}$$

which proved that the proposed combined synthetic regression estimator is unbiased.

Mean square error of $\hat{y}_{sreg}^{(t)}$

Proof

Unlike [10] that made use of calibration approach in the derivation of the variance, here, the traditional method will be adopted to obtain the variance as follows:

$$\begin{aligned} MSE(\hat{y}_{sreg}^{(t)}) &= E[\hat{y}_{sreg}^{(t)} - E(\hat{y}_{sreg}^{(t)})]^2 \\ &= E[\sum_{g=1}^G W_{dg} \bar{Y}_g^{(t)} + B_s (\bar{X}_d^{(t)} - \sum_{g=1}^G W_{dg} \bar{X}_g^{(t)}) - \bar{Y}_d^{(t)}]^2 \\ &= E[\sum_{g=1}^G W_{dg} \bar{Y}_g^{(t)} (1 + e_1) + B_s (\bar{X}_d^{(t)} - \sum_{g=1}^G W_{dg} \bar{X}_g^{(t)} (1 + e_2)) - \bar{Y}_d^{(t)}]^2 \\ &= E[\sum_{g=1}^G W_{dg} \bar{Y}_g^{(t)} e_1 - B_s (\sum_{g=1}^G W_{dg} \bar{X}_g^{(t)} e_2)]^2 \end{aligned}$$

Expanding the binomial, we have

Recall that $E(e_1) = E(e_2) = 0$

therefore,

$$\begin{aligned} &= \sum_{g=1}^G W_{dg}^2 \bar{Y}_g^{(t)2} E(e_1^2) - 2B_s \sum_{g=1}^G W_{dg} \bar{Y}_g^{(t)} \bar{X}_g^{(t)} E(e_1 e_2) + B_s^2 \sum_{g=1}^G W_{dg}^2 \bar{X}_g^{(t)2} E(e_2^2) \\ &= \sum_{g=1}^G W_{dg}^2 \bar{Y}_g^{(t)2} \left[\left(\frac{1}{n_g} - \frac{1}{N_g}\right) \frac{S_{y_g}^2}{\bar{Y}_g^{(t)2}}\right] + B_s^2 \sum_{g=1}^G W_{dg}^2 \bar{X}_g^{(t)2} \left(\frac{1}{n_g} - \frac{1}{N_g}\right) \frac{S_{x_g}^2}{\bar{X}_g^{(t)2}} - 2B_s \sum_{g=1}^G W_{dg} \bar{Y}_g^{(t)} \bar{X}_g^{(t)} \left(\frac{1}{n_g} - \frac{1}{N_g}\right) \rho_g \frac{S_{y_g} S_{x_g}}{\bar{Y}_g^{(t)} \bar{X}_g^{(t)}} \\ &= \sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_g} - \frac{1}{N_g}\right) S_{y_g}^2 + B_s^2 \sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_g} - \frac{1}{N_g}\right) S_{x_g}^2 - 2B_s \sum_{g=1}^G W_{dg} \left(\frac{1}{n_g} - \frac{1}{N_g}\right) \rho_g S_{y_g} S_{x_g} \\ V(\hat{y}_{sreg}^{(t)}) &= \sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_g} - \frac{1}{N_g}\right) S_{y_g}^2 + B_s^2 \sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_g} - \frac{1}{N_g}\right) S_{x_g}^2 - 2B_s \sum_{g=1}^G W_{dg} \left(\frac{1}{n_g} - \frac{1}{N_g}\right) \rho_g S_{y_g} S_{x_g} \quad (12) \end{aligned}$$

Optimum choice of B_s

Differentiating equation(12) with respect to B and equating to zero, we have

$$\frac{\partial V(\hat{y}_{sreg}^{(t)})}{\partial B_s} = 2B_s \sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_g} - \frac{1}{N_g} \right) S_{x_g}^2 - 2 \sum_{g=1}^G W_{dg} \left(\frac{1}{n_g} - \frac{1}{N_g} \right) \rho_{.g} S_{y_g} S_{x_g} = 0$$

$$\hat{B}_s = \frac{\sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_g} - \frac{1}{N_g} \right) \rho_{.g} S_{y_g} S_{x_g}}{\sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_g} - \frac{1}{N_g} \right) S_{x_g}^2}$$

Substitute \hat{B}_s in Eq. 12

$$\begin{aligned} \text{MSE}(\hat{y}_{sreg}^{(t)}) &= \sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_g} - \frac{1}{N_g} \right) S_{y_g}^2 + \left(\frac{\sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_g} - \frac{1}{N_g} \right) \rho_{.g} S_{y_g} S_{x_g}}{\sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_g} - \frac{1}{N_g} \right) S_{x_g}^2} \right)^2 \sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_g} - \frac{1}{N_g} \right) S_{x_g}^2 - 2 \left(\frac{\sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_g} - \frac{1}{N_g} \right) \rho_{.g} S_{y_g} S_{x_g}}{\sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_g} - \frac{1}{N_g} \right) S_{x_g}^2} \right) \sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_g} - \frac{1}{N_g} \right) \rho_{.g} S_{y_g} S_{x_g} \\ &= \sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_g} - \frac{1}{N_g} \right) S_{y_g}^2 - \frac{\left[\sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_g} - \frac{1}{N_g} \right) \rho_{.g} S_{y_g} S_{x_g} \right]^2}{\sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_g} - \frac{1}{N_g} \right) S_{x_g}^2} \\ \text{MSE}(\hat{y}_{sreg}^{(t)}) &= \sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_g} - \frac{1}{N_g} \right) S_{y_g}^2 - \frac{\sum_{g=1}^G W_{dg}^4 \rho_{.g}^2 \left(\frac{1}{n_g} - \frac{1}{N_g} \right)^2 S_{y_g}^2 S_{x_g}^2}{\sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_g} - \frac{1}{N_g} \right) S_{x_g}^2} \end{aligned} \quad (13)$$

3.5. Modified Direct Estimator

A direct estimator by [9] under longitudinal survey is hereby modified by incorporating an auxiliary variable through the concept of calibration as follows

$$\hat{y}_{dc}^{(t)} = \sum_{g=1}^G W_{dg} \bar{y}_{dg}^{(t)} \quad (14)$$

Let the calibration estimator be given as

$$\hat{y}_{dc}^{(t)} = \sum_{g=1}^G \Omega_{dg} \bar{y}_{dg}^{(t)} \quad (15)$$

where Ω_{dg} is the calibration weight chosen such that the distance function

$$\frac{\sum_{g=1}^G (\Omega_{dg} - W_{dg})^2}{2W_{dg} q_{dg}} \text{ is minimized subject to the constraint}$$

$$\sum_{g=1}^G \Omega_{dg} \bar{x}_{dg}^{(t)} = \bar{X}_d^{(t)} \quad (16)$$

The Lagrange function ϕ given as

$$\phi = \frac{\sum_{g=1}^G (\Omega_{dg} - W_{dg})^2}{2W_{dg} q_{dg}} - \lambda [\sum_{g=1}^G \Omega_{dg} \bar{x}_{dg}^{(t)} - \bar{X}_d^{(t)}] \text{ is differentiated with respect to } \Omega_{dg} \text{ and equated to zero such that}$$

$$\Omega_{dg} = W_{dg} [1 + \lambda \bar{x}_{dg}^{(t)} q_{dg}] \quad (17)$$

$$\text{with } \lambda = \frac{\bar{X}_d^{(t)} - \sum_{g=1}^G W_{dg} \bar{x}_{dg}^{(t)}}{\sum_{g=1}^G W_{dg} q_{dg} \bar{x}_{dg}^{(t)2}}, \text{ then with proper substitution in Eq.}$$

17, the calibration direct estimator is given as

$$\hat{y}_{dc}^{(t)} = \sum_{g=1}^G W_{dg} \bar{y}_{dg}^{(t)} + \left[\frac{\sum_{g=1}^G W_{dg} q_{dg} \bar{x}_{dg}^{(t)} \bar{y}_{dg}^{(t)}}{\sum_{g=1}^G W_{dg} q_{dg} \bar{x}_{dg}^{(t)2}} \right] (\bar{X}_d^{(t)} - \sum_{g=1}^G W_{dg} \bar{x}_{dg}^{(t)}) \quad (18)$$

Eq. 18 is known as Calibrated Direct Estimator

3.6. Calibrated Direct Ratio and Regression Estimators

Consider the following two cases, by making appropriate substitutions, combined direct ratio and regression estimators are obtain as follows

Proposed combined direct ratio estimator

Case 1: Assuming the tuning parameter in Eq. 18 is

$$q_{dg} = \frac{1}{\bar{x}_{dg}^{(t)}}$$

Then the combined direct ratio estimator is given as:

$$\hat{y}_{dr}^{(t)} = \frac{\sum_{g=1}^G W_{dg} \bar{y}_{dg}^{(t)}}{\sum_{g=1}^G W_{dg} \bar{x}_{dg}^{(t)}} \bar{X}_d^{(t)} \quad (19)$$

Proposition 2a: The direct combined ratio estimator in Eq. 19 is biased with MSE given as follows:

$$\text{Bias}(\hat{y}_{dr}^{(t)}) = \sum_{g=1}^G W_{dg} \bar{y}_{dg}^{(t)} \left[\left(\frac{1}{n_{dg}} - \frac{1}{N_{dg}} \right) C_{x_{dg}}^2 - \left(\frac{1}{n_{dg}} - \frac{1}{N_{dg}} \right) \rho_{.g} C_{y_{dg}} C_{x_{dg}} \right]$$

$$\begin{aligned} \text{MSE}(\hat{y}_{dr}^{(t)}) &= \sum_{g=1}^G W_{dg}^2 \bar{y}_{dg}^{(t)2} \left[\left(\frac{1}{n_{dg}} - \frac{1}{N_{dg}} \right) \{ C_{x_{dg}}^2 - 2\rho_{.g} C_{y_{dg}} C_{x_{dg}} \} + \right. \\ &\quad \left. \left(\frac{1}{n_g} - \frac{1}{N_g} \right) C_{x_{dg}}^2 \right] \end{aligned}$$

Proof:

Bias of $\hat{y}_{dr}^{(t)}$

Let

$$\begin{cases} \bar{y}_{dg}^{(t)} = \bar{Y}_d^{(t)}(1 + e_0) \\ \bar{x}_{dg}^{(t)} = \bar{X}_d^{(t)}(1 + e_2) \end{cases} \quad (20)$$

Such that

$$E(e_0) = E(e_2) = 0, E(e_0^2) = f_{dg} C_{ydg}^{2(t)}, E(e_2^2) = f_{dg} C_{xdg}^{2(t)}$$

$$\text{Where, } f_{dg} = \left(\frac{1}{n_{dg}} - \frac{1}{N_{dg}} \right), C_{xdg}^{2(t)} = \frac{S_{xdg}^{2(t)}}{\bar{X}_d^{2(t)}}, C_{ydg}^{2(t)} = \frac{S_{ydg}^{2(t)}}{\bar{Y}_d^{2(t)}},$$

$$E(e_0 e_2) = f_{dg} \rho_{dg} C_{ydg} C_{xdg}$$

Therefore Eq. 19 becomes

$$\begin{aligned} \hat{y}_{dr}^{(t)} &= \frac{\sum_{g=1}^G W_{dg} \bar{y}_{dg}^{(t)} (1 + e_0)}{\sum_{g=1}^G W_{dg} \bar{x}_{dg}^{(t)} (1 + e_2)} \bar{X}_d^{(t)} \\ &= \sum_{g=1}^G W_{dg} \bar{Y}_d^{(t)} (1 + e_0) (1 + e_2)^{-1} \end{aligned}$$

$$MSE(\hat{y}_{dr}^{(t)}) = \sum_{g=1}^G W_{dg}^2 \bar{Y}_d^{(t)2} \left(\frac{1}{n_{dg}} - \frac{1}{N_{dg}} \right) \left[C_{ydg}^2 + C_{xdg}^2 - 2\rho_{dg} C_{ydg} C_{xdg} \right] \quad (23)$$

Case 2: Assuming the turning parameter in Eq. 18 is $q_{dg} = 1$, yields a direct combined regression estimator as follows

$$\hat{y}_{dreg}^{(t)} = \sum_{g=1}^G W_{dg} \bar{y}_{dg}^{(t)} + B_d (\bar{X}_d^{(t)} - \sum_{g=1}^G W_{dg} \bar{x}_{dg}^{(t)}) \quad (24)$$

which is in the form of a direct combined regression estimator under longitudinal survey.

where

$$B_d = \frac{\sum_{g=1}^G W_{dg} \bar{x}_{dg}^{(t)} \bar{y}_{dg}^{(t)}}{\sum_{g=1}^G W_{dg} \bar{x}_{dg}^{(t)2}}$$

$$\begin{aligned} \hat{y}_{dreg}^{(t)} &= \sum_{g=1}^G W_{dg} \bar{Y}_d^{(t)} (1 + e_0) + B_d (\bar{X}_d^{(t)} - \sum_{g=1}^G W_{dg} \bar{X}_{dg}^{(t)} (1 + e_2)) \\ &= \sum_{g=1}^G W_{dg} \bar{Y}_d^{(t)} + \sum_{g=1}^G W_{dg} \bar{Y}_d^{(t)} e_0 + B_d (\bar{X}_d^{(t)} - \sum_{g=1}^G W_{dg} \bar{X}_{dg}^{(t)} - \sum_{g=1}^G W_{dg} \bar{X}_{dg}^{(t)} e_2) \end{aligned}$$

Taking expectation of both sides

$$\begin{aligned} E(\hat{y}_{dreg}^{(t)}) &= \sum_{g=1}^G W_{dg} \bar{Y}_d^{(t)} + \sum_{g=1}^G W_{dg} \bar{Y}_d^{(t)} E(e_0) + B_d (\bar{X}_d^{(t)} - \sum_{g=1}^G W_{dg} \bar{X}_{dg}^{(t)} - \sum_{g=1}^G W_{dg} \bar{X}_{dg}^{(t)} E(e_2)) \\ &= \sum_{g=1}^G W_{dg} \bar{Y}_d^{(t)} + B_d (\bar{X}_d^{(t)} - \sum_{g=1}^G W_{dg} \bar{X}_{dg}^{(t)}) \\ &= \bar{Y}_d^{(t)} + B_d (\bar{X}_d^{(t)} - \bar{X}_d^{(t)}) \\ &\approx \bar{Y}_d^{(t)} \end{aligned} \quad (25)$$

By Taylor's series approximation and taking expectation, we have

$$\hat{y}_{dr}^{(t)} - \bar{Y}_d^{(t)} = \sum_{g=1}^G W_{dg} \bar{Y}_d^{(t)} (e_0 - e_2 + e_2^2 - e_0 e_2) \quad (21)$$

$$\begin{aligned} Bias(\hat{y}_{dr}^{(t)}) &= \sum_{g=1}^G W_{dg} \bar{Y}_d^{(t)} \left[\left(\frac{1}{n_{dg}} - \frac{1}{N_{dg}} \right) C_{x_{dg}}^2 - \left(\frac{1}{n_{dg}} - \frac{1}{N_{dg}} \right) \rho_{dg} C_{y_{dg}} C_{x_{dg}} \right] \end{aligned} \quad (22)$$

This shows that the proposed direct combined ratio estimator under longitudinal survey is biased.

MSE of $\hat{y}_{dr}^{(t)}$

To obtain the MSE, we proceed as follows:

By squaring both sides of Eq. 21 and taking expectation to the first order of approximation gives

$$E(\hat{y}_{dr}^{(t)} - \bar{Y}_d^{(t)})^2 = E\{\sum_{g=1}^G W_{dg}^2 \bar{Y}_d^{(t)2} (e_0^2 + e_2^2 - 2e_0 e_2)\}$$

Proposition 2b: The direct combined regression estimator in Eq. 24 is asymptotically unbiased and has mean square error given as follows:

$$\begin{aligned} MSE(\hat{y}_{sr}^{(t)}) &= \sum_{g=1}^G W_{dg}^2 \left[\left(\frac{1}{n_{dg}} - \frac{1}{N_{dg}} \right) S_{y_{dg}}^2 \right] + B_d^2 \sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_{dg}} - \frac{1}{N_{dg}} \right) S_{x_{dg}}^2 \\ &\quad - 2B_d \sum_{g=1}^G W_{dg} \left(\frac{1}{n_{dg}} - \frac{1}{N_{dg}} \right) \rho_{dg} S_{y_{dg}} S_{x_{dg}} \end{aligned}$$

Proof

Unbiasedness of $\hat{y}_{dreg}^{(t)}$

To show that $E(\hat{y}_{dreg}^{(t)}) \approx \bar{Y}_d^{(t)}$

Substituting the large sample approximation earlier defined for $\bar{y}_{dg}^{(t)}$ and $\bar{x}_{dg}^{(t)}$ in Eq. 24

which proved that the proposed combined direct regression estimator is asymptotically unbiased.

MSE of $\hat{y}_{dreg}^{(t)}$

To obtain the MSE, we proceed as follows

$$MSE(\hat{y}_{dreg}^{(t)}) = E[\hat{y}_{dreg}^{(t)} - E(\hat{y}_{dreg}^{(t)})]^2$$

$$MSE(\hat{y}_{dreg}^{(t)}) = \sum_{g=1}^G W_{dg}^2 \left[\left(\frac{1}{n_{dg}} - \frac{1}{N_{dg}} \right) S_{y_{dg}}^2 \right] + B_d^2 \sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_g} - \frac{1}{N_g} \right) S_{x_g}^2 - 2B_d \sum_{g=1}^G W_{dg} \left(\frac{1}{n_{dg}} - \frac{1}{N_{dg}} \right) \rho_{dg} S_{y_g} S_{x_g} \quad (26)$$

Optimum Choice of B_d

Differentiating Eq. 26 with respect to B_d and equating to zero, we have

$$\hat{B}_d = \frac{\sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_{dg}} - \frac{1}{N_{dg}} \right) \rho_{dg} S_{y_g} S_{x_g}}{\sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_g} - \frac{1}{N_g} \right) S_{x_g}^2}$$

Substituting for \hat{B}_d in Eq. 26 with proper simplification

$$MSE(\hat{y}_{dreg}^{(t)}) = \sum_{g=1}^G W_{dg}^2 \left[\left(\frac{1}{n_{dg}} - \frac{1}{N_{dg}} \right) S_{y_{dg}}^2 \right] - \frac{\left[\sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_{dg}} - \frac{1}{N_{dg}} \right) \rho_{dg} S_{y_g} S_{x_g} \right]^2}{\sum_{g=1}^G W_{dg}^2 \left(\frac{1}{n_g} - \frac{1}{N_g} \right) S_{x_g}^2} \quad (27)$$

4. Empirical Investigation

As stated in section 3.1, real-life data obtained from the Household Finances and Consumption Survey (HFCS) and the Integrated Household Survey (IHS) which is a quarterly survey is partitioned into WAVE 1, 3 & 4 for 2019 (before COVID-19), WAVE 1, 2, 3 & 4 for 2020 (during COVID-19) and WAVE 2, 3 & 4 for 2021 (After COVID-19). The population is comprised of selected households spread across the 37 states (domains) of Nigeria. Again, the population is sub-divided into two groups (strata) of Urban and Rural dwellers.

Performance of Estimators Using R

A simulation of real-life data using R is computed to demonstrate the performance of the respective domain estimators in each WAVE of 2019, 2020 and 2021 which signifies before, during and after the COVID-19 pandemic. The simulation study is carried out using simple random sampling without replacement for 1000 replications. The domain estimators, estimate the mean population of the expenditure pattern in each state of the federation over a period of time. The results of performance of the estimators for mean estimates, average relative bias (ARB), average mean square error (AMSE) and average coefficient of variation (ACV) are presented in Tables.

Table 1. Performance of Estimators in WAVE 1 of 2019.

STATES	\bar{y}_d^1	\bar{y}_s^1	\hat{y}_{sr}^1	\hat{y}_{sreg}^1	\hat{y}_{dr}^1	\hat{y}_{dreg}^1
AVE	2577.157	2015.51	34683.49	34720.94	44568.35	43248.92
AMSE	1309547910	1378131713	287718902	288046917	642199144	625137340
ARB	0.9357807	0.9340782	0.1771801	0.1991829	0.3373016	0.1606586
ACV	0.9357807	0.9340782	0.5082464	0.5090257	0.3329459	0.322272

STATES	\bar{y}_d^1	\bar{y}_s^1	$\hat{\bar{y}}_{sr}^1$	$\hat{\bar{y}}_{sreg}^1$	$\hat{\bar{y}}_{dr}^1$	$\hat{\bar{y}}_{dreg}^1$
\bar{Y}	39282.73					

Table 2. Performance of Estimators in WAVE 3 of 2019.

STATES	\bar{y}_d^3	\bar{y}_s^3	$\hat{\bar{y}}_{sr}^3$	$\hat{\bar{y}}_{sreg}^3$	$\hat{\bar{y}}_{dr}^3$	$\hat{\bar{y}}_{dreg}^3$
AVE	2178.997	2208.07	40178.4	40027.83	38493.16	35490.35
AMSE	1386350590	1748851321	550436965	546505645	455246660	482783142
ARB	0.9444336	0.9389944	0.1481006	0.1699649	0.1919592	0.05885609
ACV	0.9444336	0.9142011	0.3601048	0.3575546	0.2292409	0.2547304
\bar{Y}		40545.98				

Table 3. Performance of Estimators in WAVE 4 of 2019.

STATES	\bar{y}_d^4	\bar{y}_s^4	$\hat{\bar{y}}_{sr}^4$	$\hat{\bar{y}}_{sreg}^4$	$\hat{\bar{y}}_{dr}^4$	$\hat{\bar{y}}_{dreg}^4$
AVE	1273.127	1510.836	24060.98	24115.95	24619.51	23859.38
AMSE	548963629	754299143	144487937	144552210	54491744	64321232
ARB	0.9618719	0.942419	0.059321	0.029888	0.09293	0.034675
ACV	0.9618719	0.917871	0.308376	0.308372	0.20121	0.21209
\bar{Y}	27066.87					

Table 4. Performance of Estimators in WAVE 1 of 2020.

STATES	\bar{y}_d^1	\bar{y}_s^1	$\hat{\bar{y}}_{sr}^1$	$\hat{\bar{y}}_{sreg}^1$	$\hat{\bar{y}}_{dr}^1$	$\hat{\bar{y}}_{dreg}^1$
AVE	3591.126	4090.677	71765.07	71920.58	65215.42	63894.74
AMSE	12034853638	4136670273	872370420	876983729	370357420	375904744
ARB	0.9691173	0.9265719	0.1668074	0.1927824	0.2211566	0.02324393
ACV	0.9691173	0.9265719	0.3626012	0.3640058	0.226493	0.2387889
\bar{Y}	68125.18					

Table 5. Performance of Estimators in WAVE 2 of 2020.

STATES	\bar{y}_d^2	\bar{y}_s^2	$\hat{\bar{y}}_{sr}^2$	$\hat{\bar{y}}_{sreg}^2$	$\hat{\bar{y}}_{dr}^2$	$\hat{\bar{y}}_{dreg}^2$
AVE	2877.962	3099.407	59054.76	59075.77	56846.09	55346.86
AMSE	4245321488	450340476502	438061756359	438061811658	438812099301	438901849839
ARB	0.9504211	0.9435805	0.003031283	0.01510526	0.07992032	0.05786398
ACV	0.9504211	0.9435805	0.3335228	0.3336408	0.1694406	0.1621179

STATES	\bar{y}_d^2	\bar{y}_s^2	$\hat{\bar{y}}_{sr}^2$	$\hat{\bar{y}}_{sreg}^2$	$\hat{\bar{y}}_{dr}^2$	$\hat{\bar{y}}_{dreg}^2$
\bar{Y}	61961.9					

Table 6. Performance of Estimators in WAVE 3 of 2020.

STATES	\bar{y}_d^3	\bar{y}_s^3	$\hat{\bar{y}}_{sr}^3$	$\hat{\bar{y}}_{sreg}^3$	$\hat{\bar{y}}_{dr}^3$	$\hat{\bar{y}}_{dreg}^3$
AVE	3457.075	3622.027	59961.07	59279.41	67578.26	65086.97
AMSE	5044134292	4213740517	647710500	644994313	252824732	213196627
ARB	0.9526255	0.9437981	0.05997906	0.05852164	0.03490459	0.01047999
ACV	0.9526255	0.9201944	0.276252	0.2749713	0.1820564	0.1674836
\bar{Y}	66306.01					

Table 7. Performance of Estimators in WAVE 4 of 2020.

STATES	\bar{y}_d^4	\bar{y}_s^4	$\hat{\bar{y}}_{sr}^4$	$\hat{\bar{y}}_{sreg}^4$	$\hat{\bar{y}}_{dr}^4$	$\hat{\bar{y}}_{dreg}^4$
AVE	3540.935	3731.182	64751.25	64518.61	71761.71	68599.05
AMSE	4671464540	5017727857	776371953	775853458	234529499	255142698
ARB	0.9562904	0.943985	0.01867075	0.008393917	0.08672708	0.01442189
ACV	0.9562904	0.9189337	0.2930774	0.2919856	0.150288	0.1563659
\bar{Y}	69369.19					

Table 8. Performance of Estimators in WAVE 2 of 2021.

STATES	\bar{y}_d^2	\bar{y}_s^2	$\hat{\bar{y}}_{sr}^2$	$\hat{\bar{y}}_{sreg}^2$	$\hat{\bar{y}}_{dr}^2$	$\hat{\bar{y}}_{dreg}^2$
AVE	2039.53	6386.783	123297.2	123025.9	131436.6	123849.2
AMSE	14198268143	12994022704	2869261614	2854990305	1861227650	1838743398
ARB	0.9899013	0.9422168	0.06667438	0.08294576	0.0474055	0.06375305
ACV	0.98172353	0.9422168	0.3418827	0.341164	0.2841045	0.311278
\bar{Y}	116682.6					

Table 9. Performance of Estimators in WAVE 3 of 2021.

STATES	\bar{y}_d^3	\bar{y}_s^3	$\hat{\bar{y}}_{sr}^3$	$\hat{\bar{y}}_{sreg}^3$	$\hat{\bar{y}}_{dr}^3$	$\hat{\bar{y}}_{dreg}^3$
AVE	2026.005	5623.038	105641.4	105652	117592.9	113097.5
AMSE	11486839433	11270120207	2645228863	2645580100	1043052037	1025665996
ARB	0.9801287	0.94534	0.03357499	0.04656683	0.06105538	0.07066468
ACV	0.9801287	0.9201338	0.394137	0.3941594	0.2422864	0.2320182

STATES	\bar{y}_d^3	\bar{y}_s^3	\hat{y}_{sr}^3	\hat{y}_{sreg}^3	\hat{y}_{dr}^3	\hat{y}_{dreg}^3
\bar{Y}	106581.5					

Table 10. Performance of Estimators in WAVE 4 of 2021.

STATES	\bar{y}_d^4	\bar{y}_s^4	\hat{y}_{sr}^4	\hat{y}_{sreg}^4	\hat{y}_{dr}^4	\hat{y}_{dreg}^4
AVE	1127.653	3663.538	62406.66	62619.98	67594.14	67302.54
AMSE	5567310780	4500549019	790417659	790608619	208194739	258796222
ARB	0.9827764	0.9420642	0.03440734	0.07937525	0.08019187	0.01224156
ACV	0.9827764	0.9178592	0.3701064	0.3716396	0.1887388	0.1903148
\bar{Y}	66540.06					

5. Discussion

Considering the result of analysis in Tables 1, 2 and 3, the population mean expenditure of households stands at 39,282.73k in WAVE 1, 40,545.98k in WAVE 3 and 27,066.87k in WAVE 4 of 2019. As specified in the result, the modified synthetic estimators \hat{y}_{sr}^t and \hat{y}_{sreg}^t and the calibrated direct estimators \hat{y}_{dr}^t and \hat{y}_{dreg}^t outperformed the existing synthetic and direct estimators \bar{y}_s^t and \bar{y}_d^t respectively with smaller average mean square error (AMSE), average relative bias (ARB) and average coefficient of variation (ACV). However, there is no clear cut consistency in the performance of the the modified synthetic and calibrated direct estimators in WAVE 1, 3 & 4 of 2019, though the later had higher AMSE than the former, the calibrated direct estimators were less biased with favorable ACV than the synthetic estimators. More so, the calibrated direct ratio estimator \hat{y}_{dr}^t had a favorable ACV of 23%, and 20% in WAVE 3 and 4 respectively while the calibrated direct regression estimator \hat{y}_{dreg}^t had ACV of 25%, and 21% in WAVE 3, and 4 respectively. This performance was better, favorable and preferred for small area estimation than that of the modified synthetic estimators whose ACV recorded 36% and 31% in WAVE 3 and 4 respectively, and is considered not appropriate for small area estimation.

The results as shown in Tables 4, 5, 6 and 7, show that the population mean expenditure of households in Nigeria stands at 68,125.18k in WAVE 1, 61,961.90k in WAVE 2, 66,306.01k in WAVE 3 and 69,369.19k in WAVE 4 of 2020. Furthermore, the Tables recorded that the modified synthetic estimators and the calibrated direct estimators performed better than the existing synthetic and direct estimators with smaller AMSE, negligible ARB and a favorable ACV. However, although modified synthetic estimators showed dominance in WAVE 2 with smaller average mean square

error and negligible bias, the calibrated direct estimators performed better than the modified synthetic estimators in terms of smaller AMSE and less ARB in WAVE 1, 3 & 4 of 2020. In addition, the calibrated direct estimators had a favorable ACV of 23%, 17%, 18%, and 15% in WAVE 1, 2, 3, and 4 respectively for the calibrated direct ratio estimator \hat{y}_{dr}^t and ACV of 23%, 16%, 17%, and 16% in WAVE 1, 2, 3, and 4 respectively for the calibrated direct regression estimator \hat{y}_{dreg}^t which is considered to be appropriate for small area estimation. This performance agrees with the literature by [17] that the benchmark for the percentage average coefficient of variation for small area estimation is 25%. Judging from this, the modified synthetic estimators could not be preferred for estimation in WAVE 1, 2, 3, and 4 of 2020 because of the high ACV of 36%, 33%, 28% and 29% respectively for both \hat{y}_{sr}^t and \hat{y}_{sreg}^t .

Tables 8, 9 and 10 revealed that the population mean expenditure of households stands at 116,682.60k in WAVE 2, 106,581.50k in WAVE 3 and 66,540.06k in WAVE 4 of 2021. Again, the result further affirmed the overwhelming performance of the modified synthetic estimators and the calibrated direct estimators over the existing synthetic and direct estimators with smaller AMSE, less ARB and preferable ACV. However, the calibrated direct estimators performed better than the modified synthetic estimators in terms smaller AMSE and negligible ARB in WAVE 2, 3 & 4 of 2021. In addition, the calibrated direct estimators seems more preferred and appropriate in small area estimation with ACV of 28%, 24%, and 19% in WAVE 2, 3, and 4 respectively for the calibrated direct ratio estimator \hat{y}_{dr}^t and ACV of 31%, 23% and 19% in WAVE 2, 3, and 4 respectively for the calibrated direct regression estimator \hat{y}_{dreg}^t though their ACV for WAVE 2 is a bit outside the prescribed benchmark. Judging from this, the modified synthetic estimators are not appropriate for estimation in WAVE 2, 3, and 4 of 2021 because of the high ACV of

34%, 39% and 37% respectively for both \hat{y}_{sr}^t and \hat{y}_{sreg}^t . The performance of the calibrated direct estimators agrees with the literature that when the sample size is large in any domain of interest, the direct estimator performs better than the synthetic estimator. This result further strengthens the result by [18] on the use of auxiliary variable and that of [19] on the usefulness of calibration estimation technique.

6. Conclusion

This work was conceived to establish the pattern of household consumption-expenditure in Nigeria before, during and after COVID-19 pandemic using the existing synthetic calibration estimators and the suggested direct calibration estimators for small areas. It is observed that the suggested calibration estimators have provided a more reliable estimates to compensate against the instability of the synthetic estimators and the higher variance of the direct estimators. In fact, the calibrated direct ratio/regression estimators have shown prominence in performance over the synthetic ratio/regression estimators. This establishes the fact that when there are reasonable number of observations in a domain of study, it is appropriate and profitable to use the calibrated direct estimators for a reliable and more precise estimates unlike the synthetic estimators whose strength lies in borrowing of strength from other domains with the same structural characteristics. As seen from Tables 1-10, the performance of the existing synthetic estimator has been very poor in estimating the population mean probably due to lack of auxiliary variable in the formulation of the estimators. However, the calibrated synthetic ratio/regression estimators, though could provide a reasonable estimates of the population mean, fail short of the conditions for estimation in small area because they produce coefficient of variation (CV) higher than 25% as stipulated by Molina and Rao (2010). Consequently, the gains made on the performance of the suggested calibrated direct ratio/regression estimators cannot be overemphasized. The performance is outstanding in the average mean estimates, AMSE, ARB and ACV of the population mean expenditure across all the waves of 2019, 2020 and 2021. This is an indication to the fact that auxiliary variable (income) incorporated into the existing estimator through calibration techniques by Iseh and Enang (2021) has been very fruitful. This study has made a practical illustration to confirm the submission by Iseh and Enang (2021) whose synthetic estimator provided estimates in areas of interest with small/no sample data and was unstable when the sample size is large while the calibrated direct estimator performs better with direct observation when the sample size is large. In addition, as seen from the results, in 2019 before COVID-19 pandemic, the average population expenditure of households in WAVE 1, 3 and 4 was below 41,000.00k and could be attributed to more rural respondents participation. In 2020, during COVID-19 the average population expenditure of households rose from 68,125.18k naira in WAVE 1, to 69,369.19k in WAVE 4, which suggests that

households consumed more while they were at home. Again, in 2021, after the COVID-19 pandemic era, the average population expenditure of households rose to 116,682.60k in WAVE 2 and declines a bit to 106,581.50k in WAVE 3 and falls back to 66,540.06k in WAVE 4. This pattern signifies that households consumed more during COVID-19 period while at home and the burden lessens after the pandemic. The period under study indicates an increase in population mean expenditure of both rural and urban areas of Nigeria during COVID-19 pandemic. However, a decrease in the average expenditure was noticed in WAVE 4.

Based on the theoretical and empirical validations of this study, it suffices to recommend calibration techniques should be adopted to enhance efficiency in existing estimators that has no auxiliary variable (s) and when surveys of this nature are carried out, there should be equal number of rural and urban households' participation. Also, since this study has established the use of auxiliary variable that is strongly correlated with the study variable in domain estimation where there is small/no sample data in areas of interest, it suffices to say that when study of this nature are carried out, auxiliary variable that is strongly correlated with the study variable should be used. In addition, further study is required in the future for a combination of the two (direct and synthetic) calibrated estimators for optimal result in some domains of interest.

Conflicts of Interest

The author declares no conflicts of interest.

References

- [1] Sarndal, C. E., Swensson, B. and Wretman, J. (1992). Model assisted survey sampling, New York, Springer-Verlag.
- [2] Rao, J. N. K. (2003). Small Area Estimation, Wiley, New Jersey, US.
- [3] Lahiri, P. (2003). On the Impact of Bootstrap in Survey Sampling and Small Area Estimation. *Statistical Science*, 18(2): 199-210.
- [4] Lumley, T. and Scott, A. (2017). Fitting regression models to survey data, *Statistical Science*, 32, 265–278.
- [5] Rao, J. N. K. (2005). Inferential issues in small area estimation: Some new developments. *Statistics in Transition* 7 513–526.
- [6] Gonzalez, M. E. (1973). 'Use and Evaluation of synthetic Estimates ', *Proceedings of the Social Statistics Section, American Statistical Association, USA*, 33-36.
- [7] Marker, D. A. (1999). Organization of small area estimators using generalized linear regression framework. *Journal of Official Statistics*, 15, 1-24.
- [8] Hidiogrou, M. A. and Patak, Z. (2004). Small area Estimation: Theory and Practice. *Statistical Innovation and Research Division*, 16 floor Section D. R. H. Coats Building. Tunney's Pasture, Ottawa, Ontario, KIA OT6, Canada.

- [9] Singh, B. and Sisodia, B. V. S. (2011). Small area estimation in longitudinal surveys. *J. of Rel. and Stat. Stu.*, 4(2): 83-91.
- [10] Iseh, M. J. and Enang, E. I. (2021). A calibration synthetic estimator of population mean in small area under stratified sampling design. *Transition in Statistics new series*. Vol. 22(3): 15–30. <https://doi.org/10.21307/stattrans-2021-025>
- [11] Ghosh, M. and Rao, J. N. K. (1994). Small area estimation: An appraisal with discussion, *Statistical Science*, 9, 55–76.
- [12] Lahiri, P. and Meza, J. (2002). *Encyclopedia of Environments*, volume 4, Chapter Small Area Estimation, pages 2010-2014. Wiley.
- [13] Pfeffermann, D. (2002). Small area estimation—new developments and directions. *International Statistical Review* 70 125– 143.
- [14] Hidirolou, M. A. and Patak, Z. (2006). Domain Estimation using Linear Regression. *Survey Methodology*, 30(1): 67-78.
- [15] Sarndal, C. E., Swensson, B. and Wretman, J. (2003). *Model Assisted Survey Sampling*, Springer, New York, US.
- [16] Pandey, K. K. and Tikkiwal, G. C. (2007). The Generalized Class of Composite Method of Estimation for Crop Acreage in Small Domain, *International Journal of Statistics and Systems* 5: 319–33.
- [17] Molina, I., Rao, J. N. K., (2010). Small Area Estimation Poverty Indicators, *Canadian Journal of statistics*, Vol. 38: 369–385.
- [18] Khare, B. B., Ashutosh, and Rai, P. K. (2021). A Comparative Study of a Class of Direct Estimators for Domain Mean with a Direct Ratio Estimator for Domain Mean using Auxiliary Character, *Statistics In Transition new series*, 2021; 22: 189–200, <https://doi.org/10.21307/stattrans>
- [19] Ikot, E. E. and Iseh, M. J. (2024). Calibration for efficiency of ratio estimator in domains of study with sub-sampling the nonrespondents. *Biom Biostat int j.* 2024; 42-50. <https://doi.org/10.15406/bbij.13.00413>
- [20] Avila, J. L., Huerta, M., Leiva, V., Riquelme, M. and Trujillo, L. (2020). The Fay–Herriot model in small area estimation: EM algorithm and application to official data, *REVSTAT*, 18(5), 613–635.
- [21] Särndal, C. E. (1984). “Design-consistent versus Model-dependent Estimation for Small Domains.” *Journal of the American Statistical Association* 79 (387): 624-631. <https://doi.org/10.2307/2288409>
- [22] Iseh, M. J. and Bassey, K. J. (2021a). A New Calibration Estimator of Population Mean for Small Area with Nonresponse. *Asian Journal of Probability and Statistics*, Vol. 12: 14-51.
- [23] Iseh, M. J. and Bassey, K. J. (2021b). Calibration estimator for population mean in small sample size with Non-response. *European Journal of Statistics and Probability*, Vol. 9(1): 32-42.
- [24] Iseh, M. J and Bassey M. O. (2024). Smoothing of Estimators of Population mean using Calibration Technique with Sample Errors. *Journal of Modern Applied Statistical Methods*. Vol. 23(1), 1-20 <https://doi.org/10.56801/Jmasm.V23.i1.2>