Hindawi
Journal of Mathematics
Volume 2022, Article ID 4530489, 10 pages
https://doi.org/10.1155/2022/4530489

Research Article

On Quantiles Estimation Based on Stratified Sampling Using Multiplicative Bias Correction Approach

Nicholas Makumi,[1,2] **Romanus Odhiambo Otieno**,[3] **George Otieno Orwa**,[2] **and Alexis Habineza**[1]

[1]*Pan African University, Institute for Basic Sciences, Technology and Innovation (PAUSTI), Nairobi, Kenya*
[2]*Department of Statistics and Actuarial Sciences, JKUAT, Nairobi, Kenya*
[3]*Meru University of Science and Technology, Meru, Kenya*

Correspondence should be addressed to Nicholas Makumi; nicholas.makumi@jkuat.ac.ke, Romanus Odhiambo Otieno; rodhiambo@must.ac.ke, George Otieno Orwa; gorwa@jkuat.ac.ke, and Alexis Habineza; alexhabk87@gmail.com

Received 1 March 2022; Accepted 7 April 2022; Published 13 May 2022

Academic Editor: A. Ghareeb

Copyright © 2022 Nicholas Makumi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the context of stratified sampling, we develop a nonparametric regression technique to estimating finite population quantiles in model-based frameworks using a multiplicative bias correction strategy. Furthermore, the proposed estimator's asymptotic behavior is presented, and when certain conditions are met, the estimator is observed to be asymptotically unbiased and asymptotically consistent. Simulation studies were conducted to determine the proposed estimator's performance for the three quartiles of two fictitious populations under varied distributional assumptions. Based on relative biases, mean-squared errors, and relative root-mean-squared errors, the proposed estimator can be extremely satisfactory, according to these findings.

1. Introduction

Many of the activities conducted by official statistics institutes are based on surveys conducted on a finite population employing stratified random sampling with no replacement. According to Thompson [1], stratified simple random sampling is described as follows: the population is subdivided into various mutually exclusive and exhaustive subgroups or strata, each of which denotes a known portion of the entire population. The researcher selects certain instances from each stratum into the sample using random sampling [2], and the results of these distinct samples are merged appropriately to yield an estimate of some specified population parameter. These surveys gather information on three categories of variables: binary variables, categorical variables with more than two modalities, and continuous quantitative variables. When numerous homogeneous and mutually exclusive strata or subpopulations are found in a population, stratified sampling is an appropriate strategy. Stratification can help to increase sample representativeness by minimizing sampling error. The greater the difference between the strata, the greater the accuracy gain. Furthermore, certain strata may be smaller in size but significant in the study. In these circumstances, thorough sampling is advised, which means that all individuals from these strata will be included in the sample.

In sample surveys connected to agriculture, markets, industries, and social research, for example, multiple characteristics are typically observed out of each selected unit of population. For economic and efficiency reasons, stratified random sampling is preferable to alternative survey designs for gathering information from a heterogeneous population. The stratified sampling principle is only associated with desirable properties of estimators developed from stratified simple random samples, as well as the best (optimal) sample size to be selected from various strata, to either maximize the accuracy of designed estimators for a fixed amount or to reduce survey fees for a fixed specificity of estimators. When several features are discerned from each chosen unit of a finite population in stratified sampling, the

sample allocation dilemma gets much more challenging. An allocation that is optimal for one attribute may not be optimal for others unless the features are significantly related.

Scientists are frequently really into estimating cumulative distribution functions from analytical survey data. Sedransk and Sedransk [3] studied the suitability of using estimated cumulative distribution functions to compare patient treatment at radiation therapy centers using a huge nationwide survey of cancer patient medical data. Functions of the cumulative distribution function, such as quantiles and the interquartile range, are also of relevance. The Bureau of Labor Statistics, for example, publishes median salaries for wage and salary workers in the periodical news on a regular basis. The medians are derived from a stratified multistage subsample of the Current Population Survey. Despite the fact that large-scale surveys almost always use some sort of stratified cluster sampling, much of the research on quantile estimation for finite populations is limited to simple random sampling or stratified random sampling.

Simple random sampling (SRS) is widely utilized only when variables' values do not really change significantly, and the population is homogeneous. SRS is among the most basic sampling procedures in many ways, and no further information is required. Furthermore, when using SRS to create a sample, sample weights are not really required for evaluating data from a survey using, for example, regression or multivariate analysis. A downside of SRS is the complexity in managing accuracy and the inefficiencies of not using supplemental data, which could result in enormous samples that are unneeded. Furthermore, because no supplementary information is used, there is always the potential of a skewed sample.

Stratification is widely used to enhance the accuracy of estimates and to ensure that the sample within a survey region is sufficiently distributed through subpopulations. Sometimes, it is a characteristic of designs used in soil surveys and research in soil science. The population (e.g., the survey area) is divided into classes that are mutually exclusive or strata that divide the population into survey area categories. In each stratum, samples are selected independently. More reliable estimators can be obtained when the variance within each stratum of the feature of interest is small compared to the variation within strata. In addition, if subpopulations of interest are identified by strata, an allocation scheme can be implemented to ensure that a sufficient number of sample units for making inferences on these subpopulations are located within each stratum.

The benefit of stratified sampling is that the accuracy can be determined in each stratum. Furthermore, practical features of response, measurement, and auxiliary information may change from one subgroup to the next, and this information can help stratify the population and increase efficiency. Geographic territories can be utilized as various geographical strata for administrative purposes.

The simple random sampling (SRS) design is the most commonly used in the literature. To acquire a representative sample of the population of interest, a more organized sampling approach, such as stratified sampling or systematic sampling, might be used in practice. In many agricultural and environmental studies, as well as more recently in human populations and reliability analyses (see, e.g., Samawi and Al-Sagheer [4]), the actual measurement of a sampling unit might be more expensive than its physical acquisition. As a result, when all available sampling units contribute to the selection process but just a small fraction (experimental units) is used for actual quantification, significant cost reductions can be gained in survey sampling and experimental research. The stratified simple random sampling (SSRS) approach can be used to accomplish this.

In the literature, much emphasis has been placed on the p th quantile estimation problem. The cdf estimator is required to estimate the p th quantile, according to the definition. Although the empirical distribution function is the most well-known nonparametric estimation for the cdf, it is a step function and thus insufficiently smooth.

Majority of contributions in literature use simple random sampling (SRS) to estimate the α th quantile utilizing kernel density function; for information we direct the reader see for example Nadaraya [5], Lio and Padgett [6], Jones [7]. Furthermore, some studies addressed the estimation of the pth quantile using the SSRS scheme. For example, Samawi et al. [8] developed an estimation technique for population quantiles predicted on stratified simple random sampling (SSRS) as well as stratified ranked set sampling (SRSS) using the empirical distribution function of a stratified population, and Eftekharian and Samawi [9] recently introduced kernel-based estimators of population quantiles based on SSRS and SRSS.

Kernel estimating methods have long claimed that the smoothing bandwidth of the kernel determines the effectiveness of the method more than the choice of kernel. The vast majority of kernels used are symmetric kernels that are preconfigured. This method may be beneficial for estimating boundless support curves, but it is ineffective for compact support curves with discontinuous border points. For curves of this sort, a set kernel shape causes a boundary bias. The weight allocation of the fixed symmetric kernel outside the distribution support generates this boundary bias when smoothing close to the border.

Boundary bias is a widely known challenge, and various researchers have offered methods to minimize it. The reader is recommended to [10–16]. In this study, we present a nonparametric estimator for the quantile function of a finite population predicted on SSRS, addressing the problem of boundary bias in quantile estimation using a multiplicative bias-corrected technique described in [17]. This method has two distinguishing characteristics. One is that it ensures a precise estimate, and finally, it reduces estimate bias while increasing variance by a negligible amount.

2. Notation and Basic Concepts

Let U_N denote finite population subdivided into L subgroups, with N_h being the known number of units within every stratum and that $N = \sum_{h-1}^{L} N_h < \infty$:

$$\begin{aligned} U_N &= \{u_{hi}\colon h = 1, \ldots, L, i = 1, \ldots, N_h\} \\ &= \{U_{N_1}, \ldots, U_{N_L}\}, \end{aligned} \tag{1}$$

where $U_{N_h} = u_{h1}, \ldots, u_{hN_h}$ for $h = 1, \ldots, L$. Assume that every unit in $U_N = \cup_{h-1}^{L} U_{N_h}$ is linked with a unique value of the feature, Y. The numbering of the items in each stratum, U_{N_h}, is considered to be independent of Y. For $h = 1, \ldots, L$ and $i = 1, \ldots, N_h$, let Y_{hi} signify the value of Y associated with unit u_{hi-}. Let F_N be the Y distribution function in the U_N population:

$$F_N(t) = N^{-1} \sum_{h=1}^{L} \sum_{h=1}^{N_h} I\left(Y_{hi} \le t\right), \tag{2}$$

where

$$I\left(Y_{hi} \le t\right) = \begin{cases} 1, & \text{if } Y_{hi} \le t, \\ 0, & \text{if elsewhere.} \end{cases} \tag{3}$$

Conversely, the distribution function of Y in every stratum can be used to define $F_N(t)$:

$$F_N(t) = \frac{1}{N} \sum_{h=1}^{L} N_h F_{N_h}(t), \tag{4}$$

where

$$F_{N_h}(t) = N_h^{-1} \sum_{h=1}^{N_h} I\left(Y_{hi} \le t\right). \tag{5}$$

The population's αth quantile of Y is described as

$$Q(\alpha) = \inf\{t: F_N(t) > \alpha\}, \tag{6}$$

where α takes the values between 0 and 1. Simple random samples of predetermined size $n_1, \ldots, n_L$ are taken from the L stratum interdependently with no replacement. Let y_{hi} denote the values of the characteristic Y corresponding with the units in the sample from stratum $h\,(h = 1, \ldots, L)$ for $i = 1, \ldots, n_h$. For the combined sample of $n = \sum_{h-1}^{L} n_h$, the weighted empirical distribution function is expressed by

$$F_n(t) = N^{-1} \sum_{h=1}^{L} \sum_{h=1}^{n_h} w_h I\left(y_{hi} \le t\right), \tag{7}$$

in which the weight designated unit hi is $w_h = N_h (n_h N)^{-1} = W_h n_h^{-1}, W_h = N_h N^{-1}$, and

$$I\left(y_{hi} \le t\right) = \begin{cases} 1, & \text{if } y_{hi} \le t, \\ 0, & \text{if } y_{hi} > t. \end{cases} \tag{8}$$

The weight, w_h, is inversely proportional to the likelihood that i^{th} unit in stratum h will be included in the sample. It is also possible to write the weighted empirical distribution function as

$$F_n(t) = \frac{1}{N} \sum_{h=1}^{L} N_h F_{n_h}(t), \tag{9}$$

where

$$F_{n_h}(t) = n_h^{-1} \sum_{h=1}^{n_h} I\left(y_{hi} \le t\right). \tag{10}$$

The α sample quantile is denoted as

$$Q(\alpha) = \inf\{t: F_n(t) > \alpha\}. \tag{11}$$

The following assumptions were considered:

(i) The underlying population of the hth stratum has the cdf as $F_h(\cdot)$ which is Hölder continuous with a square-integrable second derivative for any $h = 1, \ldots, L$.

(ii) $K_h(\cdot)$ is an absolutely continuous function, such that $\lim_{z\_-a} K_h(x) = 0$ and $\lim_{x \longrightarrow a} K_h(x) = 1$, $h = 1, 2, \ldots, L$.

(iii) The kernel function $k_h(x)$ satisfies the following conditions for any x

$$\begin{aligned} k_h(x) &= k_h(-x), \\ \int_{-a}^{a} k_h(x)\mathrm{d}x &= 1, \\ \int_{-a}^{a} x^2 k_h(x)\mathrm{d}x &\neq 0. \end{aligned} \tag{12}$$

3. MBC Quantile Estimation for SSRS

In this section, we describe the multiplicative bias correction distribution function based on SSRS which was proposed by Onsongo et al. [18] and later use it to introduce quantile estimator along with its asymptotic properties. For the h^{th} stratum, let $X_{hi}, i = 1, 2, \ldots, N_h$ be the auxiliary variable with associated survey measurement $Y_{hi}, i = 1, 2, \ldots, N_h$ from a predominant univariate distribution function. Suppose a simple random sample of size n_h is taken from stratum h^{th} without replacement, with the sample fraction $f_h = n_h/N_h \longrightarrow 0$ as $n_h \longrightarrow N_h$ as well as $N_h \longrightarrow \infty$. Accordingly, for a finite population, the empirical distribution function is therefore defined as

$$F_N(t) = \frac{1}{N} \sum_{i=1}^{N} \Delta\left(t - y_i\right). \tag{13}$$

For a stratified population, the associated estimator of a distribution function is defined as

$$\begin{aligned} \widehat{F}_{Ny}^{S}(t) &= \frac{1}{N} \sum_{i=1}^{H} N_h \left(\frac{1}{N_h} \sum_{i=1}^{N_h} \Delta\left(t - y_{hi}\right) \right) \\ &= \frac{1}{N} \sum_{i=1}^{H} N_h F_{hy}(t), \end{aligned} \tag{14}$$

in which Δ is perhaps step function of a particular set, t, α-quantile, and i represent measurements taken from h^{th} stratum. $F_{hy}(t)$ is the h^{th} stratum distribution function for the random variable Y. Suppose s represents a set of data n_h units selected from h^{th} stratum using simple random sampling with no replacement, $j \in r \in h = (N_h - s)$ represent nonsampled units in h^{th} stratum. Suppose that survey

variables were generated with the help of a super population model, which is represented by

$$y_{hi} = m(x_{hi}) + \sigma(x_{hi})\epsilon_{hi}, \tag{15}$$

where $\varepsilon_{hi} \sim ii\ dN(0, \sigma^2(x_{hi}))$ and

$$\mathrm{Cov}(y_{hi}, y_{hj}) = \begin{cases} \sigma^2(x_{hi}), & \text{if } i = 1, 2, \ldots, N, \\ 0, & \text{elsewhere.} \end{cases} \tag{16}$$

As a result, the predicted form of the empirical distribution function for a stratified population is obtained using the model-based technique.

$$F^S_{N_y}(t) = \frac{1}{N}\sum_{h=1}^{L} N_h \left(\frac{1}{N_h} \left[\sum_{i\epsilon s\epsilon h} I(y_{hi} \le t) + \sum_{j\epsilon r\epsilon h} I(y_{hj} \le t) \right] \right). \tag{17}$$

The second term of equation (17) is not known, and the concern is determining how to accurately estimate it. Onsongo et al. [18] suggested a multiplicative bias-corrected estimator for finite population distribution under stratified sampling to estimate equation (17).

$$\widehat{F}^S_{\mathrm{MBC}}(t) = \frac{1}{N}\sum_{h=1}^{L} N_h \left(\frac{1}{N_h} \left[\sum_{j\epsilon s\epsilon h} I(y_{hi} \le t) + \sum_{j\epsilon r\epsilon h} \widehat{G}\left(t - \widehat{\mu}_{n_h}(x_{hj})\right) \right] \right), \tag{18}$$

where the term $\widehat{\mu}(x_{hj})$ represents nonparametric estimator under model-based technique for $\mu(x_{hj})$ and $\widehat{G}(t - \widehat{\mu}(x_{hj}))$ denotes residual estimated distribution function, where residuals are given by $e_{hj} = y_{hj} - \widehat{\mu}(x_{hj})$ for h th stratum. According to Onsongo et al. [18], $\widehat{F}^S_{\mathrm{MBC}}(t)$ leads to an unbiased estimator for $F_N(t)$ and variance is expressed as

$$\mathrm{Var}\left[\widehat{F}^{(S)}_{\mathrm{MBC}}(t) - F_N(t)\right] =$$

$$\frac{1}{N^2}\left[\sum_{i\epsilon s} \left\{ \sum_{j-1}^{N_h - n_h} \sum_{k-1}^{N_h - n_h} w^*_{ij} w^*_{ik} \begin{bmatrix} D_i\left(t - \max\left(\widehat{\eta}_{hj}, \widehat{\eta}_{hk}\right)\right) \\ -D_i\left(t - \widehat{\eta}_{hj}\right) D_i\left(t - \widehat{\eta}_{hk}\right) \end{bmatrix} \right\} + (N_h - n_h) P\left(y_{hj} \le t\right)\left[1 - P\left(y_{hj} \le t\right)\right] \right]. \tag{19}$$

We have that $\widehat{F}^{(S)}_{\mathrm{MBC}}(t) \longrightarrow^P F_N(t)$. Hence, the SSRS MBC-based estimator of cdf, $\widehat{F}^{(S)}_{\mathrm{MBC}}(\cdot)$, can be considered for estimating quantile function. As with $0 < \alpha < 1$, the α th quantile of the underlying distribution $F_N(\cdot)$ is defined as follows:

$$Q^S_{N_y}(\alpha) = \inf\{t: F_N(t) \ge \alpha\}, \tag{20}$$

and is alternatively denoted by $F^{-1}_N(\alpha)$. Based on a sample from SSRS with size n and using an approach similar to that used by Eftekharian and Samawi [9], it is immediate that an MBC estimator of the α th quantile is proposed as

$$\widehat{Q}^S_{\mathrm{MBC}}(\alpha) = \inf\left\{t \in U: \widehat{F}^S_{\mathrm{MBC}}(t) > \alpha\right\} = \widehat{F}^{(S)-1}_{\mathrm{MBC}}(\alpha), \tag{21}$$

where α is an index taking values between $1/N$ and $(N-1)/N$. That is, $\widehat{Q}^S_{\mathrm{MBC}}(\alpha)$ is the smallest value of t for which at least $100\alpha\%$ of the population y_i values are less than or equal to that value. Furthermore, from (18), $\widehat{Q}^S_{\mathrm{MBC}}(\alpha)$ can be computed by solving $\widehat{F}^S_{\mathrm{MBC}}(\widehat{Q}^S_{\mathrm{MBC}}(\alpha)) = \alpha$. However, under assumption (i), it can be easily seen that $\widehat{F}^S_{\mathrm{MBC}}(\widehat{Q}^S_{\mathrm{MBC}}(\alpha))$ is twice differentiable at $\widehat{Q}^S_{\mathrm{MBC}}(\alpha)$.

Now, assume that n be proportionally allocation into L strata, then using Taylor series expansion of the function $F^S_{\mathrm{MBC}}(Q^S_{N_y}(\alpha))$ around $\widehat{Q}^S_{\mathrm{MBC}}(\alpha)$, we can write

$$\widehat{F}^S_{\mathrm{MBC}}\left(Q^S_{N_y}(\alpha)\right) = \alpha + \left[\widehat{Q}^S_{\mathrm{MBC}}(\alpha) - Q^S_{N_y}(\alpha)\right] f^S_{\mathrm{MBC}}\left(Q^S_{N_y}(\alpha)\right) + R_n, \tag{22}$$

where $F^S_{\mathrm{MBC}} = f^S_{\mathrm{MBC}} > 0$ and $R_n = O(n^{-1/2})$ [19] become negligible as $n \longrightarrow \infty$. From equation (22), Bahadur's representation [20] of the estimator, $\widehat{Q}^S_{\mathrm{MBC}}(\alpha)$, is given by

$$\widehat{Q}^S_{\mathrm{MBC}}(\alpha) = Q^S_{N_y}(\alpha) + \frac{\alpha - \widehat{F}^S_{\mathrm{MBC}}\left(Q^S_{N_y}(\alpha)\right)}{f^S_{\mathrm{MBC}}\left(Q^S_{N_y}(\alpha)\right)} + R_n. \tag{23}$$

As with $0 < \alpha < 1$, the proportion of individuals in the population that are less than or equal to the population quantile is as follows:

$$n\widehat{F}^S_{\mathrm{MBC}}\left(Q^S_{N_y^{(\alpha)}}\right) = \sum_{i=1}^{n} z_i, \tag{24}$$

where $z_i = I\left\{y_i \le Q^S_{N_y}(\alpha)\right\}$ follows a hypergeometric distribution with the parameters N, $N\widehat{F}^S_{\mathrm{MBC}}(Q^S_{N_y}(\alpha))$, and n. Using findings from Francisco [21], the estimator's expectation and variance are computed as

$$E\left[\widehat{F}^S_{\mathrm{MBC}}\left(Q^S_{N_y^{(\alpha)}}\right)\right] = Z_N = N^{-1}\sum_{i=1}^{N} Z_i = F^S_{N_y}\left(Q^S_{N_y}(\alpha)\right) = \alpha, \tag{25}$$

$$\mathrm{Var}\left[\widehat{F}^S_{\mathrm{MBC}}\left(Q^S_{N_y}(\alpha)\right)\right] = (1-f)(n-1)^{-1}\alpha(1-\alpha). \tag{26}$$

4. Properties of the Proposed Estimator

4.1. Asymptotic Unbiasedness of the Proposed Estimator. Now, consider the bias for the nonparametric estimator $\widehat{Q}^S_{\mathrm{MBC}}(\alpha)$ defined by

$$E\left[\widehat{Q}^S_{\mathrm{MBC}}(\alpha) - Q^S_{N_y}(\alpha)\right] = E\left[\widehat{Q}^S_{\mathrm{MBC}}(\alpha)\right] - E\left[Q^S_{N_y}(\alpha)\right]. \tag{27}$$

Then, from equation (23), it follows that

$$E\left[\widehat{Q}^S_{\mathrm{MBC}}(\alpha)\right] = E\left[Q^S_{N_y}(\alpha) + \frac{\left(\alpha - \widehat{F}^S_{\mathrm{MBC}}\left(Q^S_{N_y}(\alpha)\right)\right)}{f^S_{\mathrm{MBC}}\left(Q^S_{N_y}(\alpha)\right)} + O\left(n^{-(1/2)}\right)\right]. \tag{28}$$

Using the results of equation (25), it can be easily seen that

$$E\left[\widehat{Q}^{S}_{\mathrm{MBC}}(\alpha)\right] = Q^{S}_{N_y}(\alpha) + O\left(n^{-(1/2)}\right). \tag{29}$$

Since $O(n^{-1/2})$ becomes negligible as $n \longrightarrow \infty$ [19], the right-hand side of equation (29) tends to 0, and so, $\widehat{Q}^{S}_{\mathrm{MBC}}(\alpha)$ is asymptotically unbiased.

4.2. Asymptotic Variance of the Proposed Estimator. The variance of $\widehat{Q}_{\mathrm{MBC}}(\alpha)$ will now be computed as follows. From equation (23), taking variance on both sides, we have

$$\mathrm{Var}\left[\widehat{Q}^{S}_{\mathrm{MBC}}(\alpha)\right] = \mathrm{Var}\left[Q^{S}_{N_y}(\alpha) + \frac{1}{f^{S}_{\mathrm{MBC}}\left(Q^{S}_{N_y}(\alpha)\right)}\left(\alpha - \widehat{F}^{S}_{\mathrm{MBC}}\left(Q^{S}_{N_y}(\alpha)\right)\right) + O\left(n^{-(1/2)}\right)\right]. \tag{30}$$

Applying the results of equation (26), it is immediate that

$$\begin{aligned}\mathrm{Var}\left[\widehat{Q}^{S}_{\mathrm{MBC}}(\alpha)\right] &= \left[\frac{1}{f^{S}_{\mathrm{MBC}}\left(Q^{S}_{N_y}(\alpha)\right)}\right]^2 (1-f)(n-1)^{-1}\alpha(1-\alpha) \\ &= \frac{1-f}{n-1}\alpha(1-\alpha)\left[f^{S}_{\mathrm{MBC}}\left(Q^{S}_{N_y}(\alpha)\right)\right]^{-2}.\end{aligned} \tag{31}$$

4.3. Asymptotic Mean-Squared Error. Asymptotic MSE of the estimator $\widehat{Q}^{S}_{\mathrm{MBC}}(\alpha)$ is expressed as

$$\mathrm{MSE}\left(\widehat{Q}^{S}_{\mathrm{MBC}}(\alpha)\right) = \mathrm{Var}\left(\widehat{Q}^{S}_{\mathrm{MBC}}(\alpha)\right) + \left[\mathrm{Bias}\left(\widehat{Q}^{S}_{\mathrm{MBC}}(\alpha)\right)\right]^2. \tag{32}$$

From equations (29) and (31), the following results are immediate consequences:

$$\mathrm{MSE}\left(\widehat{Q}^{S}_{\mathrm{MBC}}(\alpha)\right) = \frac{1-f}{n-1}\alpha(1-\alpha)\left[f^{S}_{\mathrm{MBC}}\left(Q^{S}_{N_y}(\alpha)\right)\right]^{-2} + O\left(\frac{1}{n}\right). \tag{33}$$

Equation (33) tends to zero as $n \longrightarrow \infty$, and thus, $\mathrm{MSE}(\widehat{Q}^{S}_{\mathrm{MBC}}(\alpha)) \longrightarrow 0$. This shows that $\widehat{Q}^{S}_{\mathrm{MBC}}(\alpha)$ is a consistent estimator of $Q^{S}_{N_y}(\alpha)$. Furthermore, $\widehat{Q}^{S}_{\mathrm{MBC}}(\alpha)$ has an asymptotic normal distribution as in Serfling [22].

$$N\left(Q^{S}_{N_y}(\alpha), \frac{1-f}{n-1}\alpha(1-\alpha)\left[f^{S}_{\mathrm{MBC}}\left(Q^{S}_{N_y}(\alpha)\right)\right]^{-2}\right). \tag{34}$$

5. Empirical Study

5.1. Description of the Population. In this part, simulation studies were carried out to investigate the performance of the proposed multiplicative bias-corrected quantile estimator for a stratified population. Two data variables, linear and cosine, were used to simulate a population of size 1000. The linear function was constructed using a linear model that has the following relationship.

$$Y_i = 1 + 2(x_i - 0.5) + e_i. \tag{35}$$

The cosine function, which has the relationship

$$Y_i = \cos\left(1 + 2(x_i - 0.5)^2\right) + e_i, \tag{36}$$

was used to get the second study variable or mean function. The supplementary variable X was considered to have a uniform distribution on a range of $[0, 1]$. The error term e_i is perceived as a standard normal variable that follows $e_i \sim N(0, 1)$.

To investigate the proposed estimator's practical performance, each of the populations (i.e. $Y_i's$) was subdivided into 5 equal, disjoint, and mutually exclusive subgroups $(N_h = 200, h = 1, 2, \ldots, 5)$, which are made as homogeneous as possible to ensure that units in each stratum vary little from each other. Thereafter, a sample of size $n = 200$ was drawn, with each stratum providing a sample size of $n_h = 40$, $(h = 1, 2, \ldots, 5)$ employing simple random sampling with no replacement for each scenario. The Epanechnikov kernel, defined by,

$$K(v) = \frac{3}{4}\left(1 - u^2\right)I_{\{|v|\le 1\}}, \tag{37}$$

was employed for kernel smoothing on the different populations.

5.2. Estimators Included in the Empirical Study. We compare the MBC quantile estimator under SSRS defined by (21) to some of the popular quantile estimators under SSRS proposed in the literature since one of our goals is to develop estimators with desirable features with respect to bias, variance, and asymptotic mean-squared error. For comparison purposes, the following estimators were used, and first, we include in our study estimator of [8] defined as

$$\widehat{\zeta}_p = \widehat{F}^{-1}_{\mathrm{SSRS}}(p), \tag{38}$$

where $\widehat{F}_{\mathrm{SSRS}}(x) = (1/n_h)\sum_{h=1}^{L}\sum_{i=1}^{n_h} W_h I(X_{hi} \le x)$. We also include in our empirical study kernel-based estimator of the quantile based on SSRS which is proposed by Eftekharian and Samawi [9].

$$\widehat{\xi}^{\mathrm{SSRS}}_p = \widehat{F}^{-1}_{\mathrm{SSRS}}(p), \tag{39}$$

where, in this case, $\widehat{F}_{\mathrm{SSRS}}(x) = (1/n_h)\sum_{h=1}^{L}\sum_{i=1}^{n_h} W_h K_h(x - X_{ha}/d_h)$ Finally, in our empirical study, we include Chambers and Clark estimator studied in [23].

$$\widehat{F}^{rob}_{Ny}(t) = \frac{1}{N}\left[\sum_{i\in s} d_{it} + \sum_{j\in r}\widehat{G}\left(t - b_v z_j\right)\right]. \tag{40}$$

Table 1: Unconditional biases, relative mean errors, and relative root-mean-squared errors.

				Linear			Cosine
α	Estimator	Bias	RME	RRMSE	Bias	RME	RRMSE
0.25	ESQE	1.4245	1.4801	0.2156	0.4288	1.4423	0.4470
	RCQE	0.3268	1.0781	0.1571	1.3966	1.0086	0.3126
	MBCQE	0.1176	0.7877	0.1148	0.10001	0.7998	0.2479
0.5	ESQE	0.5715	0.9747	0.1264	0.4929	0.8758	0.2214
	RCQE	0.9571	1.0140	0.1315	0.8559	0.9203	0.2326
	MBCQE	0.5266	0.8639	0.1121	0.4434	0.7843	0.1982
0.75	ESQE	1.0373	1.6103	0.1892	1.2112	1.5113	0.3169
	RCQE	1.6081	1.2953	0.1522	1.5092	1.4223	0.2983
	MBCQE	1.0315	1.1818	0.1388	0.7838	0.9709	0.2037

Table 2: Comparison of empirical quantile estimator with other estimators.

	Quantile estimates			
Mean functions	Estimators	0.25	0.50	0.75
	$Q(p)$	6.8636	7.7089	8.5115
Linear	ESQE	5.4392	6.7519	6.9034
	RCQE	7.1905	7.1374	7.4742
	MBCQE	6.7460	7.1823	7.4801
	$Q(p)$	3.2265	3.9561	4.7677
Cosine	ESQE	1.8298	3.1001	3.2586
	RCQE	3.6553	3.4631	3.5565
	MBCQE	3.1265	3.5127	3.9839

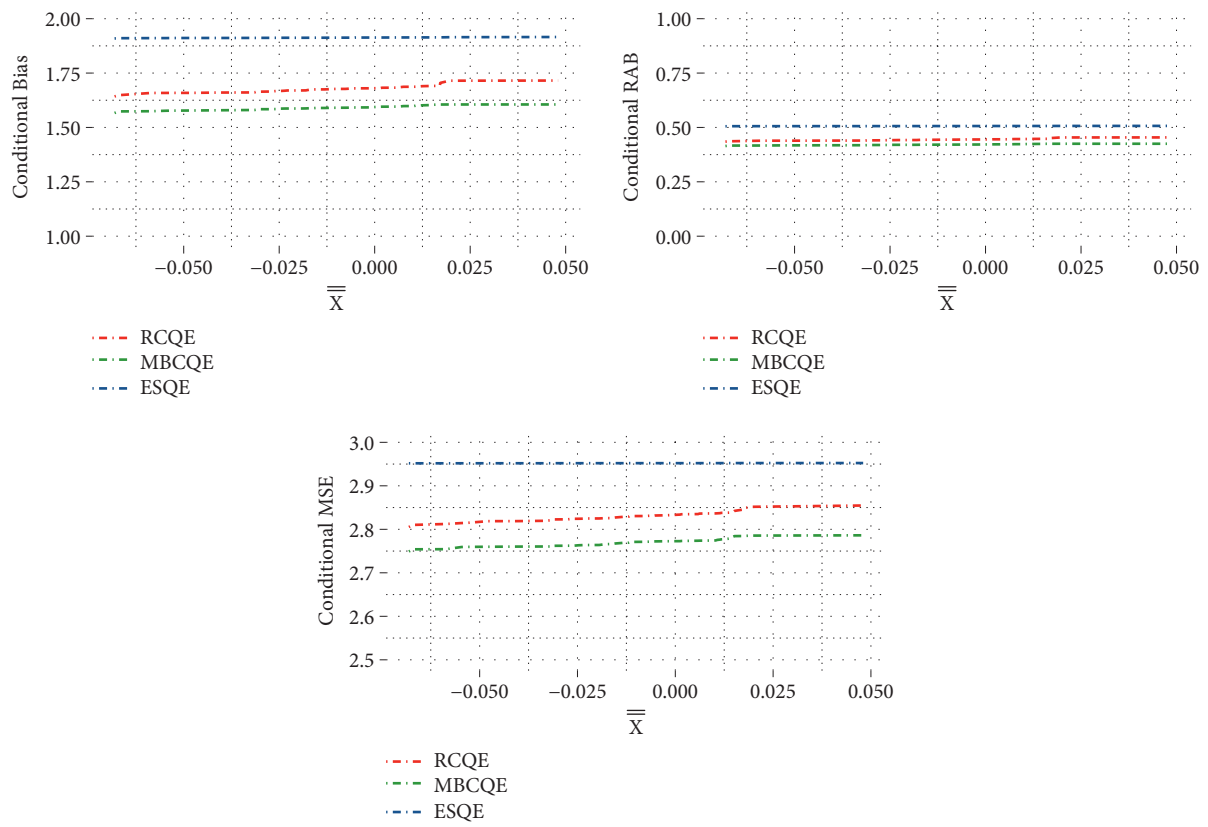

Figure 1: CB, CRAB, and CMSE for estimators: linear mean function, $\alpha = 0.25$.

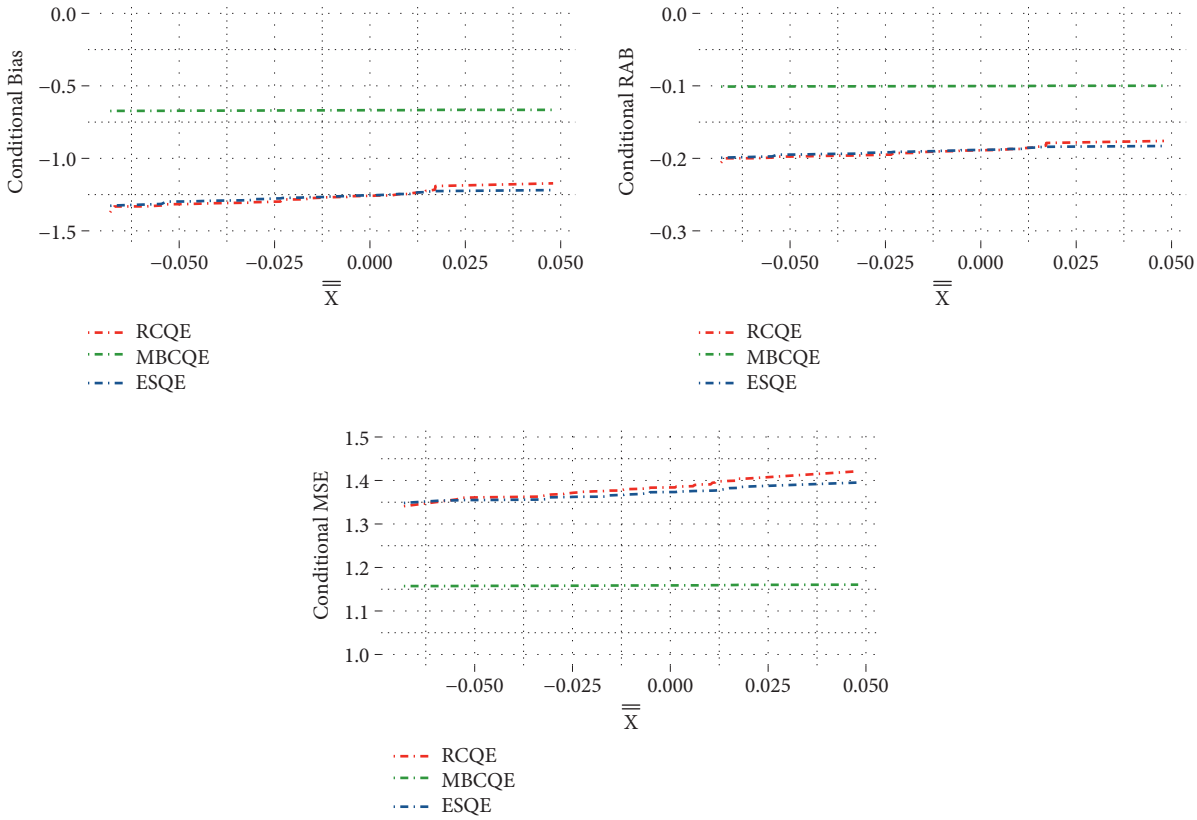

Figure 2: CB, CRAB, and CMSE for estimators: linear mean function, $\alpha = 0.5$.

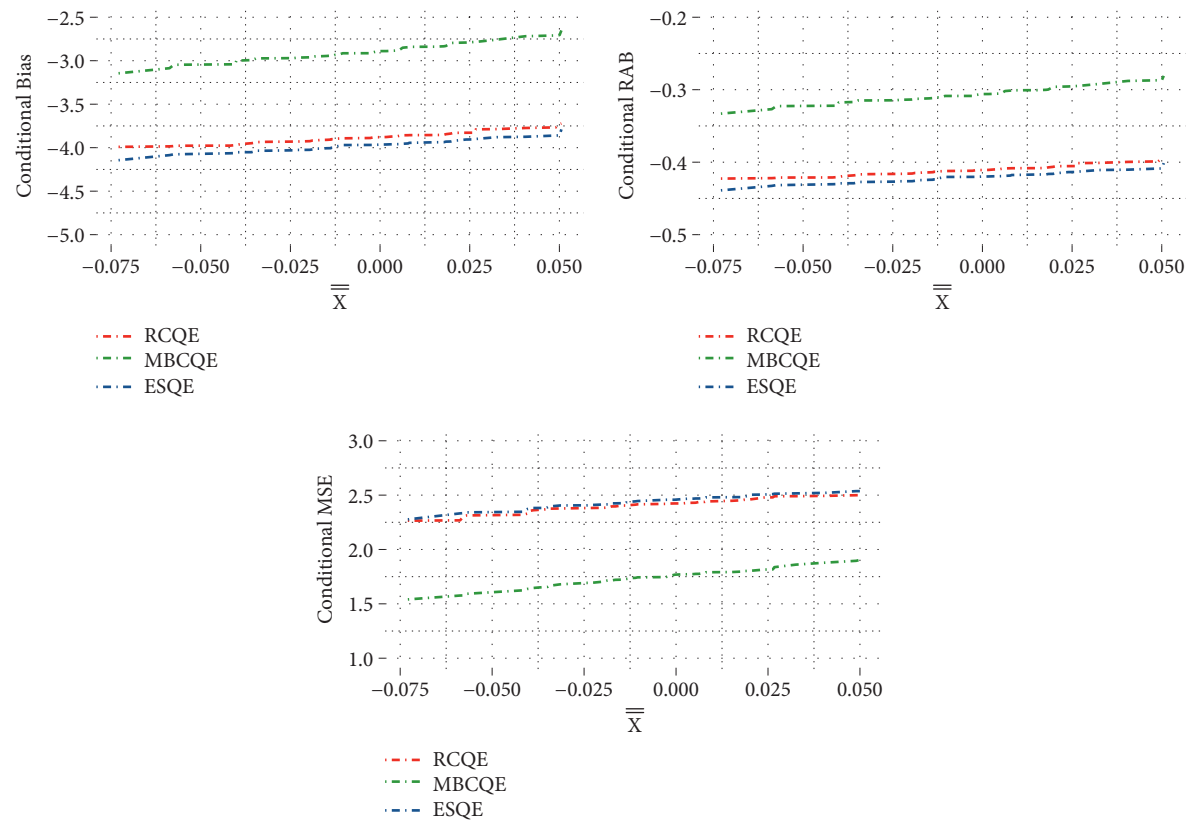

Figure 3: CB, CRAB, and CMSE for estimators: linear mean function, $\alpha = 0.75$.

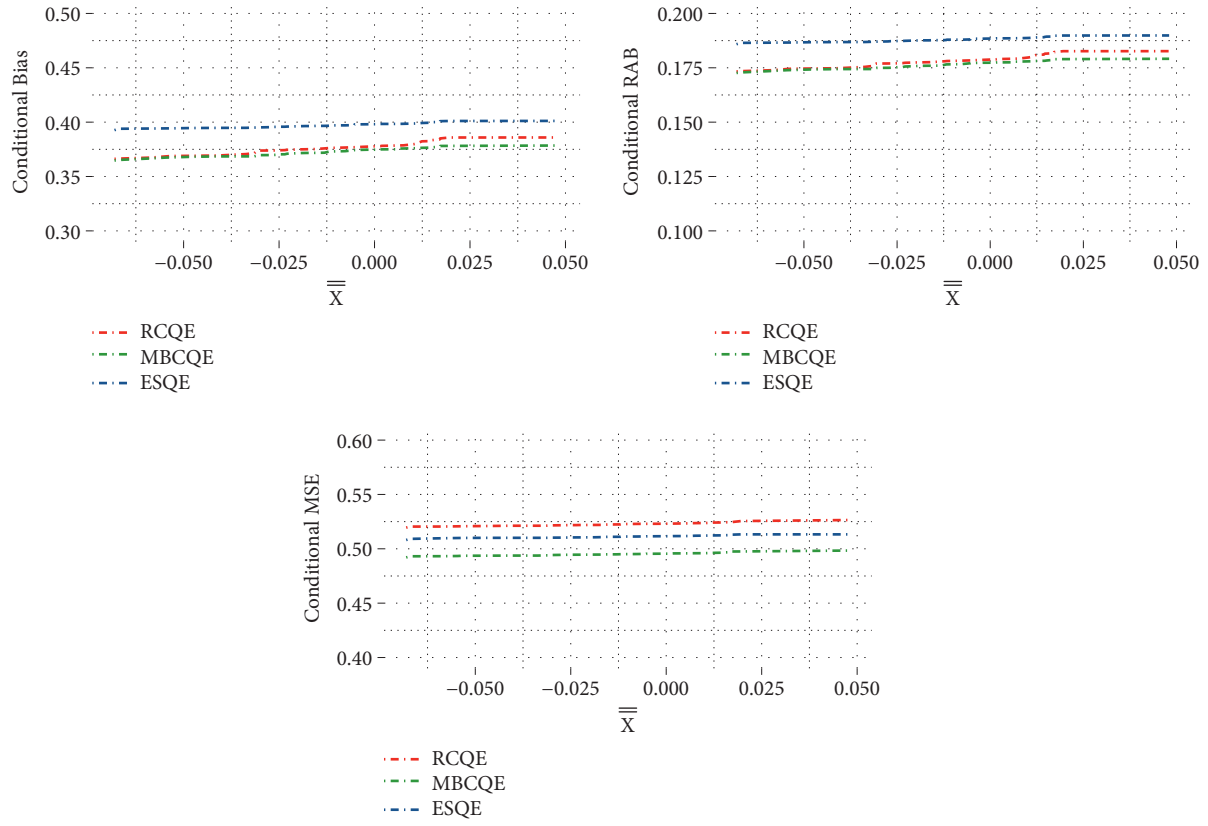

Figure 4: CB, CRAB, and CMSE for estimators: cosine mean function, $\alpha = 0.25$.

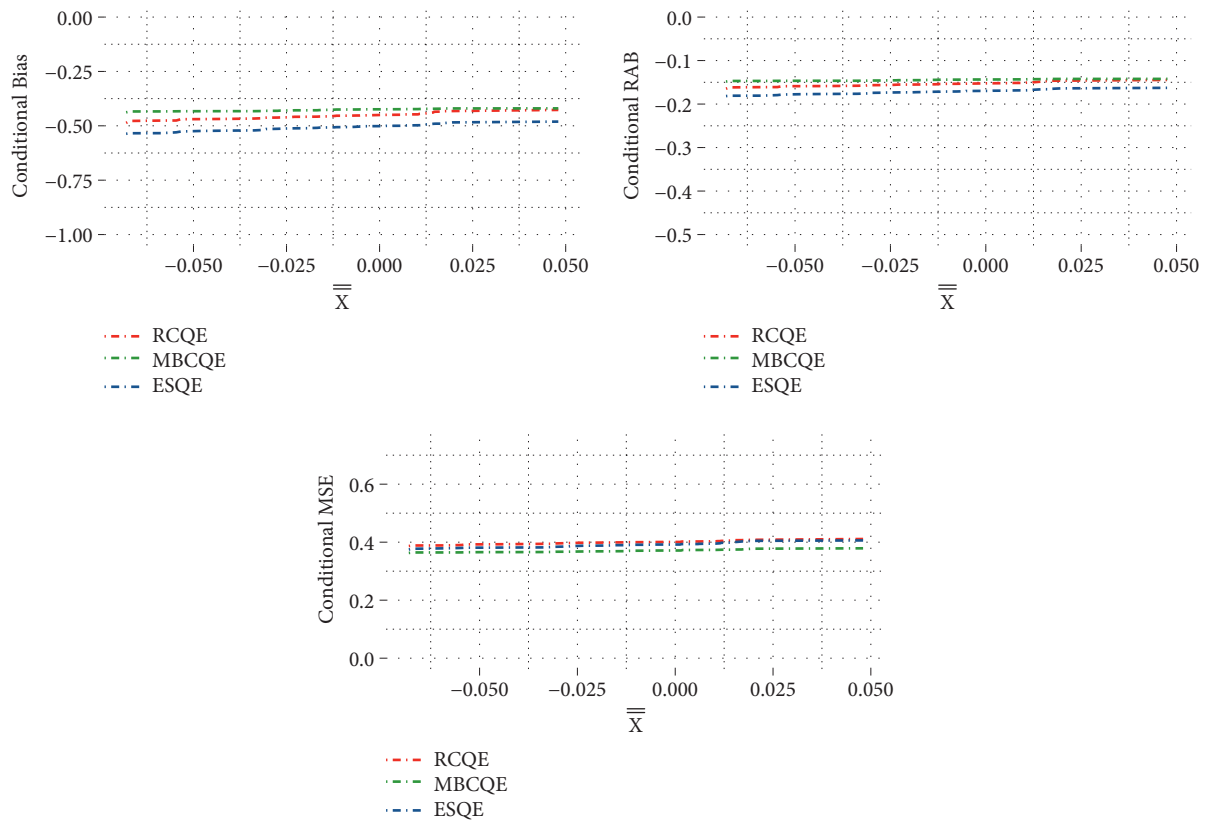

Figure 5: CB, CRAB, and CMSE for estimators: cosine mean function, $\alpha = 0.5$.

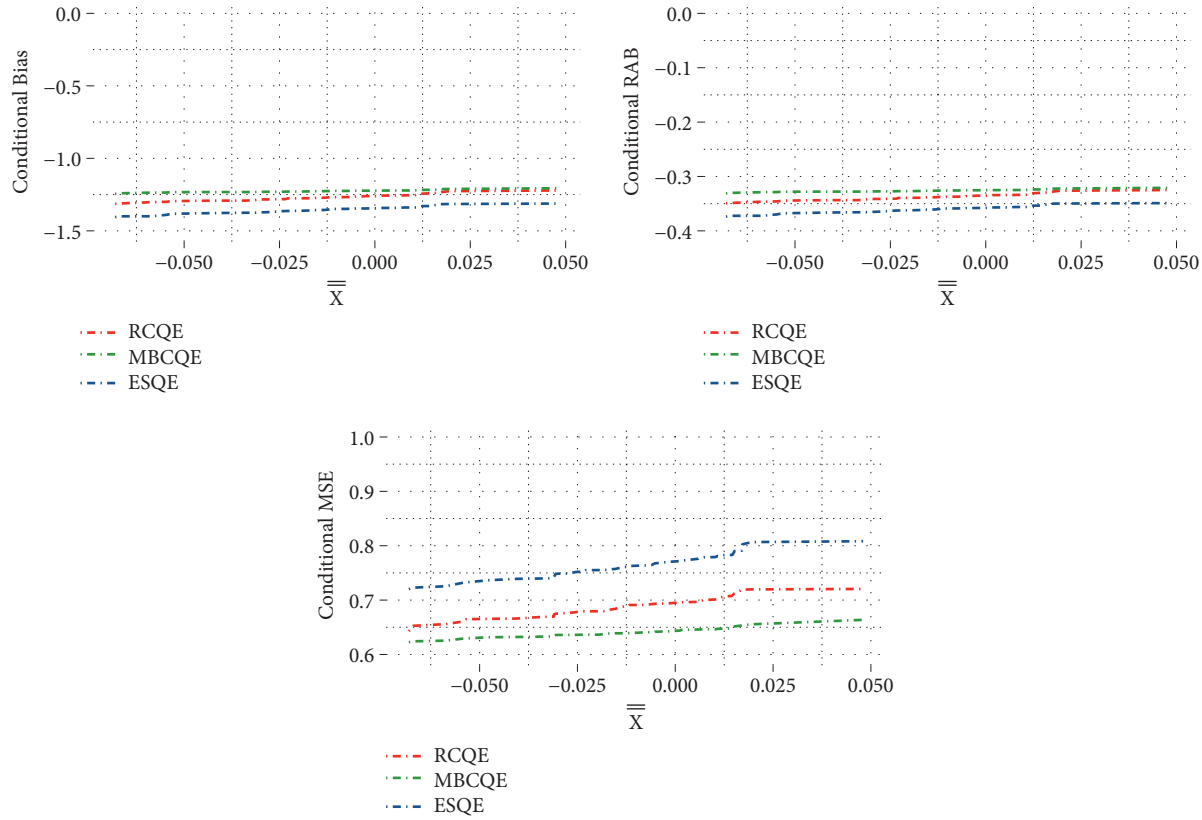

Figure 6: CB, CRAB, and CMSE for estimators: cosine mean function, $\alpha = 0.75$.

The corresponding estimator of the quantile function according to Chambers and Clark [23] was defined by

$$\widehat{Q}_{RC}(p) = \inf\left\{t: \widehat{F}_{Ny}^{rob}(t) \geq p\right\}, 0 < p < 1. \tag{41}$$

5.3. Results. The unconditional biases, unconditional relative mean error (RME), and unconditional relative root-mean-squared error (RRMSE) for the estimators for various values of the quantile α (i.e., 0.25, 0.5, and 0.75) are shown in Table 1. The findings were tabulated using linear and cosine mean functions. Additional mean functions, including bump, quadratic, cycle, and sine, can provide comparable findings and draw similar conclusions. For any estimator $\widehat{Q}_{n,X}^{S}(p)$, say, we define the relative mean error as

$$\text{RME} = \frac{1}{Q^S(p)}\left\{\frac{1}{1000}\sum_{r=1}^{1000}\left(\widehat{Q}_{n,X}^{S(r)}(p) - Q^S(p)\right)\right\}, \tag{42}$$

and the relative root-mean-squared error as

$$\text{RRMSE} = \frac{1}{Q^S(p)}\sqrt{\frac{1}{1000}\sum_{r=1}^{1000}\left(\widehat{Q}_{n,X}^{S(r)}(p) - Q^S(p)\right)^2}, \tag{43}$$

where $\widehat{Q}_{n,X}^{S(r)}(p)$ is the quantile corresponding to the r^{th} simulated sample.

It is clear from Table 1 that, in terms of bias, MBCQE is less biased than ESQE and RCQE for all values of α since it exhibits a smaller bias. In terms of performance as measured by RME and RRMSE, MBCQE is better than ESQE and RCQE since it has smaller values of RME and RRMSE for both linear and cosine mean functions.

Table 2 tabulates the quantile estimates findings of the two different sets of mean function. Using $n = 200$, and $\alpha = 0.25$, 0.50, and 0.75, this table illustrates the true population quantile $Q(p)$, MBCQE, RCQE, and ESQE. Comparison of $Q(p)$ to the listed estimators suggests that MBCQE is better estimator of the true population quantile since it is close to it at all probability levels.

We now turn to the conditional performances of the estimators by studying the plots of conditional bias (CB), conditional relative absolute bias (CRAB), and conditional mean-squared error (CMSE) of the estimators plotted versus group means of the means of auxiliary variables, $\overline{\overline{X}}$ for quantile levels 0.25, 0.50, and 0.75. The objective is to determine whether significant differences exist among these various estimators. In Figures 1–6, the red, green, and blue lines, respectively, represent RCQE, MBCQE, and ESQE.

Figures 1–3 show the conditional bias (CB), conditional relative absolute bias (CRAB), and conditional mean-squared error (CMSE) when linear mean functions were considered, and Figures 4–6 show the conditional bias (CB), conditional relative absolute bias (CRAB), and conditional mean-squared error (CMSE) when a cosine mean function was used.

Expected value, bias, and MSE are functions of sample size and the quantile level, and they can be used to exhibit the performance characteristics of individual estimators. Bias and MSE are two criteria by which estimators can be compared. Estimators should have low bias and minimum MSE.

It is clear that the proposed estimator MBCQE has a lower bias and minimum MSE at all values of α-quantile, as shown in Figures 1–6 for both linear and cosine mean functions. It is evident that MBCQE outperforms all other estimators investigated. Our results indicate that the proposed estimator MBCQE performs well, both unconditionally and conditionally.

6. Conclusion

The quantile estimator based on stratified simple random sampling has been proposed. We investigated the proposed estimator's properties and discovered that it possesses asymptotic normal distributions. Under SSRS, it is also an asymptotically unbiased estimator and asymptotically consistent estimator of population quantiles. It is clear from simulation results that the quantile estimator based on SSRS results in a larger decrease of bias than the one achieved using Chambers and Clark [23], Samawi et al. [8], and Eftekharian and Samawi [9]. In terms of performance, MBCQE has consistently produced results that are more precise than existing quantile estimators. We can therefore conclude that MBCQE can be used in estimating finite population quantiles for stratified populations in various sectors since it yields very good results.

Further study on the constructing of confidence intervals for the suggested estimator can be done, and a researcher can explore other bias correction procedures in quantile estimation, including adaptive boosting and bootstrap bias reduction strategies. Furthermore, the design of quantile estimators under stratified rank set sampling, as well as the usage of complex sample designs such as cluster sampling, is a research focus of discussion.

Data Availability

The dataset used to back up the theoretical assertions was obtained through simulation using the R-GUI statistical software.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] S. K. Thompson, "Simple random sampling," *Sampling*, vol. 755, pp. 9–37, 2012.

[2] R. Scheaffer, W. Mendenhall III, L. Ott, and K. Gerow, *Survey Sampling*, Brooks/Cole–Cengage Learning, Stamford, CT, USA, 7th edition, 2012.

[3] N. Sedransk and J. Sedransk, "Distinguishing among distributions using data from complex sample designs," *Journal of the American Statistical Association*, vol. 74, no. 368, pp. 754–760, 1979.

[4] H. M. Samawi and O. A. Al-Sagheer, "On the estimation of the distribution function using extreme and median ranked set sampling," *Biometrical Journal*, vol. 43, no. 3, pp. 357–373, 2001.

[5] E. A. Nadaraya, "Some new estimates for distribution functions," *Theory of Probability and Its Applications*, vol. 9, no. 3, pp. 497–500, 1964.

[6] Y. Lio and W. Padgett, "A note on the asymptotically optimal bandwidth for nadaraya's quantile estimator," *Statistics and Probability Letters*, vol. 11, no. 3, pp. 243–249, 1991.

[7] M. C. Jones, "Estimating densities, quantiles, quantile densities and density quantiles," *Annals of the Institute of Statistical Mathematics*, vol. 44, no. 4, pp. 721–727, 1992.

[8] H. Samawi, A. Chatterjee, J. Yin, and H. Rochani, "On quantiles estimation based on different stratified sampling with optimal allocation," *Communications in Statistics-Theory and Methods*, vol. 48, no. 6, pp. 1529–1544, 2019.

[9] A. Eftekharian and H. Samawi, "On kernel-based quantile estimation using different stratified sampling schemes with optimal allocation," *Journal of Statistical Computation and Simulation*, vol. 91, no. 5, pp. 1040–1056, 2021.

[10] T. Gasser and H.-G. Müller, "Kernel estimation of regression functions," in *Smoothing Techniques for Curve Estimation*-Springer, Berlin, Germany, 1979.

[11] R. John, "Boundary modification for kernel regression," *Communications in Statistics-Theory and Methods*, vol. 13, no. 7, pp. 893–900, 1984.

[12] M. C. Jones, "Simple boundary correction for kernel density estimation," *Statistics and Computing*, vol. 3, no. 3, pp. 135–146, 1993.

[13] J. S. Marron and D. Ruppert, "Transformations to reduce boundary bias in kernel density estimation," *Journal of the Royal Statistical Society: Series B*, vol. 56, no. 4, pp. 653–671, 1994.

[14] H.-G. Müller, "Smooth optimum kernel estimators near endpoints," *Biometrika*, vol. 78, no. 3, pp. 521–530, 1991.

[15] H.-G. Muller and J.-L. Wang, "Hazard rate estimation under random censoring with varying kernels and bandwidths," *Biometrics*, vol. 50, no. 1, pp. 61–76, 1994.

[16] E. F. Schuster, "Incorporating support constraints into nonparametric estimators of densities," *Communications in Statistics-Theory and Methods*, vol. 14, no. 5, pp. 1123–1136, 1985.

[17] O. Linton and J. P. Nielsen, "A multiplicative bias reduction method for nonparametric regression," *Statistics & Probability Letters*, vol. 19, no. 3, pp. 181–187, 1994.

[18] W. M. Onsongo, R. O. Otieno, and G. O. Orwa, "Nonparametric estimation of distribution function for stratified populations," *International Journal of Probability and Statistics*, vol. 7, no. 5, pp. 125–129, 2018.

[19] J. Kiefer, "On bahadur's representation of sample quantiles," *The Annals of Mathematical Statistics*, vol. 38, no. 5, pp. 1323–1342, 1967.

[20] R. R. Bahadur, "A note on quantiles in large samples," *The Annals of Mathematical Statistics*, vol. 37, no. 3, pp. 577–580, 1966.

[21] C. A. Francisco, *Estimation of Quantiles and the Interquartile Range in Complex surveys*, PhD Thesis, Iowa State University, Ames, IA, USA, 1987.

[22] R. Serfling, *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, New York, NY, USA, 1980.

[23] R. Chambers and R. Clark, *An Introduction to Model-Based Survey Sampling with Applications*, OUP, Oxford, UK, 2012.