

Bias Correction Technique for Estimating Quantiles of Finite Populations under Simple Random Sampling without Replacement

Nicholas Makumi¹, Romanus Odhiambo Otieno², George Otieno Orwa³, Festus Were³, Habineza Alexis¹

¹Pan African University, Institute for Basic Sciences, Technology and Innovation, Nairobi, Kenya

²Department of Mathematics, Meru University of Science and Technology, Meru, Kenya

³Department of Statistics and Actuarial Sciences, JKUAT, Nairobi, Kenya

Email: makuminicholas@gmail.com, rodhiambo@must.ac.ke, gorwa@jkuat.ac.ke, werref87@gmail.com, alexhabk87@gmail.com

How to cite this paper: Makumi, N., Otieno, R.O., Orwa, G.O., Were, F. and Alexis, H. (2021) Bias Correction Technique for Estimating Quantiles of Finite Populations under Simple Random Sampling without Replacement. *Open Journal of Statistics*, 11, 854-869.

<https://doi.org/10.4236/ojs.2021.115050>

Received: August 6, 2021

Accepted: October 16, 2021

Published: October 19, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In this paper, the problem of nonparametric estimation of finite population quantile function using multiplicative bias correction technique is considered. A robust estimator of the finite population quantile function based on multiplicative bias correction is derived with the aid of a super population model. Most studies have concentrated on kernel smoothers in the estimation of regression functions. This technique has also been applied to various methods of non-parametric estimation of the finite population quantile already under review. A major problem with the use of nonparametric kernel-based regression over a finite interval, such as the estimation of finite population quantiles, is bias at boundary points. By correcting the boundary problems associated with previous model-based estimators, the multiplicative bias corrected estimator produced better results in estimating the finite population quantile function. Furthermore, the asymptotic behavior of the proposed estimators is presented. It is observed that the estimator is asymptotically unbiased and statistically consistent when certain conditions are satisfied. The simulation results show that the suggested estimator is quite well in terms of relative bias, mean squared error, and relative root mean error. As a result, the multiplicative bias corrected estimator is strongly suggested for survey sampling estimation of the finite population quantile function.

Keywords

Quantile Function, Kernel Estimator, Multiplicative Bias Correction Technique, Simple Random Sampling without Replacement

1. Introduction

In recent years, the estimation of population distribution functions in the context of survey sampling has received considerable attention. A particular focus of this attention was the median, which is often considered to be a more acceptable position measure than the mean, especially when the interest variable follows a distorted distribution. Modern population mean or total estimators may typically be significantly enhanced when appropriate supplementary information is made available. Accordingly, the use of the auxiliary information in sample quantile estimators seems highly desirable. Use of known auxiliary knowledge both at the estimation stage and at the selection stage contributes to better estimation strategies in the sampling of surveys. If such information is not fully known or missing and information on the auxiliary variable(s) is relatively cheaper to obtain, one may consider taking a broad preliminary sample to estimate the auxiliary variable population mean(s).

Traditional kernel estimation methods have generally held that the performance of kernel methods depends largely on the smoothing bandwidth of the kernel, and very little depends on the type of the kernel. Most kernels used are symmetric kernels and are set once chosen. This may be useful for estimating unbounded support curves, but not for curves that have compact support and are discontinuous at boundary points. For curves of this kind, a fixed kernel shape leads to a boundary bias. This boundary bias is due to the weight allocation of the fixed symmetric kernel outside the distribution support when smoothing close to the boundary takes place. In addition, standard kernel methods yield wiggly estimates in the tail of the distribution as the reduction of the boundary bias leads to a limited bandwidth that prevents the pooling of appropriate data. Even otherwise, as noted in [1] when estimating the probability density function, the standard kernel estimator “works well for densities not far from Gaussian in shape”, however, it can perform very poorly when the shape seems far from Gaussian, particularly near the boundary.

Boundary bias is a well-known problem, and several scholars have proposed ways to eliminate it. In the context of nonparametric regression, [2] [3] [4] proposed the use of boundary kernels, while [5] used Richardson’s extrapolation to combine two kernel estimates with different bandwidths. In density estimation, [6] proposed data reflection, [7] considered empirical transformations, and [8] proposed a framework of jackknife methods for correcting boundary bias. In recent years, it has been shown by [9] [10], that in nonparametric regression, local linear smoother is free of boundary bias and achieves the optimal convergence for mean integrated squared error. It is interesting to note a local linear smoother uses a fixed kernel in its initial form, and the local least-regression implicitly employs different kernels at different places. The transformation method is among the numerous methods suggested to deal with data on $[0, +\infty]$. In order to minimize the boundary bias in the density estimation framework, [1] [11] [12], among others, studied general transformation methods. The transformation may operate under unique conditions and it is important to select the ap-

appropriate transformation by analyzing the subject matter and related studies.

The estimation of population quantiles is of great interest when a parametric form for the underlying distribution is not available. In a broad range of statistical applications, quantile estimation plays an important role: the Q-Q plot; the goodness-of-fit, the computation of extreme quantiles and value at risk in insurance business and financial risk management. Also, a large class of actuarial risk measures can be defined as functional of quantiles see ([13]). Most contributions have been made based on simple random sampling (SRS) to estimate the p th quantile using a kernel function. The reader can be referred to [14] [15] [16].

Quantile estimation has been intensively used in many fields. Most of the existing quantile estimators suffer from either a bias or an inefficiency for high probability levels. In order to correct the bias problems, [17] suggested several nonparametric quantile estimators based on the beta-kernel and applied them to transformed data. A Monte Carlo based study showed that those estimators improve the efficiency of the traditional ones, not only for light tailed distributions, but also for heavy tailed, when the probability level is close to 1, [18] used transformed kernel estimate. In their study, they overcame this inconsistency by using a new approach based on the modified Champernowne distribution which behaves as the Pareto distribution.

As a result, the aim of this paper is to develop a nonparametric estimator for the quantile function of finite populations using a bias corrected approach to address the shortcomings of previously studied estimation methods. There are two unique features about this approach. One is that it ensures an accurate estimate and the other is that it reduces the estimation bias with negligible increase in variance.

The concept of Multiplicative Bias Correction (MBC) approach was first considered in [19], and the results obtained showed that the estimator of the regression function had desirable properties compared to existing estimators, including solving the boundary problems. This form of correction is especially well suited for changing non-negative regression function because it does not change the sign of the regression function and ensures an accurate estimate and reduces the estimation bias with negligible increase in variance. As there is always a bias-variance trade off for non-parametric smoothers in finite samples, smoothers can be generated whose asymptotic bias converges to zero while maintaining the same asymptotic variance. For a deeper discussion of Multiplicative Bias Correction technique we refer the reader to [20] [21] [22] [23].

Outline of the paper

In Section 2, we propose an estimator for finite population quantile function using a bias correction technique. Asymptotic properties of the proposed estimator are derived in Section 3. Empirical study of the results is given in Section 4 and the conclusion of the findings is given in Section 5.

2. Proposed Estimator

In the sampling survey, we are time and again interested in studying the distri-

bution of a specific variable of interest, Y . The efficient technique to illustrate the distribution function is by assessing the quantiles of the distribution. By the p^{th} quantile of the distribution, we imply the value Q , which would be $P(Y \leq Q) = p$. One way of designing quantile estimators is to invert the estimator of the distribution function. Let $\hat{F}_y(t)$ denote an estimator of $P(Y \leq t)$. Since the estimator \hat{F}_y is often a step function, the form of the quantile estimator may not be smooth.

In this section we discuss a quantile estimator derived from a model-based multiplicative bias correction distribution function estimator that integrates auxiliary information. This distribution function estimator was introduced by [24]. The quantile estimator is based on inverting the [24] distribution function estimator. We derive a Bahadur representation for the quantile estimator.

Let $F(y) = P(Y < y)$ be a probability distribution function. The population quantile of order α is defined as

$$Q(\alpha) = \inf \{t : F(t) > \alpha\} \tag{1}$$

for $0 < \alpha < 1$. If F is continuous and strictly increasing, then

$$Q(\alpha) = F^{-1}(\alpha)$$

is the unique solution to Equation (1). In general, $Q(\alpha)$ satisfies

$$P\{Y < Q(\alpha)-\} < \alpha < P\{Y < Q(\alpha)\}, \tag{2}$$

or equivalently

$$F(Q(\alpha)-) < \alpha < F(Q(\alpha)). \tag{3}$$

Suppose that X_i 's for $i = 1, 2, \dots, N$ are independent and identically distributed (i.i.d) random variables with conforming survey values Y_i ($i = 1, 2, \dots, N$). By definition, Y_1, \dots, Y_N are independent, identically distributed random variables, each with common distribution function F . For all real t , the empirical population distribution function for Y_1, \dots, Y_N is defined to be

$$F_{N_y}(t) = \frac{1}{N} \sum_{i=1}^N I(y_i \leq t) \tag{4}$$

where

$$I(y_i \leq t) = \begin{cases} 1 & \text{if } y_i \leq t; \\ 0 & \text{if otherwise.} \end{cases}$$

The sample quantile of order α is defined as

$$Q_{N_y}(\alpha) = \inf \{t : F_{N_y}(t) > \alpha\} \tag{5}$$

The sample quantile of order α is a strongly consistent estimator of $Q(\alpha)$, unless $F(Q(\alpha)) = \alpha$ and $F(Q(\alpha)) = F(Q(\alpha) + \varepsilon)$ for some $\varepsilon > 0$ (i.e., unless F is flat in a right neighborhood of $Q(\alpha)$). See ([25]).

Theorem 1 ([26]). Let Y_1, \dots, Y_N be a random sample of size n with common distribution function F , and let $0 < p < 1$. If $y = q(p)$ is the unique solution

of $F(y-) < p < F(y)$, then $\hat{q}_n(p) \xrightarrow{a.s.} q(p)$ as $n \rightarrow \infty$.

Suppose a sample s of n units is drawn through simple random sampling without replacement from a finite population and $r = n - s$ be the non-sampled units of the finite population. Let Y be the survey variable associated with auxiliary variable X which are assumed to follow superpopulation model under model-based approach. A commonly used working model for the finite population is

$$Y_i = \mu(x_i) + \sigma(x_i)e_i, i = 1, 2, \dots, N \tag{6}$$

where $\sigma(x_i)$ is a known function of x_i that accounts for heteroscedasticity and e_i 's are independent and identically distributed (i.i.d) random variables with mean 0 and variance σ^2 , $E(Y_i) = \mu(x_i)$ and

$$Cov(Y_i, Y_j) = \begin{cases} \sigma^2(x_i) & \text{if } i = 1, 2, \dots, N \\ 0 & \text{elsewhere.} \end{cases}$$

Under model-based approach Equation (4) can be expressed as

$$F_{N_y}^*(t) = \frac{1}{N} \left[\sum_{i \in s} I(y_i \leq t) + \sum_{j \in r} I(y_j \leq t) \right] \tag{7}$$

where $\sum_i I(y_i \leq t)$ represent the sampled part and is known while $\sum_j I(y_j \leq t)$ is the non-sampled part which is unknown.

The problem is estimating the second term of Equation (7). To estimate Equation (7), [24] proposed a multiplicative bias corrected estimator for finite population distribution given by

$$\hat{F}_{MBC}(t) = \frac{1}{N} \left\{ \sum_{i \in s} I(y_i \leq t) + \sum_{j \in r} \hat{H}(t - \hat{\mu}(x_j)) \right\} \tag{8}$$

where $\hat{\mu}(x_j)$ is the model-based nonparametric estimator for $\mu(x_j)$ and $\hat{H}(t - \hat{\mu}(x_j))$ is the estimated distribution function of the residuals defined by $e_j = y_j - \hat{\mu}(x_j)$.

In this study, we propose a multiplicative bias corrected quantile estimator for finite population based on finite population distribution in Equation (8) given by

$$\begin{aligned} \hat{Q}_{MBC}(\alpha) &= \inf \{ t \in U : \hat{F}_{MBC}(t) > \alpha \} = \hat{F}_{MBC}(\hat{Q}_{MBC}(\alpha)) \\ &= \frac{1}{N} \left\{ \sum_{i \in s} I(y_i \leq \hat{Q}_{MBC}(\alpha)) + \sum_{j \in r} \hat{H}(\hat{Q}_{MBC}(\alpha) - \hat{\mu}(x_j)) \right\} \end{aligned} \tag{9}$$

The problem is to estimate $\hat{Q}_{MBC}(\alpha)$ for any α given. Thus, from the sample of n units of a population of size N , we observe y_1, \dots, y_n . The general method is formulated as follows: first obtain an estimator of the distribution function, $\hat{F}_{MBC}(t)$, and then estimate the quantile by taking the inverse.

3. Properties of Proposed Estimator

3.1. Asymptotic Unbiasedness

In simple random sampling, as $n\hat{F}_{MBC}(Q_{N_y}(\alpha))$ is hypergeometrically distri-

buted variable, then

$$E\left[\hat{F}_{MBC}\left(Q_{N_y}(\alpha)\right)\right]=F_{N_y}\left(Q_{N_y}(\alpha)\right)=\alpha \tag{10}$$

and

$$Var\left[\hat{F}_{MBC}\left(Q_{N_y}(\alpha)\right)\right]=\frac{1-f}{n}\alpha(1-\alpha) \tag{11}$$

If the sample size n is sufficiently large then $\hat{F}_{MBC}\left(Q_{N_y}(\alpha)\right)$ is approximately normal.

Theorem 2 ([27]). *Let x be in the interval A_1 containing $q\left(\gamma_1^0\right)$ as an interior point. Then the sample quantile,*

$$\hat{q}_r(\gamma)=q(\gamma)-\left[f\left(q(\gamma)\right)\right]^{-1}\left[F_r\left(q(\gamma)\right)-F\left(q(\gamma)\right)\right]+R_r^*(\gamma) \tag{12}$$

with $R_r^*(\gamma)=o_p\left(n_r^{-1/2}\right)$ uniformly in γ for γ in H_1 , where $H_1=\left\{\gamma:F(x)=\gamma \text{ and } x \in A_1\right\}$

Proof: For proof see [27].

We now study the properties of the $\hat{Q}_{MBC}(\alpha)$ estimator. For this, a linear approximation is needed because $\hat{Q}_{MBC}(\alpha)$ is not a continuous function. The estimator $\hat{Q}_{MBC}(\alpha)$ can be expressed asymptotically as a linear function of the estimated distribution function evaluated at the quantile $Q_{N_y}(\alpha)$ by the Bahadur representation (see [28]) together with the results from Theorem 2 above.

Let F_{MBC} be the multiplicative bias corrected distribution function of the density f_{MBC} .

Theorem 3 (Taylor’s Theorem). *Let $k \geq 1$ be an integer and let the function $f: \mathbb{R} \rightarrow \mathbb{R}$ be k times differentiable at the point $a \in \mathbb{R}$. Then there exists a function $h_k: \mathbb{R} \rightarrow \mathbb{R}$ such that*

$$f(x)=f(a)+\sum_{n=1}^k \frac{f^{(n)}(a)}{n!}(x-a)^n+h_k(x)(x-a)^k, \tag{13}$$

and $\lim_{x \rightarrow a} h_k(x)=0$. This is called the Peano form of the remainder.

Then using Taylor series expansion of the function $F_{MBC}\left(\hat{Q}_{MBC}(\alpha)\right)$ around $Q_{N_y}(\alpha)$ we can write

$$\begin{aligned} F_{MBC}\left(\hat{Q}_{MBC}(\alpha)\right) &= \sum_{n=0}^{\infty} \frac{F_{MBC}^{(n)}\left(Q_{N_y}(\alpha)\right)}{n!}\left[\hat{Q}_{MBC}(\alpha)-Q_{N_y}(\alpha)\right]^n \\ &= F_{MBC}\left(Q_{N_y}(\alpha)\right)+F_{MBC}'\left(Q_{N_y}(\alpha)\right)\left[\hat{Q}_{MBC}(\alpha)-Q_{N_y}(\alpha)\right] \\ &\quad +F_{MBC}''\left(Q_{N_y}(\alpha)\right)\left[\hat{Q}_{MBC}(\alpha)-Q_{N_y}(\alpha)\right]^2+\dots \\ &= F_{MBC}\left(Q_{N_y}(\alpha)\right)+f_{MBC}\left(Q_{N_y}(\alpha)\right)\left[\hat{Q}_{MBC}(\alpha)-Q_{N_y}(\alpha)\right]+O\left(n^{-\frac{1}{2}}\right) \end{aligned} \tag{14}$$

where $F_{MBC}'\left(Q_{N_y}(\alpha)\right)=f_{MBC}\left(Q_{N_y}(\alpha)\right)$, according to [29] since F_{MBC} contains two derivatives in a $Q_{N_y}(\alpha)$ neighborhood, this neighborhood is bound by the second derivative and $F_{MBC}'\left(Q_{N_y}(\alpha)\right)$ is positive.

Then solving for $\hat{Q}_{MBC}(\alpha)$ in Equation (14) we have the Bahadur’s repre-

sensation as

$$\begin{aligned} & \hat{Q}_{MBC}(\alpha) - Q_{N_y}(\alpha) \\ &= \frac{1}{f_{MBC}(Q_{N_y}(\alpha))} \left[F_{MBC}(\hat{Q}_{MBC}(\alpha)) - F_{MBC}(Q_{N_y}(\alpha)) \right] + O\left(n^{-\frac{1}{2}}\right) \end{aligned} \quad (15)$$

Moreover, it can be shown that

$$\begin{aligned} & F_{MBC}(\hat{Q}_{MBC}(\alpha)) - F_{MBC}(Q_{N_y}(\alpha)) \\ &= \hat{F}_{MBC}(\hat{Q}_{MBC}(\alpha)) - \hat{F}_{MBC}(Q_{N_y}(\alpha)) + O\left(n^{-\frac{1}{2}}\right) \end{aligned} \quad (16)$$

Substituting the above results of Equation (16) in Equation (15) yields

$$\begin{aligned} & \hat{Q}_{MBC}(\alpha) - Q_{N_y}(\alpha) \\ &= \frac{1}{f_{MBC}(Q_{N_y}(\alpha))} \left(\hat{F}_{MBC}(\hat{Q}_{MBC}(\alpha)) - \hat{F}_{MBC}(Q_{N_y}(\alpha)) \right) + O\left(n^{-\frac{1}{2}}\right) \\ &= \frac{1}{f_{MBC}(Q_{N_y}(\alpha))} \left(\alpha - \hat{F}_{MBC}(Q_{N_y}(\alpha)) \right) + O\left(n^{-\frac{1}{2}}\right) \end{aligned} \quad (17)$$

where $f_{MBC}(\cdot)$ denotes the derivative of the limiting value of $F_{MBC}(\cdot)$ as $N \rightarrow \infty$ and $\hat{F}_{MBC}(\hat{Q}_{MBC}(\alpha)) = \alpha$.

The linear approximation previously used by [30] [31] helps to study the asymptotic properties of the estimator. On the other hand, the estimator $\hat{Q}_{MBC}(\alpha)$ is asymptotically unbiased because $\hat{F}_{MBC}(t)$ is unbiased estimator of $\hat{F}_N(t)$ (see [24]). In this way

$$\begin{aligned} & E\left[\hat{Q}_{MBC}(\alpha) - Q_{N_y}(\alpha)\right] \\ &= E\left[\frac{1}{f_{MBC}(Q_{N_y}(\alpha))} \left(\alpha - \hat{F}_{MBC}(Q_{N_y}(\alpha)) \right) + O\left(n^{-\frac{1}{2}}\right)\right] \\ &= \frac{1}{f_{MBC}(Q_{N_y}(\alpha))} E\left(\alpha - \hat{F}_{MBC}(Q_{N_y}(\alpha))\right) + O\left(n^{-\frac{1}{2}}\right) \end{aligned} \quad (18)$$

but $E\left[\alpha - \hat{F}_{MBC}(Q_{N_y}(\alpha))\right] = 0$ and by using Equation (18) it can be seen that

$$E\left[\hat{Q}_{MBC}(\alpha)\right] = Q_{N_y}(\alpha) + O\left(n^{-\frac{1}{2}}\right) \quad (19)$$

The bias of $\hat{Q}_{MBC}(\alpha)$ is of order $O\left(n^{-\frac{1}{2}}\right)$. Thus, it converges to zero at a faster rate. Therefore, $\hat{Q}_{MBC}(\alpha)$ is asymptotically unbiased.

3.2. Asymptotic Variance

Asymptotic Variance of $\hat{Q}_{MBC}(\alpha)$ will be obtained as follows, Consider the Bahadur's representation:

$$\hat{Q}_{MBC}(\alpha) - Q_{N_y}(\alpha) = \frac{1}{f_{MBC}(Q_{N_y}(\alpha))}(\alpha - \hat{F}_{MBC}(Q_{N_y}(\alpha))) + O\left(n^{-\frac{1}{2}}\right) \quad (20)$$

Then applying variance on both side of Equation (20) we have

$$\begin{aligned} & Var\left[\hat{Q}_{MBC}(\alpha) - Q_{N_y}(\alpha)\right] \\ &= Var\left[\frac{1}{f_{MBC}(Q_{N_y}(\alpha))}(\alpha - \hat{F}_{MBC}(Q_{N_y}(\alpha))) + O\left(n^{-\frac{1}{2}}\right)\right] \\ &= \left[\frac{1}{f_{MBC}(Q_{N_y}(\alpha))}\right]^2 Var(\alpha - \hat{F}_{MBC}(Q_{N_y}(\alpha))) \\ &= \left[\frac{1}{f_{MBC}(Q_{N_y}(\alpha))}\right]^2 \{Var(\alpha) + Var(\hat{F}_{MBC}(Q_{N_y}(\alpha)))\} \\ &= \left[\frac{1}{f_{MBC}(Q_{N_y}(\alpha))}\right]^2 \frac{1-f}{n} \alpha(1-\alpha) \\ &= \frac{1-f}{n} \alpha(1-\alpha) [f_{MBC}(Q_{N_y}(\alpha))]^{-2} \end{aligned} \quad (21)$$

3.3. Asymptotic Mean Squared Error

The asymptotic mean squared error of the estimator $\hat{Q}_{MBC}(\alpha)$ is given by

$$MSE(\hat{Q}_{MBC}(\alpha)) = Var(\hat{Q}_{MBC}(\alpha)) + [Bias(\hat{Q}_{MBC}(\alpha))]^2 \quad (22)$$

Substituting Equations (19) and (21) we get

$$\begin{aligned} MSE(\hat{Q}_{MBC}(\alpha)) &= \frac{1-f}{n} \alpha(1-\alpha) [f_{MBC}(Q_{N_y}(\alpha))]^{-2} + \left[O\left(n^{-\frac{1}{2}}\right)\right]^2 \\ &= \frac{1-f}{n} \alpha(1-\alpha) [f_{MBC}(Q_{N_y}(\alpha))]^{-2} + O\left(\frac{1}{n}\right) \end{aligned} \quad (23)$$

Equation (23) tends to zero as $n \rightarrow \infty$ and thus $MSE(\hat{Q}_{MBC}(\alpha)) \rightarrow 0$. This shows that $\hat{Q}_{MBC}(\alpha)$ is asymptotically consistent.

4. Empirical Study

The main purpose of this section is to compare the performance of the proposed estimator *MBCQE* with the existing quantile estimators: *RKMQE*, *CDQE*, *FAQE* and *NWQE*. In this study, two populations are considered, which are generated from the regression model given by

$$y_i = m(x_i) + e_i$$

where $1 \leq i \leq 1000$ with the following mean functions described in **Table 1**.

A population of 1000 auxiliary values x_i are generated as independent and identically distributed uniform random variables, $x_i \in [0, 1]$. The mean functions

Table 1. Mean functions used in the simulation study.

Mean function	
Linear	$1 + 2(x_i - 0.5)$
Cosine	$\cos(1 + 2(x_i - 0.5))$

represent a class of correct and incorrect model specifications for the estimators being considered. The errors are assumed to be independent and identically distributed (i.i.d) normal random variables having mean 0 and standard deviation, $\sigma = 1$. They contain 1000 units and the population is simulated as i.i.d uniform random variables. The population values y_i 's are generated from the mean functions by adding the errors e_i 's in each of the cases. 1000 samples are simulated using simple random sampling without replacement for each case.

Nadaraya-Watson kernel weights are used in the smoothing of y_i to obtain the rough estimator, $\tilde{\mu}_n(x_i) = \sum_{j=1}^n w_i(x; l) y_j$, of the mean function (x_i) . A ratio $\beta_i = \frac{y_i}{\tilde{\mu}_n(x_i)}$ is evaluated and is smoothed further to obtain the correction factor $\hat{\alpha}(x_i)$ which is then used together with the rough estimator to obtain the multiplicative bias corrected estimator, $\hat{\mu}_n(x_i)$, of the mean function.

The existing estimators for quantile functions for finite populations that were used for comparison with our developed estimator Multiplicative Bias Corrected Quantile Estimator (MBCQE);

$$\hat{Q}_{MBC}(\alpha) = \inf \{t \in U : \hat{F}_{MBC}(t) > \alpha\}$$

are:

- 1) Chamber and Dunstan Quantile Estimator (CDQE):

$$\hat{Q}_{y,CD,\alpha} = \inf \{t | \hat{F}_{y,CD}(t) \geq \alpha\}$$

- 2) Nadaraya Watson Quantile Estimator (NWQE):

$$\tilde{Q}_{NW}(p) = \inf \{x : \hat{F}_n(x) \geq p\}, 0 < p < 1$$

- 3) Rao Kovar Mantel Quantile Estimator (RKMQE):

$$\hat{Q}_{RKM,\alpha} = \hat{F}_{rkm}^{-1}(\alpha)$$

- 4) Dorfman and Hall Quantile Estimator (FAQE):

$$\hat{Q}_{DH}(p) = \inf \{t : \hat{F}_{DH}(t) \geq p\}$$

The results of this simulation study are summarized in **Table 2**. **Table 2** shows the unconditional Biases, Relative Mean Error (RME) and Relative Root Mean Squared Error (RRMSE) for the estimators at various values of the quantile α (i.e. 0.25, 0.5 and 0.75). Linear and cosine mean functions were used to obtain the tabulated results. Similar results and conclusions can be obtained using other mean functions such as quadratic, sine, bump etc. To analyze the performance of the proposed estimator against some specified estimators, unconditional Relative

Table 2. Unconditional biases, relative mean errors and relative root mean squared errors.

α	Estimator	LINEAR			COSINE		
		Bias	RME	RRMSE	Bias	RME	RRMSE
0.25	<i>RKMQE</i>	0.6048	3.0136	0.7977	0.3820	1.0475	0.4958
	<i>CDQE</i>	1.7358	3.5052	0.9279	0.3411	1.1029	0.5220
	<i>FAQE</i>	0.7600	2.9623	0.7842	0.3438	0.9495	0.4494
	<i>NWQE</i>	-3.6280	3.6652	0.9702	-0.8407	0.9674	0.4579
	<i>MBCQE</i>	0.2452	2.9367	0.7773	0.2317	0.9384	0.4442
0.5	<i>RKMQE</i>	-0.6990	2.9746	0.4565	-0.4352	1.0471	0.3544
	<i>CDQE</i>	-1.2883	3.0075	0.4515	-0.5186	1.0699	0.3621
	<i>FAQE</i>	-1.0713	2.9777	0.4470	-0.3968	1.0682	0.3616
	<i>NWQE</i>	-4.3034	4.3185	0.6483	-1.6661	1.7740	0.6005
	<i>MBCQE</i>	-0.6840	2.8757	0.4317	-0.3854	1.0133	0.3430
0.75	<i>RKMQE</i>	-3.1533	4.13236	0.4376	-1.2097	1.5206	0.4045
	<i>CDQE</i>	-4.0437	4.8268	0.5112	-1.3309	1.6303	0.4337
	<i>FAQE</i>	-3.5487	4.5007	0.4767	-1.1978	1.5440	0.4107
	<i>NWQE</i>	-6.7790	6.7822	0.7183	-2.4528	2.4667	0.6562
	<i>MBCQE</i>	-2.9630	4.1183	0.4362	-1.1149	1.4458	0.3846

Mean Error and Relative Root Mean Squared Errors for the estimator $\hat{Q}_{n,X}(p)$ are computed as

$$RME = \frac{1}{Q(p)} \left\{ \frac{1}{N} \sum_{s=1}^N (\hat{Q}_{n,X}^{(s)}(p) - Q(p)) \right\} \tag{24}$$

and

$$RRMSE = \sqrt{\frac{1}{Q(p)} \left\{ \frac{1}{N} \sum_{s=1}^N (\hat{Q}_{n,X}^{(s)}(p) - Q(p))^2 \right\}} \tag{25}$$

where $\hat{Q}_{n,X}^{(s)}(p)$ is the quantile corresponding to the s^{th} simulated sample $\{X_1^s, X_2^s, \dots, X_n^s\}$ and N is the number of replications. The RME indicates the measure of how close the estimator being considered is from the actual value, while RRMSE indicates measure of accuracy of the estimator. For instance, an estimator, *MBCQE*, will be said to be “better” or more preferable than the other estimators if its RRMSE is comparably smaller.

Bias of a quantile estimator refers to the deviation of the expected value of the estimator from the true quantile value. All of the quantile estimators considered here are biased but comparatively *MBCQE* exhibits a smaller bias. *MBCQE* can be seen to be a very efficient estimator of the empirical quantile function at all levels of the α -quantile followed closely by *RKMQE* and *FAQE*. *CDQE* proved to be a very inefficient estimator at all levels of α .

Further, comparison of estimators was done with respect to empirical quantile function which further affirmed the results tabulated above. **Table 3** and **Table 4** give a tabulation of all the estimators listed below.

Table 3. Quantile estimates for linear mean function.

Estimators	Quantile estimates						
	0.05	0.10	0.25	0.50	0.75	0.90	0.95
$Q(p)$	1.2435	1.2435	3.7776	6.6617	9.4424	2.7891	1.2578
$RKMQE$	3.6002	3.7099	4.3825	5.9927	6.2890	2.3354	4.2224
$CDQE$	4.8990	4.8918	5.5135	5.3733	5.399	1.5787	2.3654
$FAQE$	3.4606	3.235	4.5376	5.5904	5.894	2.3354	4.2224
$NWQE$	-1.1691	-0.8517	0.1497	2.3583	2.6635	2.3354	4.2224
$MBCQE$	-0.3213	0.5437	4.0229	5.9777	6.4790	2.6474	1.2347

Table 4. Quantile estimates for cosine mean function.

Estimators	Quantile estimates						
	0.05	0.10	0.25	0.50	0.75	0.90	0.95
$Q(p)$	0.8917	1.3130	2.1128	2.9543	3.7592	4.5294	5.0166
$RKMQE$	2.469	2.5494	2.4947	2.5191	2.5495	2.5305	2.4572
$CDQE$	2.6548	2.6190	2.4539	2.4357	2.4282	2.6412	2.5945
$FAQE$	2.5477	2.3186	2.4566	2.5576	2.5615	2.7704	2.9166
$NWQE$	2.5477	2.2673	1.2721	1.2882	1.3064	1.2944	1.2680
$MBCQE$	1.6502	1.6450	2.3444	2.5690	2.6443	2.9880	3.1520

$CDQE$ overestimates the empirical quantile function at all points while $MBCQE$ give an almost perfect estimation of the empirical quantile function. On the other hand, $NWQE$ underestimates the true quantile function at some points towards the lower tail while it overestimates the same function at other points along the upper tail.

The conditional performance of the estimator was done and compared with the performance of other existing quantile estimators. To do this, 200 random samples, all of size 400, were selected and the mean of the auxiliary values x_i was computed for each sample to obtain 200 values of \bar{X} . These sample means were then sorted in ascending order and further grouped into clusters of size 20 such that a total of 10 groups were realized. Further, group means of the means of auxiliary variables was calculated to get $\bar{\bar{X}}$. Empirical means and biases were then computed for all the estimators $RKMQE$, $CDQE$, $FAQE$, $NWQE$ and $MBCQE$. The conditional biases were plotted against $\bar{\bar{X}}$ to provide a good understanding of the pattern generated. **Figures 1-6** show the behavior of the conditional biases, relative absolute biases and mean squared error realized by all the estimators of quantile functions under linear and cosine mean functions at various values of the quantile α (i.e. 0.25, 0.5 and 0.75).

In most cases, there are significant differences among the bias characteristics of the various estimators. A detailed examination of the plots reveals that $MBCQE$ and $RKMQE$ have lower levels of bias overall, as indicated by the

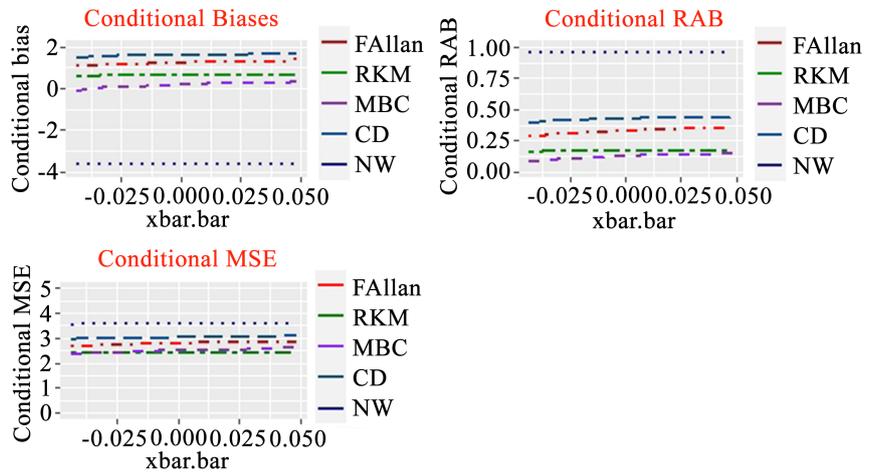


Figure 1. Conditional biases, RAB and MSE for the estimators using a linear mean function at $\alpha = 0.25$.

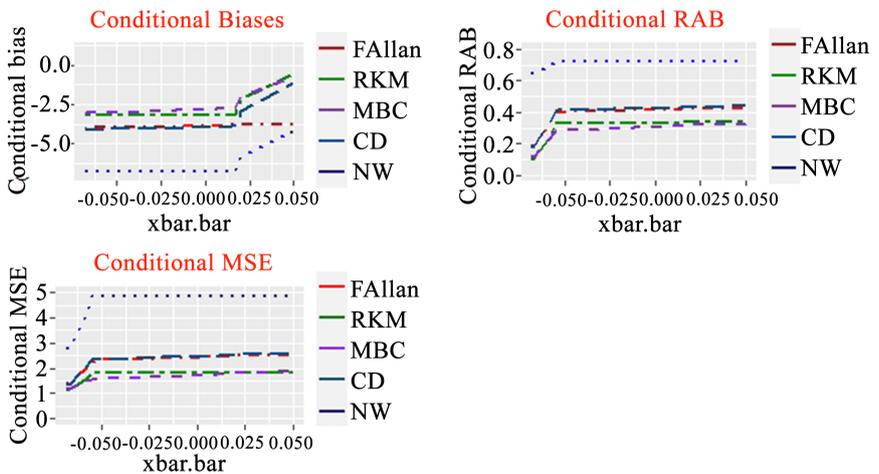


Figure 2. Conditional biases, RAB and MSE for the estimators using a linear mean function at $\alpha = 0.5$.

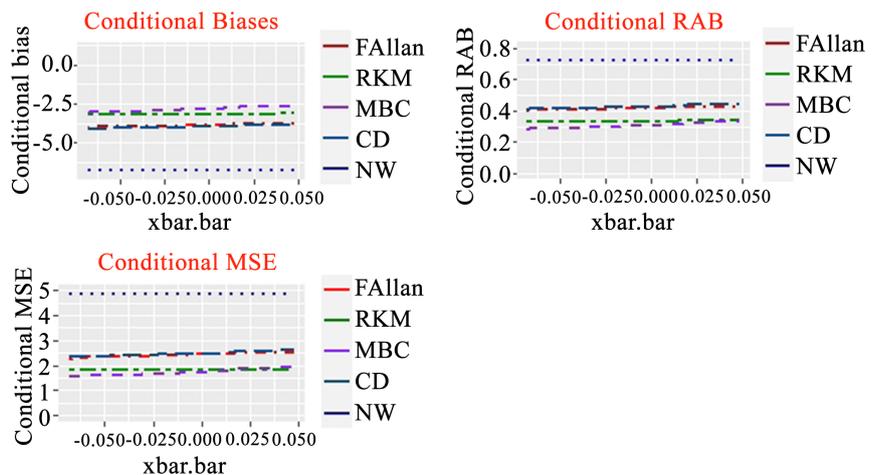


Figure 3. Conditional biases, RAB and MSE for the estimators using a linear mean function at $\alpha = 0.75$.

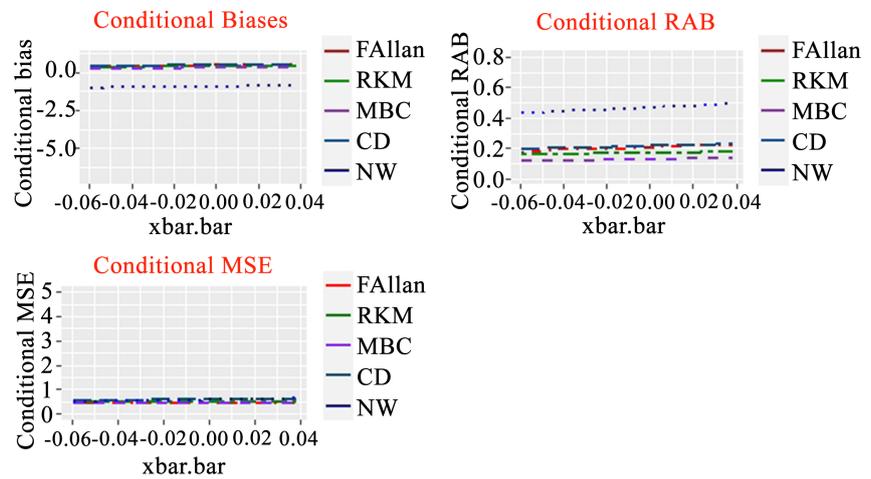


Figure 4. Conditional biases, RAB and MSE for the estimators using a cosine mean function at $\alpha = 0.25$.

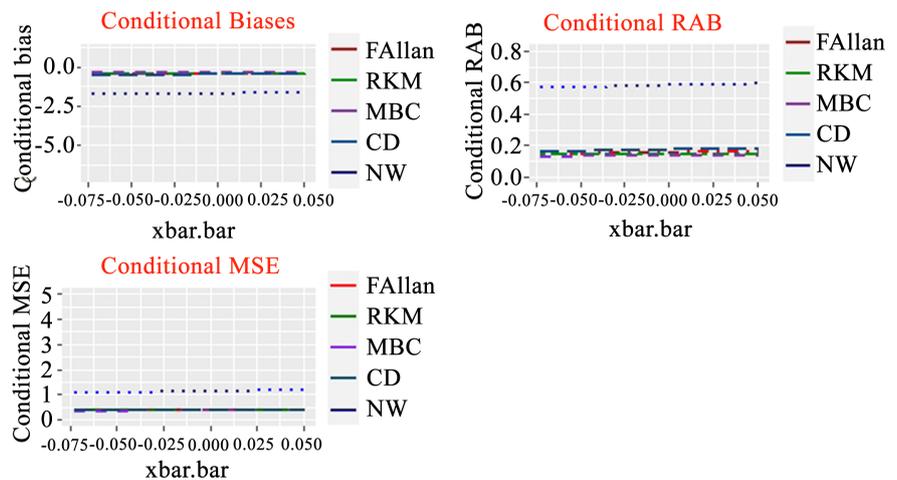


Figure 5. Conditional biases, RAB and MSE for the estimators using a cosine mean function at $\alpha = 0.5$.

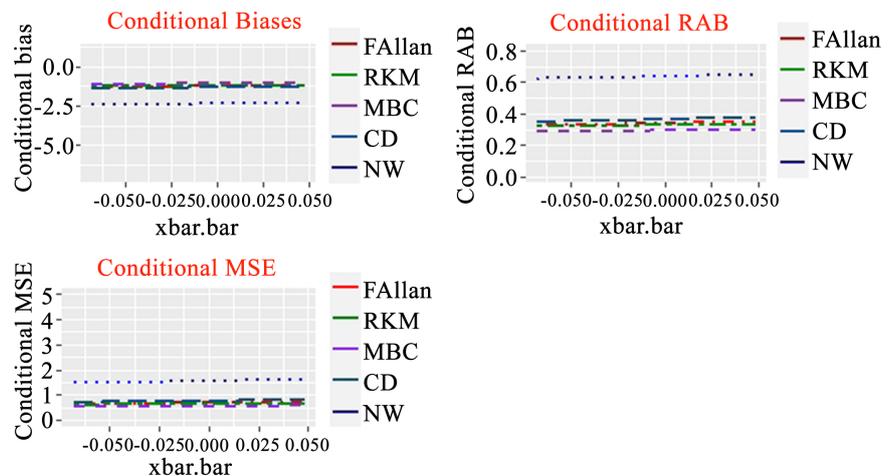


Figure 6. Conditional biases, RAB and MSE for the estimators using a cosine mean function at $\alpha = 0.75$.

proximity of plotted curves to the horizontal (no bias) line at 0.0 on the vertical axis. Interestingly, despite the rather entangled nature of some of the plots, estimator *MBCQE* emerges clearly as the least biased for nearly every group means of the means of auxiliary variables and quantile level. For the median, several estimators exhibit identical bias, and for most of the estimators, bias is not symmetrical with respect to quantile level.

Plots of Conditional MSE versus group means of the means of auxiliary variables similarly reveal coincident behavior for the quantiles. *MBCQE* and *RKMQE* produce generally the lowest MSE values. In particular, *MBCQE* yields the lowest MSE in most cases among all other estimators. *MBCQE* is consistently better than all other estimators for both bias and MSE. All of these estimators are asymptotically unbiased and they all exhibit MSE consistency in that the MSE values tend toward zero as sample size increases.

From the plots it can be seen that *MBCQE* and *RKMQE* performed equally better than all other estimators of the true quantile function and it can be seen that sample balancing does not affect the performance of the estimators.

5. Conclusions and Suggestions

In conclusion, using the results from **Table 2-4** and **Figures 1-6**, *MBCQE* was found to be an efficient estimator of the quantile function for finite population. *NWQE* was found to be very inefficient of all the estimators with large conditional bias, relative absolute bias and mean squared error compared to the other estimators. *MBCQE* can therefore be used in estimating quantile functions for various units in the population in various sectors of the economy. Finally, further work can be done on the construction of confidence intervals for the proposed estimator, and a researcher can investigate various bias correction strategies such as Adaptive Boosting and the Bootstrap bias reduction techniques in quantile function estimation.

Acknowledgements

Sincere thanks to the Pan-African University Institute of Basic Sciences, Technology and Innovation (PAUSTI) for funding this research.

Conflicts of Interest

The authors declare no conflicts of interest with respect to the publication of this article.

References

- [1] Wand, M.P., Marron, J.S. and Ruppert, D. (1991) Transformations in Density Estimation. *Journal of the American Statistical Association*, **86**, 343-353. <https://doi.org/10.1080/01621459.1991.10475041>
- [2] Gasser, T. and Müller, H.-G. (1979) Kernel Estimation of Regression Functions. In: *Smoothing Techniques for Curve Estimation*, Springer, Berlin, 23-68. <https://doi.org/10.1007/BFb0098489>

- [3] Müller, H.-G. (1991) Smooth Optimum Kernel Estimators near Endpoints. *Biometrika*, **78**, 521-530. <https://doi.org/10.1093/biomet/78.3.521>
- [4] Müller, H.-G. and Wang, J.-L. (1994) Hazard Rate Estimation under Random Censoring with Varying Kernels and Bandwidths. *Biometrics*, **50**, 61-76. <https://doi.org/10.2307/2533197>
- [5] John, R. (1984) Boundary Modification for Kernel Regression. *Communications in Statistics—Theory and Methods*, **13**, 893-900. <https://doi.org/10.1080/03610928408828728>
- [6] Schuster, E.F. (1985) Incorporating Support Constraints into Nonparametric Estimators of Densities. *Communications in Statistics—Theory and Methods*, **14**, 1123-1136. <https://doi.org/10.1080/03610928508828965>
- [7] Marron, J.S. and Ruppert, D. (1994) Transformations to Reduce Boundary Bias in Kernel Density Estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, **56**, 653-671. <https://doi.org/10.1111/j.2517-6161.1994.tb02006.x>
- [8] Jones, M.C. (1993) Simple Boundary Correction for Kernel Density Estimation Statistics and Computing. *Statistics and Computing*, **3**, 135-146. <https://doi.org/10.1007/BF00147776>
- [9] Fan, J. and Gijbels, I. (1992) Variable Bandwidth and Local Linear Regression Smoothers. *The Annals of Statistics*, **20**, 2008-2036. <https://doi.org/10.1214/aos/1176348900>
- [10] Fan, J. (1993) Local Linear Regression Smoothers and Their Minimax Efficiencies. *The Annals of Statistics*, **21**, 196-216. <https://doi.org/10.1214/aos/1176349022>
- [11] Devroye, L. (1985) Nonparametric Density Estimation. The L1 View. John Wiley, Hoboken.
- [12] Jones, L.O. and Nielsen, J. (1995) A Simple Bias Reduction Method for Density Estimation. *Biometrika*, **82**, 327-338. <https://doi.org/10.1093/biomet/82.2.327>
- [13] Denuit, M., Goovaerts, M. and Kaas, R. (2006) Actuarial Theory for Dependent Risks: Measures, Orders and Models. John Wiley & Sons, Hoboken. <https://doi.org/10.1002/0470016450>
- [14] Nadaraya, E.A. (1964) Some New Estimates for Distribution Functions. *Theory of Probability & Its Applications*, **9**, 497-500. <https://doi.org/10.1137/1109069>
- [15] Lio, Y. and Padgett, W. (1991) A Note on the Asymptotically Optimal Bandwidth for Nadaraya's Quantile Estimator. *Statistics & Probability Letters*, **11**, 243-249. [https://doi.org/10.1016/0167-7152\(91\)90150-P](https://doi.org/10.1016/0167-7152(91)90150-P)
- [16] Jones, M.C. (1992) Estimating Densities, Quantiles, Quantile Densities and Density Quantiles. *Annals of the Institute of Statistical Mathematics*, **44**, 721-727. <https://doi.org/10.1007/BF00053400>
- [17] Charpentier, A. and Oulidi, A. (2010) Beta Kernel Quantile Estimators of Heavy-Tailed Loss Distributions. *Statistics and Computing*, **20**, 35-55. <https://doi.org/10.1007/s11222-009-9114-2>
- [18] Sayah, A. (2012) Kernel Quantile Estimation for Heavy-Tailed Distributions.
- [19] Linton, O. and Nielsen, J.P. (1994) A Multiplicative Bias Reduction Method for Nonparametric Regression. *Statistics & Probability Letters*, **19**, 181-187. [https://doi.org/10.1016/0167-7152\(94\)90102-3](https://doi.org/10.1016/0167-7152(94)90102-3)
- [20] Malenje, B.M., Mokeira, W.O., Odhiambo, R. and Orwa, G.O. (2016) A Multiplicative Bias Corrected Nonparametric Estimator for a Finite Population Mean. *American Journal of Theoretical and Applied Statistics*, **5**, 317-325. <https://doi.org/10.11648/j.ajtas.20160505.21>
- [21] Hengartner, N., Matzner-Löber, E., Rouviere, L. and Burr, T. (2009) Multiplicative

Bias Corrected Nonparametric Smoothers.

- [22] Burr, T., Hengartner, N., Matzner-Løber, E., Myers, S. and Rouviere, L. (2010) Smoothing Low Resolution Gamma Spectra. *IEEE Transactions on Nuclear Science*, **57**, 2831-2840. <https://doi.org/10.1109/TNS.2010.2054110>
- [23] Stephane, K.T., Otieno, R.O. and Mageto, T. (2017) A Multiplicative Bias Correction for Nonparametric Approach and the Two Sample Problem in Sample Survey. *Open Journal of Statistics*, **7**, 1053. <https://doi.org/10.4236/ojs.2017.76073>
- [24] Onsongo, W.M., Otieno, R.O. and Orwa, G.O. (2018) Bias Reduction Technique for Estimating Finite Population Distribution Function under Simple Random Sampling without Replacement. *International Journal of Statistics and Applications*, **8**, 259-266.
- [25] Serfling, R. (1980) Approximation Theorems of Mathematical Statistics. John Wiley and Sons, New York. <https://doi.org/10.1002/9780470316481>
- [26] Francisco, C.A. (1987) Estimation of Quantiles and the Interquartile Range in Complex Surveys. *The Annals of Statistics*.
- [27] Francisco, C.A. and Fuller, W.A. (1991) Quantile Estimation with a Complex Survey Design. *The Annals of Statistics*, **19**, 454-469. <https://doi.org/10.1214/aos/1176347993>
- [28] Chambers, R.L. and Dunstan, R. (1986) Estimating Distribution Functions from Survey Data. *Biometrika*, **73**, 597-604. <https://doi.org/10.1093/biomet/73.3.597>
- [29] Bahadur, R.R. (1966) A Note on Quantiles in Large Samples. *The Annals of Mathematical Statistics*, **37**, 577-580. <https://doi.org/10.1214/aoms/1177699450>
- [30] Kuk, A.Y. and Mak, T. (1989) Median Estimation in the Presence of Auxiliary Information. *Journal of the Royal Statistical Society: Series B (Methodological)*, **51**, 261-269. <https://doi.org/10.1111/j.2517-6161.1989.tb01763.x>
- [31] Chen, J. and Wu, C. (2002) Estimation of Distribution Function and Quantiles Using the Model-Calibrated Pseudo Empirical Likelihood Method. *Statistica Sinica*, **12**, 1223-1239.