Journal of the Nigerian Association of Mathematical Physics
***Volume 56* (March – May *2020 Issue*), pp25 – 32**
© J. of NAMP

A MODIFIED ESTIMATOR FOR QUARTILE RANKED SET SAMPLING

Apantaku F.S., Olayiwola O.M., Akintunde A. A., Dawodu G.A. and Adeyemo R.G.

Department of Statistics, Federal University of Agriculture, Abeokuta, Nigeria

Abstract

Quartile Ranked Set Sampling (QRSS) is an approach to data collection for situations where taking the actual measurements for sample observations was difficult but mechanisms for ranking a set of sample units using their quartiles was relatively easy and reliable. It usually involves sampling in cycles and picking the quartiles per circle. This is time consuming and susceptible to sampling errors due to the number of cycles. This study derived a Shrinkage Quartile Rank Set Estimator (SQRSE) which is time efficient and enhances reduction of sampling error. The QRSS was used to select random samples without replacement and ranking was done. The process was repeated for 10 cycles to form 10 sample combinations and from each sample combination, first quartile (Q1) and third quartile (Q3) were determined. The first half values of Q1 were combined to the last half values of Q3 to form shrinkage rank set samples. The shrinkage estimator for QRSS was derived as a convolution of the Quartile Rank Set Estimator (QRSE) and Shrinkage Estimator. The statistical properties of SQRSE were examined and compared with QRSE. Shrinkage Quartile Rank Set Estimator was found to be more efficient and appropriate for error reduction in Quartile Ranked Set Sampling.

1.0 Introduction

A shrinkage estimator is an estimator that, either explicitly or implicitly, incorporates the effects of shrinkage. In loose terms this means that a naive or raw estimate is improved by combining it with other information. One general result is that many standard estimators can be improved, in terms of mean squared error (MSE), by shrinking them towards zero (or any other fixed constant value). Assume that the expected value of the raw estimate is not zero and consider other estimators obtained by multiplying the raw estimate by a certain parameter. A value for this parameter can be specified so as to minimize the MSE of the new estimate. For this value of the parameter, the new estimate will have a smaller MSE than the raw one. Thus it has been improved. An effect here may be to convert an unbiased raw estimate to an improved biased one.

1.1 Quartile Ranked Set Sampling

The Quartile Ranked Set Sampling (QRSS) method is carried out by selecting n random samples each of size n units from the population of interest and ranking the units in each sample with respect to a variable of interest. If the sample size n is even, select for measurement from the first n / 2 samples the $q_1(n+1)^{th}$ smallest ranked unit and from the second n / 2 samples the $q_3(n+1)^{th}$ smallest ranked unit. If the sample size n is odd, select for measurement from the first $(n+1)$ / 2 samples the $q_1(n+1)^{th}$ smallest ranked unit, from the last $(n+1)$ / 2 samples the $q_3(n+1)^{th}$ smallest ranked unit and from the remaining sample the median ranked unit. The cycle can be repeated r times if needed to get a sample of size nr units. Note that we always take the nearest integer of $q_1(n+1)^{th}$ and $q_3(n+1)^{th}$ where $q_1 = 0.25$ and $q_3 = 0.75$.

2.0 Literature Review

Ranked set sampling (RSS) is a method of collecting data that improves estimation by utilizing the sampler's judgment or auxiliary information about the relative sizes of the sampling units. Prior to quantifying the data, the researcher samples from the population and then ranks the sampled units based on his or her judgment about their relative sizes on the variable of interest.

In survey sampling settings, a logical method of ranking the units is to order them based on the values of an auxiliary variable correlated with the variable of interest. Using the ranks of the units, the researcher creates a subset of the sample and quantifies the variable of interest for units in the subset and the measurements from this subset are used to estimate population parameters. The method was first proposed in[1] to increase the accuracy of crop yield estimates without increasing the number of observations that need to be quantified. In [2], it was demonstrated that for comparable sample sizes, the RSS procedure results in more accurate parameter estimators than simple random sampling (SRS). Equivalently, RSS requires fewer measured observations than SRS to attain the same level of precision. The improvement in precision is because RSS adds structure to the data in the form of the sampler's ranking that is absent in SRS. This added structure is likened to stratifying the population prior to taking a SRS. Whereas, stratified SRS uses auxiliary information from the entire population, however, RSS uses auxiliary information from only the units in the initial sample; it does not require the availability of auxiliary information for all units in the population.

RSS was applied in [3] to regeneration surveys in areas direct-seeded to longleaf pine. It was noted that the means based on both of RSS and SRS methods were not significantly different, but the computed variances of the means were very different. Four experiments were conducted at Hurley (UK) during 1983 to investigate the performance of RSS relative to SRS for estimation of herbage mass in pure grass swards, and of herbage mass and clover content in mixed grass-clover swards [4]. RSS method was applied to estimate the mean of forest, grassland and other vegetation resources in [5]. The study of the performance of RSS in estimating milk yield based on 402 sheep can be seen in [6]. In [7], multistage RSS was used to estimate the average of Olive's yields in a field in West of Jordan. Investigation on the use of the RSS in estimating of the mean and median of a population using the crop production dataset from the United State Department of Agriculture was done in [8].

Corresponding Author**:** Akintunde A.A., Email: akabosede@gmail.com, Tel: +2348134592275

It was found that the gain in efficiency for mean estimation using RSS is better for symmetric distribution than asymmetric distribution, and vice versa in the case of median estimation. In [9], RSS procedure was applied in market and consumer surveys. Application of RSS method to estimate the population mean and the ratio using a real data set on body measurement was done in [10]. The authors used the data of the weight and height of 507 individuals. McIntyre's method was coined as RSS and applied for estimating the weights of browse and herbage in a pine-hardwood forest of east Texas, USA [11]. For more about applications of RSS see [12].
Cost model for comparing RSS to stratified simple random sampling was developed in [13]. In [14], a method for determining the optimal set size taking into account the various costs associated with RSS was provided. Other studies on ranked set sampling include the works of [15-18]. Thorough reviews of the RSS literature can be found in [19].

3.0 Methodology

3.1 Survey Design

Survey design refers to everything that needs to be done to ensure a successful design. The term sample design refers to the procedure in which a sample is obtained from a given population. It is also known as sampling plan. There are different sampling techniques but the one adopted in this study is quartile ranked set sampling.

3.2 Estimation of Population Mean of Quartile Ranked Set Sampling [20]

Let $X_1, X_2,\ldots,X_n$ be n independent random variables from a probability density function $f(x)$, with mean and variance. The SRS estimator of the population mean based on a sample of size n is given by;

$$\bar{X}_{SRS} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad (1)$$

With variance

$$Var\left(\bar{X}_{SRS}\right) = \frac{\sigma^2}{n} \quad (2)$$

The RSS estimator of the population mean is given by

$$\bar{X}_{SRS} = \frac{1}{n}\sum_{i=1}^{n} X_{i(i)} \quad (3)$$

and the variance is

$$Var\left(\bar{X}_{SRS}\right) = \frac{\sigma^2}{n} - \frac{1}{n^2}\sum_{i=1}^{n}\left(\mu_{(i)} - \mu\right)^2 \quad (4)$$

Where $\mu_{(i)}$ is the mean of the ith order statistics, $X_{(i)}$ for a sample of size n.

Quartile ranked set sampling estimator of the population mean when n is even is defined as

$$\bar{X}_{QRS} = \frac{1}{n}\left[\sum_{i=1}^{n} X_{i(q_1(n+1))} + \sum_{i=\frac{n+2}{2}}^{n} X_{i(q_3(n+1))}\right] \quad (5)$$

With variance

$$Var\left(\bar{X}_{QRS}\right) = \frac{1}{n^2}\left[\sum_{i=1}^{n} \sigma^2_{i(q_1)} + \sum_{i=\frac{n+2}{2}}^{n} \sigma^2_{i(q_3)}\right] \quad (6)$$

Quartile ranked set sampling estimator of the population mean when n is odd is defined as

$$\bar{X}_{QRS} = \frac{1}{n}\left[\sum_{i=1}^{\frac{n-1}{2}} X_{i(q_1(n+1))} + \sum_{i=\frac{n+3}{2}}^{n} X_{i(q_3(n+1))} + X_{\frac{n+1}{1}\left(\frac{n+1}{2}\right)}\right] \quad (7)$$

With variance

$$Var\left(\bar{X}_{QRS\,2}\right) = \frac{1}{n^2}\left[\sum_{i=1}^{\frac{n-1}{2}} \sigma^2_{i(q_1)} + \sum_{i=\frac{n+3}{2}}^{n} \sigma^2_{i(q_3)} + \sigma^2_{\frac{n+1}{2}}\right] \quad (8)$$

3.3 Existing Estimator

Consider a ratio estimator

The population mean $\bar{Y}$ is estimated by

$$\bar{Y}_R = \frac{\bar{y}}{\bar{x}} \cdot \bar{X} \quad (9)$$

The population total $\hat{Y}$ is estimated by

$$\bar{Y} = \hat{R}X \quad (10)$$

where $\hat{R} = \frac{\bar{y}}{\bar{x}}$

$\hat{R}$ is a biased estimator

$\hat{R}$ is given as:

$$MSE\left(\hat{R}\right) = E\left(\hat{R}-R\right)^2 = V(\hat{R}) + (Bias)^2 \quad (11)$$

Assuming n is large; the bias is negligible for calculation.
Hence

$$MSE\left(\hat{R}\right) = \frac{1}{\bar{x}^2}\left[V(\bar{y}) + R^2V(\bar{x}) - 2R\,\mathrm{cov}(\bar{x},\bar{y})\right] \quad (12)$$

The computational formula is;
Where

$$\bar{y} = \frac{1}{n}\left[\sum_{i=1}^{n} y_{i(q_1(n+1))} + \sum_{i=\frac{n+2}{2}}^{n} y_{i(q_3(n+1))}\right] \quad (13)$$

$$\bar{x} = \frac{1}{n}\left[\sum_{i=1}^{n} x_{i(q_1(n+1))} + \sum_{i=\frac{n+2}{2}}^{n} x_{i(q_3(n+1))}\right] \quad (14)$$

$\bar{Y}$ is the population mean

3.4 Shinkage Estimator

Suppose a population parameter t can be estimated by using an estimator $\hat{t}$ with mean square error $MSE\left(\hat{t}\right)$. A general shrinkage estimator was defined in [21] as:

$$\hat{t}_s = \frac{\hat{t}}{1 + T^{-2}MSE\left(\hat{t}\right)} \quad (15)$$

Where $\hat{t}$ is any available estimator of parameter T. The minimum MSE is given as:

$$mse\left(\hat{t}_s\right) = \frac{MSE\left(\hat{t}\right)}{1 + T^{-2}MSE\left(\hat{t}\right)} \quad (16)$$

3.5 Quartile Ranked Set

Given that quartile rank set sampling estimator of the population mean when *n* is even is defined as

$$\hat{t}_1 = \frac{1}{n}\left[\sum_{i=1}^{\frac{n}{2}} X_{i(q_1(n+1))} + \sum_{i=\frac{n+2}{2}}^{n} X_{i(q_3(n+1))}\right] \quad (17)$$

Then when *n* is odd

$$\hat{t}_2 = \frac{1}{n}\left[\sum_{i=1}^{\frac{n-1}{2}} X_{i(q_1(n+1))} + \sum_{i=\frac{n+3}{2}}^{n} X_{i(q_3(n+1))} + X_{\frac{n+1}{2}\left(\frac{n+1}{2}\right)}\right] \quad (18)$$

$$E\left(\hat{t}_1\right) = E\left[\frac{1}{n}\left(\sum_{i=1}^{\frac{n}{2}} X_{i(q_1(n+1))} + \sum_{i=\frac{n+2}{2}}^{n} X_{i(q_3(n+1))}\right)\right]$$

$$= \frac{1}{n}\left[\sum_{i=1}^{\frac{n}{2}} E\left(X_{i(q_1(n+1))}\right) + \sum_{i=\frac{n+2}{2}}^{n} E\left(X_{i(q_3(n+1))}\right)\right]$$

$$= \frac{1}{n}\left[\sum_{i=1}^{\frac{n}{2}} \mu_{(q_1)} + \sum_{i=\frac{n+2}{2}}^{n} \mu_{(q_3)}\right] \quad (19)$$

where $\mu_{(q_1)}$ and $\mu_{(q_3)}$ are the means of the order statistics corresponding to the first and third quartiles respectively. Since the distribution is symmetric about μ, then, $\mu_{(q_1)} + \mu_{(q_3)} = 2\mu$. Therefore, we have

$$E\left(\hat{t}_1\right) = \frac{1}{n}\left(\frac{n}{2}\mu_{(q_1)} + \frac{n}{2}\mu_{(q_3)}\right)$$

$$= \frac{1}{n}\left(\frac{n}{2}\left(\mu_{(q_1)} + \mu_{(q_3)}\right)\right) \quad (20)$$

$$= \frac{1}{n}\left(\frac{n}{2}(2\mu)\right)$$

$$= \mu$$

$$E(\hat{t}_2) = E\left[\frac{1}{n}\left(\sum_{i=1}^{\frac{n-1}{2}} X_{i(q_1(n+1))} + \sum_{i=\frac{n+3}{2}}^{n} X_{i(q_3(n+1))} + X_{\frac{n+1}{2}\left(\frac{n+1}{2}\right)}\right)\right]$$

$$= \frac{1}{n}\left[\sum_{i=1}^{\frac{n-1}{2}} E\left[X_{i(q_1(n+1))}\right] + \sum_{i=\frac{n+3}{2}}^{n} E\left[X_{i(q_3(n+1))}\right] + E\left[X_{\frac{n+1}{2}\left(\frac{n+1}{2}\right)}\right]\right] \tag{21}$$

$$= \frac{1}{n}\left[\sum_{i=1}^{\frac{n-1}{2}} \mu_{(q_1)} + \sum_{i=\frac{n+3}{2}}^{n} \mu_{(q_3)} + \mu_{\left(\frac{n+1}{2}\right)}\right]$$

where $\mu_{(q_1)}$ is the mean of the first quartile for first $\left(\frac{n-1}{2}\right)$ samples and $\mu_{(q_3)}$ is the mean of the third quartile for the last $\left(\frac{n-1}{2}\right)$ samples. Since the distribution is symmetric about μ , then $\mu_{(q_1)} + \mu_{(q_3)} = 2\mu$. Therefore,

$$E(\hat{t}_2) = \frac{1}{n}\left(\frac{n-1}{2}\mu_{(q_1)} + \frac{n-1}{2}\mu_{(q_3)} + \mu_{\left(\frac{n+1}{2}\right)}\right) \tag{22}$$

$$= \frac{1}{n}\left(\frac{n-1}{2}\left(\mu_{(q_1)} + \mu_{(q_3)}\right) + \mu_{\left(\frac{n+1}{2}\right)}\right)$$

$$= \frac{1}{n}\left(\frac{n-1}{2}(2\mu) + \mu\right)$$

$$= \mu$$

$$Var(\hat{t}_1) = Var\left[\frac{1}{n}\left(\sum_{i=1}^{\frac{n}{2}} X_{i(q_1(n+1))} + \sum_{i=\frac{n+2}{2}}^{n} X_{i(q_3(n+1))}\right)\right]$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{\frac{n}{2}} Var\left(X_{i(q_1(n+1))}\right) + \sum_{i=\frac{n+2}{2}}^{n} Var\left(X_{i(q_3(n+1))}\right)\right] \tag{23}$$

$$= \frac{1}{n^2}\left(\sum_{i=1}^{\frac{n}{2}} \sigma^2_{(q_1)} + \sum_{i=\frac{n+2}{2}}^{n} \sigma^2_{(q_3)}\right)$$

$$Var(\hat{t}_2) = Var\left[\frac{1}{n}\left(\sum_{i=1}^{\frac{n-1}{2}} X_{i(q_1(n+1))} + \sum_{i=\frac{n+3}{2}}^{n} X_{i(q_3(n+1))} + X_{\frac{n+1}{2}\left(\frac{n+1}{2}\right)}\right)\right]$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{\frac{n-1}{2}} Var\left[X_{i(q_1(n+1))}\right] + \sum_{i=\frac{n+3}{2}}^{n} Var\left[X_{i(q_3(n+1))}\right] + Var\left[X_{\frac{n+1}{2}\left(\frac{n+1}{2}\right)}\right]\right]$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{\frac{n-1}{2}} \sigma^2_{i(q_1)} + \sum_{i=\frac{n+3}{2}}^{n} \sigma^2_{i(q_3)} + \sigma^2_{\left(\frac{n+1}{2}\right)}\right] \tag{24}$$

When n is even, for the proposed estimator is

$$\hat{t}_{SQRS} = \frac{\frac{1}{n}\left[\sum_{i=1}^{\frac{n}{2}} X_{i(q_1(n+1))} + \sum_{i=\frac{n+2}{2}}^{n} X_{i(q_3(n+1))}\right]}{1 + t^{-2}\frac{1}{n^2}\left[\sum_{i=1}^{\frac{n}{2}} \sigma^2_{i(q_1)} + \sum_{i=\frac{n+2}{2}}^{n} \sigma^2_{i(q_3)}\right]} \tag{25}$$

When n is odd for the proposed estimator

$$\hat{t}_{SQRS} = \frac{\frac{1}{n}\left[\sum_{i=1}^{\frac{n-1}{2}} X_{i(q_1(n+1))} + \sum_{i=\frac{n+3}{2}}^{n} X_{i(q_3(n+1))} + X_{\frac{n+1}{2}\left(\frac{n+1}{2}\right)}\right]}{1 + t^{-2}\frac{1}{n^2}\left[\sum_{i=1}^{\frac{n-1}{2}} \sigma^2_{i(q_1)} + \sum_{i=\frac{n+3}{2}}^{n} \sigma^2_{i(q_3)} + \sigma^2_{\frac{n+1}{2}}\right]} \tag{26}$$

Where $\hat{t}$ is any available estimator of parameter t, in this case is the estimate of the mean of Quartile Rank Set. Then minimum MSE of the proposed estimator for both even and odd n sample is given as:

$$mse(\hat{t}_s) = \frac{mse(t_{QRS})}{1 + t^{-2} mse(t_{QRS})} \quad (27)$$

$$mse(\hat{t}_s) = \frac{\frac{1}{n^2}\left(\sum_{i=1}^{\frac{n}{2}} \sigma^2_{(q_1)} + \sum_{i=\frac{n+2}{2}}^{n} \sigma^2_{(q_3)}\right)}{1 + t^{-2}\frac{1}{n^2}\left(\sum_{i=1}^{\frac{n}{2}} \sigma^2_{(q_1)} + \sum_{i=\frac{n+2}{2}}^{n} \sigma^2_{(q_3)}\right)} \quad (28)$$

$$mse(\hat{t}_s) = \frac{\frac{1}{n^2}\left[\sum_{i=1}^{\frac{n-1}{2}} \sigma^2_{i(q_1)} + \sum_{i=\frac{n+3}{2}}^{n} \sigma^2_{i(q_3)} + \sigma^2_{\frac{n+1}{2}}\right]}{1 + t^{-2}\frac{1}{n^2}\left[\sum_{i=1}^{\frac{n-1}{2}} \sigma^2_{i(q_1)} + \sum_{i=\frac{n+3}{2}}^{n} \sigma^2_{i(q_3)} + \sigma^2_{\frac{n+1}{2}}\right]} \quad (29)$$

3.6 Derivation of the proposed estimator

Suppose $\hat{t}$ is an estimator of parameter of T with MSE ($\hat{t}) = E(\hat{t} - \overline{Y})^2$, suppose we want to shrink MSE($\hat{t}$), according to [21] then

$$\hat{t}_s = \frac{\hat{t}}{1 + T^{-2} MSE(\hat{t})} \quad (30)$$

Which is the shrinkage estimator for ($\hat{t}$) and can be written as:

$$\hat{t}_s = d\hat{t} \quad (31)$$

where $d = \dfrac{1}{1 + T^{-2} MSE(\hat{t})}$ the shrinkage constant with (32)

Mean square error for ($\hat{t}$) is;

$$mse(\hat{t}_s) = \frac{MSE(\hat{t})}{1 + T^{-2} MSE(\hat{t})} \quad (33)$$

$$mse(\hat{t}_s) = d * MSE(\hat{t}) \quad (34)$$

Shrinking of estimator causes loss of information but the amount of information that may be lost due to shrinkage was not catered for in [21]. Suppose we can control the loss of information by introducing an exponentiated parameter such that $\hat{t}_s$ becomes

$$\hat{t}_s = d^{\alpha}\hat{t} \quad (35)$$

where $-1 < \alpha < 1$

Now considering a ratio estimator in quartile ranked set sampling we have,

$$\hat{t} = \frac{\overline{y}}{\overline{x}}\overline{X} \quad (36)$$

$$MSE(\hat{t}) = S_y^2 - 2RS_{yx} + R^2 S_x^2 \quad (37)$$

Where

$$S_y^2 = \frac{1}{n^2}\left(\sum_{i=1}^{\frac{n}{2}} \sigma^2_{y(q_1)} + \sum_{i=\frac{n+2}{2}}^{n} \sigma^2_{y(q_3)}\right) \quad (38)$$

$$S_x^2 = \frac{1}{n^2}\left(\sum_{i=1}^{\frac{n}{2}} \sigma^2_{x(q_1)} + \sum_{i=\frac{n+2}{2}}^{n} \sigma^2_{x(q_3)}\right) \quad (39)$$

$$S_{xy} = \frac{1}{n^2}\left(\sum_{i=1}^{\frac{n}{2}} \sigma_{xy(q_1)} + \sum_{i=\frac{n+2}{2}}^{n} \sigma^2_{xy(q_3)}\right) \quad (40)$$

Therefore, the mean square error for the existing estimator can be written as;

$$MSE(\hat{t}) = \left[\frac{1}{n^2}\left(\sum_{i=1}^{\frac{n}{2}} \sigma^2_{y(q_1)} + \sum_{i=\frac{n+2}{2}}^{n} \sigma^2_{y(q_3)}\right) - 2R\frac{1}{n^2}\left(\sum_{i=1}^{\frac{n}{2}} \sigma_{xy(q_1)} + \sum_{i=\frac{n+2}{2}}^{n} \sigma^2_{xy(q_3)}\right) + R^2\frac{1}{n^2}\left(\sum_{i=1}^{\frac{n}{2}} \sigma^2_{x(q_1)} + \sum_{i=\frac{n+2}{2}}^{n} \sigma^2_{x(q_3)}\right)\right] \quad (41)$$

$$\hat{t}_s = d^{\alpha}\frac{\overline{y}}{\overline{x}}\overline{X} \quad (42)$$

which is the shrinkage ratio estimator in quartile ranked set sampling, the Mean Square Error is derived as follows

$$MSE\ (\hat{t}_s) = E(\hat{t}_s - \bar{Y})^2 \quad (43)$$

$$MSE\ (\hat{t}_s) = E(\hat{t}_s - \bar{Y})^2$$
$$= E(d^{\alpha}\hat{t} - \bar{Y})^2$$
$$= (d^{\alpha})^2 MSE\ (\hat{t}) + \bar{Y}^2 (d^{\alpha} - 1)^2 \quad (44)$$

Let $K = d^{\alpha}$

Bias of this estimator was found to be

$$E(\hat{t} - \bar{Y}) = E(K^*\hat{t} - \bar{Y}) = E\left(K^* \frac{\bar{y}}{\bar{x}}\bar{X} - \bar{Y}\right)$$

$$= \bar{X}\ E\left(\frac{K^*\bar{y} - R\bar{x}}{\bar{x}}\right)$$

$$= \bar{Y}(K^* - 1) + \frac{1}{\bar{X}}\left(RS_x^2 - K\,S_{yx}\right) \quad (45)$$

The derivative of MSE was taken with respect to K and equate to zero to find the optimum value of K

$$\frac{\partial MSE\ (\hat{t}_s)}{\partial K} = 2K\left(S_y^2 - 2RS_{yx} + R^2 S_x^2\right) + 2(K-1)\ \bar{Y}^2 \quad (46)$$

$$K = d^{\alpha} = \frac{\bar{Y}^2}{\bar{Y}^2 + \left(S_y^2 - 2RS_{yx} + R^2 S_x^2\right)} \quad (47)$$

where $-1 < \alpha < 1$ and $0 < d^{\alpha} < 1$

Mean square error for the proposed estimator

$$MSE\ (\hat{t}_s) = 2K\left[\frac{1}{n^2}\left(\sum_{i=1}^{\frac{n}{2}} \sigma^2_{y(q_1)} + \sum_{i=\frac{n+2}{2}}^{n} \sigma^2_{y(q_3)}\right) - 2R\frac{1}{n^2}\left(\sum_{i=1}^{\frac{n}{2}} \sigma_{xy(q_1)} + \sum_{i=\frac{n+2}{2}}^{n} \sigma^2_{xy(q_3)}\right) + R^2\frac{1}{n^2}\left(\sum_{i=1}^{\frac{n}{2}} \sigma^2_{x(q_1)} + \sum_{i=\frac{n+2}{2}}^{n} \sigma^2_{x(q_3)}\right)\right] + \bar{Y}^2 2(K-1) \quad (48)$$

4.0 Analysis

The data used were the volume of petroleum product (in litre) used in 2017 by the 36 states in Nigeria for two quarters from the publication of National Bureau of statistics (NBS). The first quarter is our auxiliary variable (X) while the recent quarter is our variable of interest (Y). Quartile Rank Set Sampling (QRSS) was used to select random samples without replacement and ranking was done. The process was repeated for 10 cycles to form 10 sample combinations. From the sample combinations, first quartile (Q1) and third quartile (Q3) were determined. The first half values of Q1 were combined to last half values of Q3 to form shrinkage rank set samples. The repeated 10 cycles and ranking gave 10x10 sample combinations.

4.1 Descriptive Analysis

Descriptive statistics shows that Lagos state used 270,361,533 litres of petrol which is the highest used by any state in a month while Jigawa state used 3,346,985 litres which is the least volume. Box plot (figure 1) showed that the volume of petrol used in Lagos, Kano and Niger states stood as outliers. Normal Quartile-Quartile (Q-Q) plot (figure 2) was used to check the normality of the data sets.

Table 1: Summary of Descriptive Statistics

Descriptive statistics	Value
	2017
Count	37
Mean	40,777,421.81
Sample Standard Deviation	46,995,078.76
Sample Variance	2,208,537,427,782,460.00
Minimum	3,346,985
Maximum	270,361,533
Range	267,014,548

Table 2: Quartiles of the Ranked Set Sample of volume of petroleum for first and third quarters

Y		X	
1st quartile	**3rd quartile**	**1st quartile**	**3rd quartile**
2,441,019	3,512,973	3,346,985	4,221,975
3,224,009	6,546,986	4,221,975	6,651,74
3,512,973	8,625,561	5,709,534	11,369,871
6,450,990	9,365,889	6,027,928	14,331,631
8,378,785	16,545,807	6,027,928	17,192,849
13,407,971	18,593,015	13,642,201	21,597,967
13,588,237	19,427,397	15,334,345	22,838,485
13,588,237	24,636,525	20,496,688	42,472,778
17,725,194	36,155,349	16,971,240	27,441,118
19,427,397	42,931,616	16,971,240	36,805,364

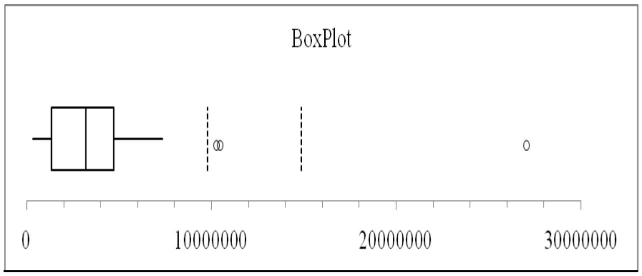

Figure 1: Box plot showed the outliers for the volume of petrol used

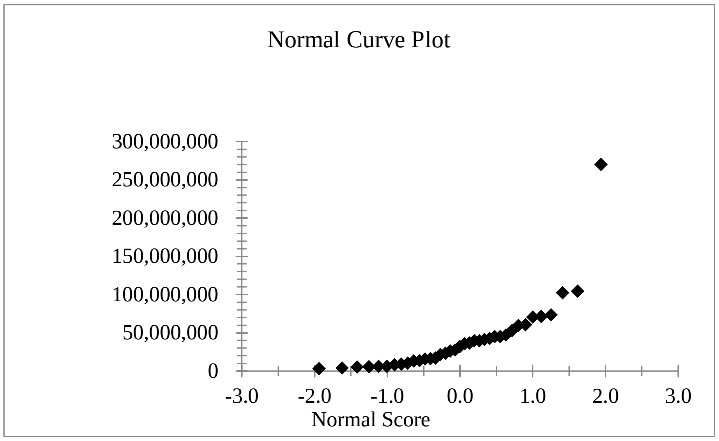

Figure 2: Normal Quartile-Quartile (Q-Q) plot to check the normality of the data set.

Then we select the final samples by taking first halve of first quartile and the remaining last halve of the third quartile

Table 3: Shrinkage quartile ranked set samples

Y	**X**
2,441,019	3,346,985
3,224,009	4,221,975
3,512,973	5,709,534
6,450,990	6,027,928
8,378,785	6,027,928
18,593,015	21,597,967
19,427,397	22,838,485
24,636,525	742,472,778
36,155,349	27,441,118
42,931,616	36,805,364

Table 4: Parameters and Estimates

Parameters	**Estimates**
$\bar{Y}$	40777421.81
$\bar{X}$	41724882.38
R	0.977293
S_y^2	2148847227031590.00
S_x^2	1977148542913970.00
S_{xy}	1.98092×10^{14}
$\hat{t}$	7890527.88
$\hat{t}_s$	7180380.37

$$MSE\ (\hat{t}) = \left(S_y^2 - 2RS_{yx} + R^2 S_x^2\right) = 1.65346 \times 10^{14} \text{ (Existing Estimator)}$$

$$K = d^{\alpha} = \frac{\bar{Y}^2}{\bar{Y}^2 + \left(S_y^2 - 2RS_{yx} + R^2 S_x^2\right)} = 0.91 \text{ (when } \alpha = 1)$$

$$MSE\ (\hat{t}_s) = K^2\left(S_y^2 - 2RS_{yx} + R^2 S_x^2\right) + \bar{Y}^2 (K-1)^2 = 1.50392 \times 10^{14} \text{ (Proposed Estimator)}$$

$$Efficiency = \frac{MSE\ (\hat{t})}{MSE\ (\hat{t}_s)} = \frac{1.65346 \times 10^{14}}{1.50392 \times 10^{14}} = 1.01 > 1$$

The shrinkage constant (K) is 0.91, the mean square error for existing estimator is 1.65 while mean square error for proposed estimator is 1.50; its efficiency is 1.01, which is 101% gain in information. Therefore, the proposed estimator is more efficient than the existing estimator in quartile ranked set sampling.

5.0 Conclusion

A shrinkage quartile rank estimator which was derived as a convolution of the quartile rank set estimator and shrinkage estimator for population mean was proposed and its properties were determined. A comparative study of the proposed estimator was made with relevant estimators. From the theoretical discussion and numerical examples, we infer that the proposed estimator is more efficient than the existing estimator by [21]. The Shrinkage estimator which is a limiting function was introduced which tends to zero when multiplied with the existing quartile rank set estimator which shrinks mean square error towards zero.

The relative efficiencies obtained from both existing and proposed mean square errors indicated that the proposed estimator is more efficient.

Normal Quartile-Quartile (Q-Q) plot were used to check the normality of the data sets. It was confirmed by the Normal Q-Q plot that the data was not Normal which is as a result of the volume used in various states. Box plot was used to check for outliers and it showed that the volume of petroleum product used in some states stood as outliers.

REFERENCES

[1] McIntyre, G. A. 1952. A method for unbiased selective sampling, using ranked sets. *Australian Journal of Agricultural Research,*Vol. 3, pp. 385-390.

[2] Dell, T. R. and Clutter, J. L. 1972. Ranked set sampling theory with order statistics background,Biometrics,Vol. 28, pp.545-555.

[3] Evans, M. J. 1967. Application of ranked set sampling to regeneration, Surveys in areas direct-seeded to long leaf pine. Master Thesis, school for Forestry and Wild-life Management, Louisiana state University, Baton Rouge, Louisiana.

[4] Cobby, J.M., Ridout, M.S., Bassett, P.J. & Large, R.V. 1985. An investigation into the use of ranked set sampling on grass and grass-clover swards. Grass and Forage Science, Vol. 40, pp. 257-63.

[5] Johnson, G.D., Paul, G.P. & Sinha, A.K. 1993. Ranked set sampling for vegetation research. Abstracta Botanica, Vol.17, pp. 87-102.

[6] Al-Saleh, M.F.; and Al-Shrafat, K. 2001. Estimation of average milk yield using ranked set sampling Environmetrics 12(4):395 – 399

[7] Al-Saleh, M. F. and Al-Omari, A.I. 2002. Multistage ranked set sampling. *Journal of Statistical Planning and Inference,* Vol.102, pp. 273-286.

[8] Husby, C. E., Stasny, E. A., and Wolfe, D. A. 2005.An application of ranked set sampling for mean and median estimation using crop production data. *Journal of Agricultural, Biological, and Environmental Statistics*,Vol.10, pp.354-373.

[9] Kowalczyk, B. 2004. Ranked set sampling and its applications in finite population studies. *Statistics in Transition*, Vol.6. pp.1031-1046.

[10] Ganeslingam, S. & Ganesh, S. 2006. Ranked set sampling versus simple random sampling in the estimation of the mean and the ratio. *Journal of Statistics and Management Systems*. Vol. 2, pp. 459-472.

[11] Halls, L. K. and Dell, T. R. 1966. Trial of ranked-set sampling for forage yields. ,*Forest Science,*Vol.12, pp. 22-26.

[12] Dong, X.F., Cui, L.R. & Liu, F.Y. 2012. A further study on reliable life estimation under ranked set sampling. Communications in Statistics –Theory and Methods, Vol. 21, pp. 3888-3902.

[13] Amarjot, Patil, G. P., Shirk, S. J., and Taillie, C. 1996. Environmental sampling with a concomitant variable: A comparison between ranked set sampling and stratified simple random sampling. *Journal of Applied Statistics,* Vol. 23, pp. 231-255.

[14] Nahhas, R. W., Wolfe, D. A., and Chen, H. 2002. Ranked set sampling: Cost and optimal set size. *Biometrics, Vol.* 58, pp. 964-971.

[15] David, H.A. & Levine, D.N. 1972. Ranked set sampling in the presence of judgment error. Biometrics, Vol. 28, pp. 553–555.

[16] Muttlak, H.A. & McDonald, L.L. 1990. Ranked set sampling with respect to concomitant variables and with biased probability of selection. Communication in Statistics-Theory and Methods, Vol. 19, No.1, pp. 205-219

[17] Lacayo, H., Neerchal, N.K. & Sinha, B.K. 2002. Ranked set sampling from a dichotomous population. Journal of Applied Statistical Science, Vol.1, No. 1 pp. 83-90.

[18] Chen, H., Stasny, E. A., and Wolfe, D. A. 2005. Ranked set sampling for efficient estimation of a population proportion. *Statistics in Medicine,*Vol. 24, pp. 3319-3329.

[19] Patil, G., Sinha, A., and Taillie, C. 1999. Ranked set sampling: a bibliography. *Environmental and Ecological Statistics,* Vol. 6, pp. 91-98.

[20] Al-Omari, A. I. 2012. "Ratio estimation of the population mean using auxiliary information in simple random sampling and median ranked set sampling," *Statistics & Probability Letters,* Vol.82, No.1883-1890.

[21] Shahbaz, M.O. and Hanif. M. 2009. A General Shrinkage Estimator in Survey Sampling. World Applied Sciences Journal 7 (5): 593-596.