

ESTIMATION OF THE POPULATION DISTRIBUTION FUNCTION USING VARIED L RANKED SET SAMPLING

MOHAMED S. ABDALLAH* 

Abstract. A generalized ranked set sampling (RSS) plan has recently been provided in the literature called varied L RSS (VLRSS). It is shown that VLRSS encompasses several existing RSS variations and also it efficiently estimates the population mean. In this article, we extend the work and consider estimating the cumulative distribution function (CDF) using VLRSS. Three new CDF estimators are proposed and their asymptotic properties are also theoretically investigated. Taking into account the information supported by the unmeasured sampling units, we also propose a general class of CDF estimators. Using small Monte Carlo experiments, we study the behavior of the proposed CDF estimators with respect to the conventional CDF estimator under RSS. It is found that the conventional RSS-based CDF is outperformed by at least one of VLRSS-based CDF estimators in most of the considered cases. Finally, an empirical example is utilized to illustrate the potential application of the proposed estimators.

Mathematics Subject Classification. 62D05, 62F03.

Received June 22, 2021. Accepted February 04, 2022.

1. INTRODUCTION

One of the most popularly used sampling methods is Ranked Set Sampling (RSS). As, it is very helpful in the situations where ranking can be done at a cost that is not problematic relative to the cost of making accurate quantification on the interested variable. For instance, assume it is required to estimate the average height of trees in a forest. Hence, a small set of selected trees can be ranked visually with respect to their heights rather than actual measurements of the height of trees. Although RSS was firstly applied in agriculture field by McIntyre [24], it has subsequently been adopted for many other branches including environmental monitoring [20], auditing [16] and medicine [23, 34].

Due to the powerful and the applicability of RSS, many studies have examined the performance of almost every standard statistical problem under RSS and its variations. We now shortly review some newly important studies as follows: Mahdizadeh and Zamanzade [22] developed a kernel-based estimator of a dynamic reliability statistic based on RSS. Zamanzade and Mahdizadeh [31, 32] addressed the population proportion under RSS and its variations. Frey and Zhang [14, 15] constructed a confidence interval for population proportion under RSS. Göçoğlu and Demirel [17] considered the point estimation for population proportion under RSS. Mahdizadeh and Zamanzade [23] concerned with the problem of interval estimation for the population quantiles under

Keywords. Ranked Set Sampling, varied L Ranked Set Sampling, distribution function, relative precision, missing data approach.

Department of Quantitative Techniques, Faculty of Commerce, Aswan University, Aswan, Egypt.

*Corresponding author: mohamed.abdallah@com.aswu.edu.eg; statisticsms.2010@gmail.com

© The authors. Published by EDP Sciences, ROADEF, SMAI 2022

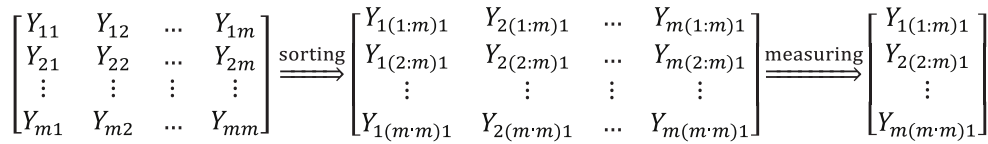


FIGURE 1. Demonstration of RSS procedure for one cycle.

RSS. Morabbi and Razmkhah [25] used two modified RSS schemes to propose new estimators for population quantiles. Zamanzade *et al.* [35] studied the problem of estimating the mean residual life under RSS. Chen *et al.* [10] introduced maximum likelihood estimator of location parameter under moving extremes RSS. Hassan *et al.* [19] considered the stress-strength reliability for generalized inverted exponential distribution using RSS and its variation methods.

To get RSS, randomly draw m sets each of size m units. Then rank the m sets separately in ascending way without numerically measuring the unit. The ranking process can be done using auxiliary variable or subjectively. Finally, quantify exactly the i th smallest unit from i th set, where $i = 1, 2, \dots, m$. These steps are considered as a cycle. Repeat, if needed, the cycle r times to obtain RSS of size $n = rm$. The RSS steps at $r = 1$ are illustrated in Figure 1, where Y_{ij} is i th unit from the j th sample. It may be important to emphasis that the ranking process may be inaccurate and contains error, this situation is called imperfect ranking. Oppositely, perfect ranking is a situation in which the ranks of the sampling units are done without errors. Symbolically, RSS is denoted by $\{Y_{i(i:m)j} : i = 1, \dots, m; j = 1, \dots, r\}$. Here, $Y_{i(i^*:m)j}$ is the i^* order statistic from the i th sample of size m in the j th cycle. The rounded parentheses of $Y_{i(i:m)j}$ implies that the perfect ranking situation is assumed. In such cases, for a fixed j , $Y_{i(i:m)j}$ follows the distribution of the i th order statistic from a sample of size m . While if the imperfect ranking situation was assumed, then the rounded parentheses will be replaced with squared ones. On the other hand, if the number of the selected sampling items is fixed across the r cycles. This case is known as balanced RSS. However unbalanced RSS allows the drawn sampling items to vary from one cycle to another. Here, balanced RSS is adopted through this paper. Let Y_1, Y_2, \dots, Y_n be a random sample of size n with a continuous probability density function $f(y)$ and cumulative distribution function (CDF) $F(y)$. Let $Y_{i(1:m)j}, Y_{i(2:m)j}, \dots, Y_{i(m:m)j}$ be the order statistics of the i th sample ($i = 1, \dots, m$) in the j th cycle ($j = 1, \dots, r$). Then $\{Y_{i(i:m)j} : i = 1, \dots, m; j = 1, \dots, r\}$ are denoted to the RSS.

The problem of estimating CDF has been well studied in the literature. Stokes and Sager [29] suggested the fundamental CDF estimator under RSS and then compared with the SRS counterpart. Thereafter, Al-Omari [1, 2] estimated CDF under L RSS and Quartile RSS respectively. Eftekharian and Razmkhah [12] introduced CDF estimator based on the kernel function under RSS. Zamanzade [30] improved a CDF estimator in the paired RSS. Ashour and Abdallah [6] developed CDF estimators using concomitant-based RSS. Under moving extreme RSS, Al-Saleh and Ahmad [5], Zamanzade *et al.* [36] and Al-Omari and Abdallah [4] proposed new CDF estimators.

Haq *et al.* [18] suggested, recently, another variation of the RSS strategy, called varied LRSS (VLRSS), that enables researchers to select more representative samples from the interested population. It is shown that VLRSS provides more precise estimates of the population mean than RSS in the case that the sample from a symmetric population. The VLRSS strategy with the set size m and the cycle number r can be explained as follows:

- Step 1.** Identify the value of the VLRSS coefficient $k = [m\alpha]$ such that $0 \leq \alpha < 0.5$, where $[x]$ is a largest integer value less than or equal to x , and m is the set size.
- Step 2.** Select randomly k samples each of size m_1 from the interested population, where $m_1 \geq m$ or $m_1 < m$ is a determined value based on cost- or budget-constraints.
- Step 3.** Select the v th smallest ranked unit from each k sets obtained by Step 2, where v is a positive integer, *i.e.* $v = 1, 2, \dots, [m_1/2]$.
- Step 4.** Repeat Step 2 for getting another k samples each of size m_1 .

- Step 5.** Select this time the $(m_1 - v + 1)$ th smallest ranked unit from each of the k sets obtained by Step 4.
- Step 6.** Select again $m(m - 2k)$ units from the interested population and allocate them into $(m - 2k)$ sets, each having m units.
- Step 7.** Select the i th smallest ranked unit from each $(m - 2k)$ sets obtained by Step 6, where $i = k + 1, \dots, m - k$.
- Step 8.** This completes one cycle of VLRSS of size m . The preceding Steps 1–7 can be repeated r times to obtain VLRSS of size $n = rm$.

The VLRSS mechanism with one cycle is illustrated in Figure 2. Symbolically, $\{Y_{i(v:m_1)j}; Y_{s(m_1-v+1:m_1)j}; Y_{t(t-k:m)j}; k \in 1, 2, \dots, [\frac{m}{2}]; v \in 1, 2, \dots, [\frac{m_1}{2}]; i = 1, 2, \dots, k; s = k + 1, k + 2, \dots, 2k; t = 2k + 1, 2k + 2, \dots, m; j = 1, \dots, r\}$ denote a VLRSS. Similar to RSS, the rounded parentheses in $\{Y_{i(v:m_1)j}; Y_{s(m_1-v+1:m_1)j}; Y_{t(t-k:m)j}\}$ emphasizes that the ranking process in VLRSS design is perfectly done. One can easily investigate that many existing variations of RSS are special cases of the VLRSS schemes, for more details see Haq *et al.* [18]. Additionally if $m_1 < m$, the number of wasted measurement items of VLRSS, $r(m_1 \times 2k + (m - 2k) \times m) - m$, will be less than the corresponding one based on RSS, $m^2 - m$. Hence VLRSS can be very useful in the case that the population is not very large. In the same sequel, Al-Omari [3] studied maximum likelihood estimators of location-scale families under VLRSS and concluded that the VLRSS-based estimators tend to outperform their RSS analogues. Motivated by these findings, we try to develop the envisaged CDF estimator under VLRSS.

The rest of the paper is organized as follows: In Section 2, the proposed CDF estimator using the empirical function is presented. Section 3 provides the proposed CDF estimator using likelihood function. Section 4 explains the proposed CDF estimator based on kernel function. Section 5 introduces a general class of CDF estimator based on missing data approach. Section 6 gives the numerical comparisons between the proposed CDF estimators with respect to the benchmark RSS estimator. In Section 7, an application to an empirical data is considered. Some concluding remarks and some potential directions for future studies are given in Section 8.

2. PROPOSED CDF ESTIMATION USING EMPIRICAL FUNCTION

The suggested CDF estimator using empirical function based on VLRSS is defined as:

$$\hat{F}_E(t) = \frac{1}{n} \sum_{j=1}^r \left(\sum_{i=1}^k I(Y_{i(v:m_1)j} \leq t) + \sum_{i=k+1}^{2k} I(Y_{i(m_1-v+1:m_1)j} \leq t) + \sum_{i=2k+1}^m I(Y_{i(i-k:m)j} \leq t) \right).$$

By setting $k = 0$ or $m_1 = m$ and $k = v = 1$, we will get the traditional CDF estimator proposed by Stockes and Sager [29] denoted by $\hat{F}_S(t)$. The following proposition shows the asymptotic properties of $\hat{F}_E(t)$.

Proposition 2.1. Let $\{Y_{i(v:m_1)j}; Y_{s(m_1-v+1:m_1)j}; Y_{t(t-k:m)j}; k \in 1, 2, \dots, [\frac{m}{2}]; v \in 1, 2, \dots, [\frac{m_1}{2}]; i = 1, 2, \dots, k; s = k + 1, k + 2, \dots, 2k; t = 2k + 1, 2k + 2, \dots, m; j = 1, \dots, r\}$ be a perfect VLRSS, then:

- (a) $\hat{F}_E(t)$ is a consistent estimator of $F(t)$ as $n \rightarrow \infty$.
- (b) $(\hat{F}_E(t) - F(t))/V(\hat{F}_E(t))$ converges in distribution to the standard normal as $n \rightarrow \infty$ for fixed k, m_1 and v .

Proof. (a) By using the fact that $I(Y_{i(v:m_1)j} \leq t)$ follows a Bernoulli distribution with success probability $B_{v, m_1-v+1}(F(t))$, one can easily investigate that the bias and the variance for $\hat{F}_E(t)$ can be derived as follows:

$$E(\hat{F}_E(t)) = \frac{1}{m} \left(k(B_{v, m_1-v+1}(F(t)) + B_{m_1-v+1, v}(F(t))) + \sum_{i=2k+1}^m B_{i-k, m-i+k+1}(F(t)) \right)$$

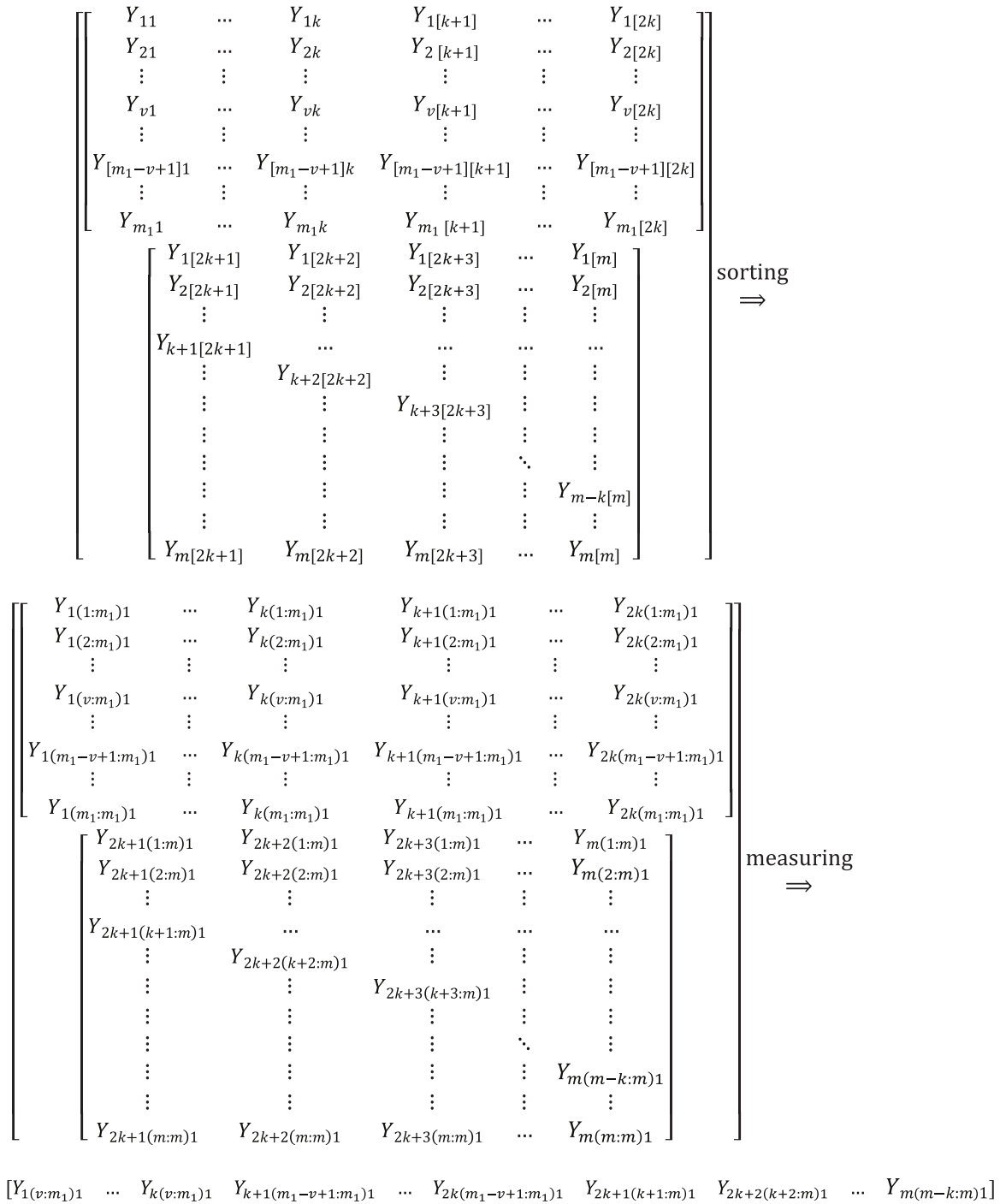


FIGURE 2. Demonstration of VLRSS procedure for one cycle.

$$= \frac{1}{m} \left(k(B_{v, m_1-v+1}(F(t)) + B_{m_1-v+1, v}(F(t))) + \sum_{i=k+1}^{m+k} B_{i-k, m-i+k+1}(F(t)) - \sum_{i=k+1}^{2k} B_{i-k, m-i+k+1}(F(t)) - \sum_{i=m+1}^{m+k} B_{i-k, m-i+k+1}(F(t)) \right).$$

Under the assumption of the consistency of the ranking process which implies that:

$$\begin{aligned} \frac{1}{m} \sum_{i=1}^m F_{(i:m)}(t) &= \frac{1}{m} \sum_{i=1}^m B_{i, k-i+1}(F(t)) = F(t) \\ E(\hat{F}_E(t)) &= \frac{1}{m} \left(k(B_{v, m_1-v+1}(F(t)) + B_{m_1-v+1, v}(F(t))) + mF(t) - \sum_{i=k+1}^{2k} B_{i-k, m-i+k+1}(F(t)) - \sum_{i=m+1}^{m+k} B_{i-k, m-i+k+1}(F(t)) \right). \end{aligned}$$

Hence the bias of $\hat{F}_E(t)$, $Bias(\hat{F}_E(t))$, will be formulated as:

$$\begin{aligned} Bias(\hat{F}_E(t)) &= \frac{1}{m} \left(k(B_{v, m_1-v+1}(F(t)) + B_{m_1-v+1, v}(F(t))) - \sum_{i=k+1}^{2k} B_{i-k, m-i+k+1}(F(t)) - \sum_{i=m+1}^{m+k} B_{i-k, m-i+k+1}(F(t)) \right). \end{aligned}$$

On the other hand, the variance of $\hat{F}_E(t)$, $V(\hat{F}_E(t))$, can be written as:

$$\begin{aligned} V(\hat{F}_E(t)) &= \frac{1}{rm^2} \left(k(B_{v, m_1-v+1}(F(t))(1 - B_{v, m_1-v+1}(F(t))) + B_{m_1-v+1, v}(F(t)) \right. \\ &\quad \left. \times (1 - B_{m_1-v+1, v}(F(t)))) + \sum_{i=2k+1}^m (B_{i-k, m-i+k+1}(F(t))(1 - B_{i-k, m-i+k+1}(F(t)))) \right), \end{aligned}$$

where $B_{a,b}(y)$ is the CDF of the Beta distribution with parameters a and b at the point y . Thus, the bias of $\hat{F}_E(t)$ tends to zero as $m \rightarrow \infty$. While the variance of $\hat{F}_E(t)$ tends to zero as either m or $r \rightarrow \infty$. This advocates that the consistency of $\hat{F}_E(t)$ which completes the proof.

(b) Let:

$$Z_j = \frac{1}{m} \left(\sum_{i=1}^k I(Y_{i(v:m_1)_j} \leq t) + \sum_{i=k+1}^{2k} I(Y_{i(m_1-v+1:m_1)_j} \leq t) + \sum_{i=2k+1}^m I(Y_{i(i-k:m)_j} \leq t) \right); \quad j = 1, 2, \dots, r.$$

Since Z_j 's are iid with finite mean and variance, then by the Central Limit Theorem $\frac{\bar{Z} - E(\hat{F}_E(t))}{\sqrt{V(\hat{F}_E(t))}}$ follows the standard normal, where $\bar{Z} = \frac{1}{r} \sum_{j=1}^r Z_j$ which achieves to the desirable result. □

3. CDF ESTIMATION USING LIKELIHOOD FUNCTION

In this part, we consider the maximum likelihood estimation (MLE) for estimation of CDF to enhance the precision of $\hat{F}_E(t)$, as the latter does not efficiently incorporate the information supported by the ranking process. Let

$$Y_i^V(t) = \begin{cases} \sum_{j=1}^r I(Y_{i(v:m_1)j} \leq t) & 1 \leq i \leq k \\ \sum_{j=1}^r I(Y_{i(m_1-v+1:m_1)j} \leq t) & k+1 \leq i \leq 2k \\ \sum_{j=1}^m I(Y_{i(i-k:m)j} \leq t) & 2k+1 \leq i \leq m \end{cases} .$$

To facilitate the notation, $Y_i^V(t)$ is written as Y^V . Assuming the perfectness, the elements of Y^V have a Bernoulli distribution, then the corresponding likelihood functions can be written as:

$$\begin{aligned} \mathbb{L}(F) &= \prod_{i=1}^k \binom{r}{y_i^V} (B_{v,m_1-v+1}(F(t)))^{y_i^V} (1 - B_{v,m_1-v+1}(F(t)))^{r-y_i^V} \\ &\times \prod_{i=k+1}^{2k} \binom{r}{y_i^V} (B_{m_1-v+1,v}(F(t)))^{y_i^V} (1 - B_{m_1-v+1,v}(F(t)))^{r-y_i^V} \\ &\times \prod_{i=2k+1}^m \binom{r}{y_i^V} (B_{i-k,m-i+k+1}(F(t)))^{y_i^V} (1 - B_{i-k,m-i+k+1}(F(t)))^{r-y_i^V} . \end{aligned}$$

Then, the log-likelihood function is:

$$\begin{aligned} \mathcal{L}(F) &\propto \sum_{i=1}^k y_i^V \ln(B_{v,m_1-v+1}(F(t))) + (r - y_i^V) \ln(1 - B_{v,m_1-v+1}(F(t))) \\ &+ \sum_{i=k+1}^{2k} y_i^V \ln(B_{m_1-v+1,v}(F(t))) + (r - y_i^V) \ln(1 - B_{m_1-v+1,v}(F(t))) \\ &+ \sum_{i=2k+1}^m y_i^V \ln(B_{i-k,m-i+k+1}(F(t))) + (r - y_i^V) \ln(1 - B_{i-k,m-i+k+1}(F(t))). \end{aligned}$$

Therefore our second proposed CDF estimator, $\hat{F}_M(t)$, based on VLRSS is obtained by maximizing $\mathbb{L}(F)$ or equivalently maximizing $\mathcal{L}(F)$ as shown below:

$$\begin{aligned} \frac{\partial \mathcal{L}(F)}{\partial F} &= \sum_{i=1}^k \frac{(y_i^V - rB_{v,m_1-v+1}(F(t)))b_{v,m_1-v+1}(F(t))}{B_{v,m_1-v+1}(F(t))(1 - B_{v,m_1-v+1}(F(t)))} \\ &+ \sum_{i=k+1}^{2k} \frac{(y_i^V - rB_{m_1-v+1,v}(F(t)))b_{m_1-v+1,v}(F(t))}{B_{m_1-v+1,v}(F(t))(1 - B_{m_1-v+1,v}(F(t)))} \\ &+ \sum_{i=2k+1}^m \frac{(y_i^V - rB_{i-k,m-i+k+1}(F(t)))b_{i-k,m-i+k+1}(F(t))}{B_{i-k,m-i+k+1}(F(t))(1 - B_{i-k,m-i+k+1}(F(t)))} , \end{aligned}$$

where $b_{a,b}(y)$ is the pdf of the Beta distribution with parameters a and b at the point y . One can easily verify that $\frac{\partial^2 \mathcal{L}(F)}{\partial F^2} < 0$, $\frac{\partial \mathcal{L}(F)}{\partial F}$ becomes positive as $F \rightarrow 0$ and negative as $F \rightarrow 1$. Therefore, $\mathcal{L}(F)$ is strictly concave in $F(t)$. Although $\hat{F}_M(t)$ does not have a closed form, it can be simply obtained by using any iterative technique such as optimize function in R package. The corresponding Fisher information number about $F(t)$ under VLRSS is given by:

$$I_V = -E\left(\frac{d^2 \log L_1(F)}{dF^2}\right) = r(I_{V1} + I_{V2} + I_{V3}),$$

where:

$$I_{V1} = \left(\frac{k(b_{v,m_1-v+1}(F(t)))^2}{B_{v,m_1-v+1}(F(t))(1 - B_{v,m_1-v+1}(F(t)))} \right),$$

$$I_{V2} = \left(\frac{k(b_{m_1-v+1,v}(F(t)))^2}{B_{m_1-v+1,v}(F(t))(1 - B_{m_1-v+1,v}(F(t)))} \right),$$

and $I_{V3} = \sum_{i=2k+1}^m \frac{(b_{i-k,m-i+k+1}(F(t)))^2}{B_{i-k,m-i+k+1}(F(t))(1 - B_{i-k,m-i+k+1}(F(t)))}$.

Hence, the variance of $\hat{F}_M(t)$ can be expressed as:

$$\sigma_M^2 = V(\hat{F}_M(t)) = \frac{1}{r(I_{V1} + I_{V2} + I_{V3})}$$

The following proposition shows the asymptotic properties of $\hat{F}_M(t)$.

Proposition 3.1. *Let $Y_{i(v:m_1)j}; Y_{s(m_1-v+1:m_1)j}; Y_{t(t-k:m)j}; k \in 1, 2, \dots, [\frac{m}{2}]; v \in \{1, 2, \dots, [\frac{m_1}{2}]\}; i = 1, 2, \dots, k; s = k + 1, k + 2, \dots, 2k; t = 2k + 1, 2k + 2, \dots, m; j = 1, \dots, r$ be a perfect VLRSS. For a fixed m, k, m_1 and v , when $r \rightarrow \infty$, then $\sqrt{r}(\hat{F}_M(t) - F)$ converges in distribution to the normal distribution with zero mean and variance $r\sigma_M^2$.*

Proof. Let:

$$L(s) = \mathcal{L}\left(F + \frac{s}{\sqrt{r}}\right) - \mathcal{L}(F)$$

Then:

$$L(s) = \left[\sum_{i=1}^k y_i^V \ln \left(\frac{B_{v,m_1-v+1}\left(F(t) + \frac{s}{\sqrt{r}}\right)}{B_{v,m_1-v+1}(F(t))} \right) + (r - y_i^V) \ln \left(\frac{1 - B_{v,m_1-v+1}\left(F(t) + \frac{s}{\sqrt{r}}\right)}{1 - B_{v,m_1-v+1}(F(t))} \right) \right]$$

$$+ \left[\sum_{i=k+1}^{2k} y_i^V \ln \left(\frac{B_{m_1-v+1,v}\left(F(t) + \frac{s}{\sqrt{r}}\right)}{B_{m_1-v+1,v}(F(t))} \right) + (r - y_i^V) \ln \left(\frac{1 - B_{m_1-v+1,v}\left(F(t) + \frac{s}{\sqrt{r}}\right)}{1 - B_{m_1-v+1,v}(F(t))} \right) \right]$$

$$+ \left[\sum_{i=2k+1}^m y_i^V \ln \left(\frac{B_{i-k,m-i+k+1}\left(F(t) + \frac{s}{\sqrt{r}}\right)}{B_{i-k,m-i+k+1}(F(t))} \right) + (r - y_i^V) \ln \left(\frac{1 - B_{i-k,m-i+k+1}\left(F(t) + \frac{s}{\sqrt{r}}\right)}{1 - B_{i-k,m-i+k+1}(F(t))} \right) \right]$$

$$= A(s) + B(s) + C(s),$$

where the three terms between brackets will be denoted respectively by $A(s), B(s)$ and $C(s)$. Using Taylor series expansion, the logarithm function of the first term $A(s)$ can be approximated as:

$$A(s) \approx \sum_{i=1}^k y_i^V \left(\frac{B_{v,m_1-v+1}\left(F(t) + \frac{s}{\sqrt{r}}\right) - B_{v,m_1-v+1}(F(t))}{B_{v,m_1-v+1}(F(t))} \right.$$

$$\left. - \frac{1}{2} \left(\frac{B_{v,m_1-v+1}\left(F(t) + \frac{s}{\sqrt{r}}\right) - B_{v,m_1-v+1}(F(t))}{B_{v,m_1-v+1}(F(t))} \right)^2 \right)$$

$$\begin{aligned}
 & - \sum_{i=1}^k (r - y_i^V) \left(\frac{B_{v,m_1-v+1} \left(F(t) + \frac{s}{\sqrt{r}} \right) - B_{v,m_1-v+1}(F(t))}{1 - B_{v,m_1-v+1}(F(t))} \right. \\
 & \left. + \frac{1}{2} \left(\frac{B_{v,m_1-v+1} \left(F(t) + \frac{s}{\sqrt{r}} \right) - B_{v,m_1-v+1}(F(t))}{1 - B_{v,m_1-v+1}(F(t))} \right)^2 \right).
 \end{aligned}$$

Which is equivalent to:

$$\begin{aligned}
 A(s) & \approx \sum_{i=1}^k \left(B_{v,m_1-v+1} \left(F(t) + \frac{s}{\sqrt{r}} \right) - B_{v,m_1-v+1}(F(t)) \right) \left(\frac{y_i^V}{B_{v,m_1-v+1}(F(t))} - \frac{(r - y_i^V)}{1 - B_{v,m_1-v+1}(F(t))} \right) \\
 & - \frac{1}{2} \sum_{i=1}^k \left(B_{v,m_1-v+1} \left(F(t) + \frac{s}{\sqrt{r}} \right) - B_{v,m_1-v+1}(F(t)) \right)^2 \\
 & \times \left(\frac{y_i^V}{(B_{v,m_1-v+1}(F(t)))^2} + \frac{(r - y_i^V)}{(1 - B_{v,m_1-v+1}(F(t)))^2} \right) \\
 & = U(s)_A - \frac{1}{2} V(s)_A,
 \end{aligned}$$

where:

$$U(s)_A = \sum_{i=1}^k \left(B_{v,m_1-v+1} \left(F(t) + \frac{s}{\sqrt{r}} \right) - B_{v,m_1-v+1}(F(t)) \right) \left(\frac{y_i^V}{B_{v,m_1-v+1}(F(t))} - \frac{(r - y_i^V)}{1 - B_{v,m_1-v+1}(F(t))} \right),$$

and

$$\begin{aligned}
 V(s)_A & = \sum_{i=1}^k \left(B_{v,m_1-v+1} \left(F(t) + \frac{s}{\sqrt{r}} \right) - B_{v,m_1-v+1}(F(t)) \right)^2 \\
 & \times \left(\frac{y_i^V}{(B_{v,m_1-v+1}(F(t)))^2} + \frac{(r - y_i^V)}{(1 - B_{v,m_1-v+1}(F(t)))^2} \right).
 \end{aligned}$$

Noting that as $r \rightarrow \infty$:

$$\frac{\sqrt{r}}{s} \left(B_{v,m_1-v+1} \left(F(t) + \frac{s}{\sqrt{r}} \right) - B_{v,m_1-v+1}(F(t)) \right) \rightarrow \frac{\partial B_{v,m_1-v+1}(F(t))}{\partial F} = b_{v,m_1-v+1}(F(t)).$$

Recalling that: y_i^V , $1 \leq i \leq k$ has a binominal distribution with the success probability $B_{v,m_1-v+1}(F(t))$, then as $r \rightarrow \infty$:

$$\begin{aligned}
 U(s)_A & = s \sum_{i=1}^k \left(\frac{B_{v,m_1-v+1} \left(F(t) + \frac{s}{\sqrt{r}} \right) - B_{v,m_1-v+1}(F(t))}{\frac{s}{\sqrt{r}} \sqrt{B_{v,m_1-v+1}(F(t))(1 - B_{v,m_1-v+1}(F(t)))}} \right) \\
 & \times \left(\frac{y_i^V - r B_{v,m_1-v+1}(F(t))}{\sqrt{r B_{v,m_1-v+1}(F(t))(1 - B_{v,m_1-v+1}(F(t)))}} \right) \\
 & \xrightarrow{d} sN(0, I_{LV1}) = sU_A
 \end{aligned}$$

where \xrightarrow{d} stands for converges in distribution, furthermore:

$$E(V(s)_A) = \sum_{i=1}^k \left(B_{v,m_1-v+1} \left(F(t) + \frac{s}{\sqrt{r}} \right) - B_{v,m_1-v+1}(F(t)) \right)^2 \\ \times \left(\frac{E(y_i^V)}{(B_{v,m_1-v+1}(F(t)))^2} + \frac{(r - E(y_i^V))}{(1 - B_{v,m_1-v+1}(F(t)))^2} \right).$$

As $r \rightarrow \infty$, we can get:

$$E(V(s)_A) = s^2 \sum_{i=1}^k (b_{v,m_1-v+1}(F(t)))^2 \left(\frac{1}{(B_{v,m_1-v+1}(F(t)))} + \frac{1}{(1 - B_{v,m_1-v+1}(F(t)))} \right) = s^2 V_A,$$

where: $U_A = N(0, I_{V_1})$ and $V_A = I_{V_1}$. Therefore, we can write:

$$A(s) \xrightarrow{d} sU_A - \frac{1}{2}s^2V_A.$$

By using similar arguments in above, one can simply show that:

$$B(s) \approx U(s)_B - \frac{1}{2}V(s)_B \text{ and also } C(s) \approx U(s)_C - \frac{1}{2}V(s)_C,$$

where:

$$U(s)_B = \sum_{i=k+1}^{2k} \left(\frac{B_{m_1-v+1,v} \left(F(t) + \frac{s}{\sqrt{r}} \right) - B_{m_1-v+1,v}(F(t))}{\sqrt{B_{m_1-v+1,v}(F(t))(1 - B_{m_1-v+1,v}(F(t)))}} \right) \\ \times \left(\frac{y_i^V - rB_{m_1-v+1,v}(F(t))}{\sqrt{B_{m_1-v+1,v}(F(t))(1 - B_{m_1-v+1,v}(F(t)))}} \right), \\ V(s)_B = \sum_{i=k+1}^{2k} \left(B_{m_1-v+1,v} \left(F(t) + \frac{s}{\sqrt{r}} \right) - B_{m_1-v+1,v}(F(t)) \right)^2 \\ \times \left(\frac{y_i^V}{(B_{m_1-v+1,v}(F(t)))^2} + \frac{(r - y_i^V)}{(1 - B_{m_1-v+1,v}(F(t)))^2} \right), \\ U(s)_C = \sum_{i=2k+1}^m \left(\frac{B_{i-k,m-i+k+1} \left(F(t) + \frac{s}{\sqrt{r}} \right) - B_{i-k,m-i+k+1}(F(t))}{\sqrt{B_{i-k,m-i+k+1}(F(t))(1 - B_{i-k,m-i+k+1}(F(t)))}} \right) \\ \times \left(\frac{y_i^V - rB_{i-k,m-i+k+1}(F(t))}{\sqrt{B_{i-k,m-i+k+1}(F(t))(1 - B_{i-k,m-i+k+1}(F(t)))}} \right),$$

and

$$V(s)_C = \sum_{i=2k+1}^m \left(B_{i-k,m-i+k+1} \left(F(t) + \frac{s}{\sqrt{r}} \right) - B_{i-k,m-i+k+1}(F(t)) \right)^2 \\ \times \left(\frac{y_i^V}{(B_{i-k,m-i+k+1}(F(t)))^2} + \frac{(r - y_i^V)}{(1 - B_{i-k,m-i+k+1}(F(t)))^2} \right).$$

Furthermore, we have as $r \rightarrow \infty$:

$$U(s)_B \xrightarrow{d} sU_B, U(s)_C \xrightarrow{d} sU_C, V(s)_B \xrightarrow{p} s^2V_B \text{ and } V(s)_C \xrightarrow{p} s^2V_C,$$

where \xrightarrow{p} stands for convergence in probability, $U_B = N(0, I_{V_2}), U_C = N(0, I_{V_3}), V_B = I_{V_2}$ and $V_C = I_{V_3}$. Hence we can conclude that as $r \rightarrow \infty$:

$$L(s) = \mathcal{L}\left(F + \frac{s}{\sqrt{r}}\right) - \mathcal{L}(F) \xrightarrow{p} -\frac{1}{2}s^2(V_A + V_B + V_C) + s(U_A + U_B + U_C).$$

Maximizing $L(s)$ with respect to s can be done as:

$$\hat{s} = \operatorname{argmax}_s L(s) = \frac{(U_A + U_B + U_C)}{(V_A + V_B + V_C)}.$$

It is clear that $\hat{s} \xrightarrow{p} N\left(0, \frac{1}{I_{V_1} + I_{V_2} + I_{V_3}}\right)$. Since $\mathcal{L}(F)$ is also maximized at $\hat{F}_M(t)$, then: $F + \frac{\hat{s}}{\sqrt{r}} = \hat{F}_M(t) \Rightarrow \hat{s} = \sqrt{r}(\hat{F}_M(t) - F)$, for which the proof is completed. □

4. CDF ESTIMATION USING KERNEL FUNCTION

The CDF estimators based on empirical function are step function in opposite to the population CDF which is a smooth function. This motivates us to use the kernel approach to produce a smooth estimator for the population CDF. The suggested CDF estimator using kernel function based on VLRSS is defined as:

$$\hat{F}_K(t) = \frac{1}{n} \sum_{j=1}^r \left(\sum_{i=1}^k K\left(\frac{t - Y_{i(v:m_1)j}}{\hat{h}}\right) + \sum_{i=k+1}^{2k} K\left(\frac{t - Y_{i(m_1-v+1:m_1)j}}{\hat{h}}\right) + \sum_{i=2k+1}^m K\left(\frac{t - Y_{i(i-k:m)j}}{\hat{h}}\right) \right),$$

where \hat{h} is the smoothing parameter to be specified, $K(t) = \int_{-\infty}^t k(y) dy$, $k(\cdot)$ is the kernel function satisfies certain conditions: (1) $k(\cdot)$ is a bounded symmetric function about zero over the interval $[-a, a]$, (2) $\int k(y) dy = 1$, (3) $\int y^2 k(y) dy \neq 0$. Hereafter, Epanechnikov function is adopted as a kernel function. Further, the iterative algorithm produced by Polansky and Baker [28] coded as R-package, Kerdies, is used for estimating \hat{h} . The following proposition shows the asymptotic properties of $\hat{F}_K(t)$.

Proposition 4.1. *Let $\{Y_{i(v:m_1)j}; Y_{s(m_1-v+1:m_1)j}; Y_{t(i-k:m)j}; k \in 1, 2, \dots, [\frac{m}{2}]; v \in 1, 2, \dots, [\frac{m_1}{2}]; i = 1, 2, \dots, k; s = k + 1, k + 2, \dots, 2k; t = 2k + 1, 2k + 2, \dots, m; j = 1, \dots, r\}$ be a perfect VLRSS and $h \rightarrow 0$ as $n \rightarrow \infty$, then:*

- (a) $\hat{F}_K(t)$ is a consistent estimator of $F(t)$.
- (b) $(\hat{F}_K(t) - F(t))/V(\hat{F}_K(t))$ converges in distribution to the standard normal.

Proof. (a)

$$E(\hat{F}_K(t)) = \frac{k}{m} \left(\left[\int_{-\infty}^{\infty} K\left(\frac{t - y_1}{h}\right) f_{(v:m_1)}(y_1) dy_1 \right] + \left[\int_{-\infty}^{\infty} K\left(\frac{t - y_2}{h}\right) f_{(m_1-v+1:m_1)}(y_2) dy_2 \right] \right) + \frac{1}{m} \left[\int_{-\infty}^{\infty} \sum_{i=2k+1}^m K\left(\frac{t - y_3}{h}\right) f_{(i-k:m)}(y_3) dy_3 \right] \tag{4.1}$$

$$= \frac{k}{m}(E_1 + E_2) + \frac{1}{m}E_3, \tag{4.2}$$

where the three terms between the brackets shown in (4.1) will be denoted respectively by E_1, E_2 and E_3 . Since:

$$E_1 = \int_{-\infty}^{t-ah} K\left(\frac{t-y_1}{h}\right) f_{(v:m_1)}(y_1) dy_1 + \int_{t-ah}^{t+ah} K\left(\frac{t-y_1}{h}\right) f_{(v:m_1)}(y_1) dy_1 + \int_{t+ah}^{\infty} K\left(\frac{t-y_1}{h}\right) f_{(v:m_1)}(y_1) dy_1.$$

Recalling the fact that:

$$K(y) = \begin{cases} 0 & y \leq -a \\ 1 & y \geq a \end{cases}.$$

Then:

$$\begin{aligned} E_1 &= \int_{-\infty}^{t-ah} f_{(v:m_1)}(y_1) dy_1 + \int_{t-ah}^{t+ah} K\left(\frac{t-y_1}{h}\right) f_{(v:m_1)}(y_1) dy_1 \\ &= F_{(v:m_1)}(t-ah) + \int_{t-ah}^{t+ah} K\left(\frac{t-y_1}{h}\right) f_{(v:m_1)}(y_1) dy_1 \\ &= F_{(v:m_1)}(t-ah) + h \int_{-a}^a K(z) f_{(v:m_1)}(t-hz) dz. \end{aligned} \tag{4.3}$$

Under the assumption $h \rightarrow 0$ as $n \rightarrow \infty$; we will obtain:

$$E_1 \approx F_{(v:m_1)}(t). \tag{4.4}$$

Similarly, one can easy show that under the assumption $h \rightarrow 0$ as $n \rightarrow \infty$:

$$E_2 \approx F_{(m_1-v+1:m_1)}(t) \quad \text{and} \quad E_3 \approx \sum_{i=2k+1}^m F_{(i-k:m)}(t). \tag{4.5}$$

By substituting (4.4) and (4.5) in (4.2), we will get:

$$\begin{aligned} E(\hat{F}_K(t)) &= \frac{1}{m} \left(kF_{(v:m_1)}(t) + kF_{(m_1-v+1:m_1)}(t) + \sum_{i=2k+1}^m F_{(i-k:m)}(t) \right) \\ &= \frac{1}{m} \left(kF_{(v:m_1)}(t) + kF_{(m_1-v+1:m_1)}(t) + \sum_{i=k+1}^{m+k} F_{(i-k:m)}(t) \right. \\ &\quad \left. - \sum_{i=k+1}^{2k} F_{(i-k:m)}(t) - \sum_{i=m+1}^{m+k} F_{(i-k:m)}(t) \right). \end{aligned}$$

Under the assumption of the consistency of the ranking process, we will obtain:

$$Bias(\hat{F}_K(t)) \approx \frac{1}{m} \left(kF_{(v:m_1)}(t) + kF_{(m_1-v+1:m_1)}(t) - \sum_{i=k+1}^{2k} F_{(i-k:m)}(t) - \sum_{i=m+1}^{m+k} F_{(i-k:m)}(t) \right). \tag{4.6}$$

On the other hand, the asymptotic variance:

$$V(\hat{F}_K(t)) = \frac{1}{m^2 r} \left(k \left(\left[V \left(K \left(\frac{t - Y_{i(v:m_1)j}}{h} \right) \right) \right] + \left[V \left(K \left(\frac{t - Y_{i(m_1-v+1:m_1)j}}{h} \right) \right) \right] \right) \right)$$

$$\begin{aligned}
 & + \sum_{i=2k+1}^m \left[V \left(K \left(\frac{t - Y_{i(i-k:m)j}}{h} \right) \right) \right] \\
 & = \frac{1}{m^2 r} (k(V_1 + V_2)) + V_3 \tag{4.7}
 \end{aligned}$$

where the three terms between the brackets shown in (4.7) will be denoted respectively by V_1, V_2 and V_3 . Since:

$$V_1 = V \left(K \left(\frac{t - Y_{i(v:m_1)j}}{h} \right) \right) = E \left(K^2 \left(\frac{t - Y_{i(v:m_1)j}}{h} \right) \right) - \left(E \left(K \left(\frac{t - Y_{i(v:m_1)j}}{h} \right) \right) \right)^2.$$

Similar to (4.3), we get:

$$\begin{aligned}
 E \left(K^2 \left(\frac{t - Y_{i(v:m_1)j}}{h} \right) \right) & = \int_{-\infty}^{t-ah} K^2 \left(\frac{t - y_1}{h} \right) f_{(v:m_1)}(y_1) dy_1 + \int_{t-ah}^{t+ah} K^2 \left(\frac{t - y_1}{h} \right) f_{(v:m_1)}(y_1) dy_1 \\
 & + \int_{t+ah}^{\infty} K^2 \left(\frac{t - y_1}{h} \right) f_{(v:m_1)}(y_1) dy_1 \\
 & = F_{(v:m_1)}(t - ah) + h \int_{-a}^a K^2(z) f_{(v:m_1)}(t - hz) dz.
 \end{aligned}$$

On the other hand,

$$\begin{aligned}
 \left(E \left(K \left(\frac{t - Y_{i(v:m_1)j}}{h} \right) \right) \right)^2 & = \left(\int_{-\infty}^{t-ah} K \left(\frac{t - y_1}{h} \right) f_{(v:m_1)}(y_1) dy_1 + \int_{t-ah}^{t+ah} K \left(\frac{t - y_1}{h} \right) f_{(v:m_1)}(y_1) dy_1 \right. \\
 & \left. + \int_{t+ah}^{\infty} K \left(\frac{t - y_1}{h} \right) f_{(v:m_1)}(y_1) dy_1 \right)^2.
 \end{aligned}$$

After simple algebra, we can get:

$$\left(E \left(K \left(\frac{t - Y_{i(v:m_1)j}}{h} \right) \right) \right)^2 = \left(F_{(v:m_1)}(t - ah) + h \int_{-a}^a K(z) f_{(v:m_1)}(t - hz) dz \right)^2.$$

Under the assumption $h \rightarrow 0$ as $n \rightarrow \infty$; we will obtain:

$$V_1 \approx F_{(v:m_1)}(t) - (F_{(v:m_1)}(t))^2 = F_{(v:m_1)}(t)(1 - F_{(v:m_1)}(t)). \tag{4.8}$$

Likewise, one can easily investigate that:

$$V_2 \approx F_{(m_1-v+1:m_1)}(t)(1 - F_{(m_1-v+1:m_1)}(t)) \quad \text{and} \quad V_3 \approx \sum_{i=2k+1}^m F_{(i-k:m)}(t)(1 - F_{(i-k:m)}(t)). \tag{4.9}$$

By substituting (4.8) and (4.9) in (4.7), we will get:

$$\begin{aligned}
 V(\hat{F}_K(t)) & \approx \frac{1}{m^2 r} \left(k(F_{(v:m_1)}(t)(1 - F_{(v:m_1)}(t)) + F_{(m_1-v+1:m_1)}(t)(1 - F_{(m_1-v+1:m_1)}(t))) \right. \\
 & \left. + \sum_{i=2k+1}^m F_{(i-k:m)}(t)(1 - F_{(i-k:m)}(t)) \right). \tag{4.10}
 \end{aligned}$$

In the light of (4.6) and (4.10), one can conclude that for a fixed k, m_1, v , the bias of $\hat{F}_K(t)$ tends to zero as $m \rightarrow \infty$ and the variance of $\hat{F}_K(t)$ tends to zero as either m or $r \rightarrow \infty$ coming up $MSE(\hat{F}_K(t)) \rightarrow 0$ which completes the proof, where MSE refers to the mean square error.

(b) Let:

$$W_j = \frac{1}{m} \left(\sum_{i=1}^k K \left(\frac{t - Y_{i(v:m_1)j}}{\hat{h}} \right) + \sum_{i=k+1}^{2k} K \left(\frac{t - Y_{i(m_1-v+1:m_1)j}}{\hat{h}} \right) + \sum_{i=2k+1}^m K \left(\frac{t - Y_{i(i-k:m)j}}{\hat{h}} \right) \right) \quad j = 1, 2, \dots, r.$$

Following the same procedure explained in the proof of Proposition 2.1 (b), the desirable result will be obtained. □

5. CDF ESTIMATION USING UNMEASURED UNITS

In this part, we shall incorporate all the potential information provided by the unmeasured items to construct a general class of CDF estimators. Taking motivation from the aforesaid CDF estimators, our proposed class of CDF estimators are based on $2krm_1 + rm(m - 2k)$ units rather than m units. It was shown by many studies that combining the information supported by the measured items together with the unmeasured items may be improve the precision of the estimators. For instance, Kvam and Samaniego [21], Ozturk [26] and Ashour and Abdallah [8] used missing data mechanism for estimating CDF under RSS. Frey and Wang [13] adopted an iterative algorithm for estimation the mean of the exponential distribution. Ozturk [27] and Ashour and Abdallah [7] also used an iterative algorithm for parametric estimation of location and scale parameters under RSS.

Assume that we have an observation $y_{(i)}$ is less (greater) than a given t , this implies that all the observations $y_{(j)} j < i (j > i)$ are also less (greater) than t . Whereas, one can use the lower (upper) truncated CDF of $y_{(i)}$ to find out the probability that the observations $y_{(j)} j > i (j < i)$ is less than t . For more details on this mechanism, an interested reader can pursue Ashour and Abdallah [8]. The suggested CDF estimator taken into account the measured as well as the unmeasured items based on VLRSS given by:

$$\begin{aligned} \hat{F}^*(t) = & \frac{1}{2krm_1 + rm(m - 2k)} \left(\left[\sum_{j=1}^r \sum_{i=1}^k \left(\left(v + (m_1 - v) \frac{\hat{F}(t) - \hat{F}(Y_{i(v:m_1)j})}{1 - \hat{F}(Y_{i(v:m_1)j})} \right) I(Y_{i(v:m_1)j} \leq t) \right. \right. \\ & \left. \left. + \left((v - 1) \frac{\hat{F}(t)}{\hat{F}(Y_{i(v:m_1)j})} \right) I(Y_{i(v:m_1)j} > t) \right) \right] \\ & + \left[\sum_{j=1}^r \sum_{i=k+1}^{2k} \left(\left((m_1 - v + 1) + (v - 1) \frac{\hat{F}(t) - \hat{F}(Y_{i(m_1-v+1:m_1)j})}{1 - \hat{F}(Y_{i(m_1-v+1:m_1)j})} \right) I(Y_{i(m_1-v+1:m_1)j} \leq t) \right. \right. \\ & \left. \left. + \left((m_1 - v) \frac{\hat{F}(t)}{\hat{F}(Y_{i(m_1-v+1:m_1)j})} \right) I(Y_{i(m_1-v+1:m_1)j} > t) \right) \right] \\ & + \left[\sum_{j=1}^r \sum_{i=2k+1}^m \left(\left((i - k) + (m - i + k) \frac{\hat{F}(t) - \hat{F}(Y_{i(i-k:m)j})}{1 - \hat{F}(Y_{i(i-k:m)j})} \right) I(Y_{i(i-k:m)j} \leq t) \right. \right. \\ & \left. \left. + \left((i - k - 1) \frac{\hat{F}(t)}{\hat{F}(Y_{i(i-k:m)j})} \right) I(Y_{i(i-k:m)j} > t) \right) \right] \right) \end{aligned} \tag{5.1}$$

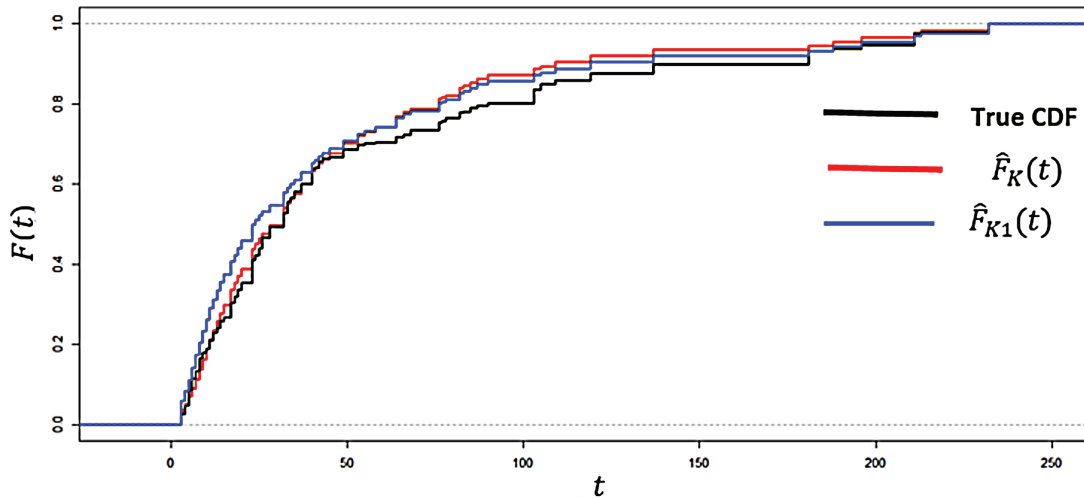


FIGURE 3. The population CDF and the CDF estimators based on kernel function based on tree dataset.

where $\hat{F}(x)$ is a consistent CDF estimator needed to be determined. It is notable that replacing $\hat{F}(t)$ in (5.1) with a suitable CDF estimator produces a new CDF estimator. Therefore, it is expectedly that the performance of $\hat{F}^*(t)$ will depend strongly on the quality of the selected CDF estimator. Another important point to be mentioned that $\hat{F}^*(t)$ is derived under the perfectness assumption, thus violating diametrically this condition may suffers of lack of efficiency. The following proposition shows the consistency of $\hat{F}^*(t)$.

Proposition 5.1. *Let $Y_{i(v:m_1)j}; Y_{s(m_1-v+1:m_1)j}; Y_{t(t-k:m)j}; k \in \{1, 2, \dots, [\frac{m}{2}]\}; k \in \{1, 2, \dots, [\frac{m_1}{2}]\}; i = 1, 2, \dots, k; s = k + 1, k + 2, \dots, 2k; t = 2k + 1, 2k + 2, \dots, m; j = 1, \dots, r$ be a perfect VLRSS and $\hat{F}(x) \xrightarrow{p} F(x)$. For a fixed m, k, m_1 and v , when $r \rightarrow \infty$, then $\hat{F}^*(t) \xrightarrow{p} F(t)$.*

Proof. Let $\hat{F}^*(t)$ be written as:

$$\hat{F}^*(t) = \frac{1}{2krm_1 + rm(m - 2k)}(A + B + C), \tag{5.2}$$

where the three terms between brackets in (5.1) will be denoted respectively by A, B and C .

Under the assumption of the consistency of $\hat{F}(x)$, A can be expressed as:

$$A \cong \sum_{j=1}^r \sum_{i=1}^k \left(\left(v + (m_1 - v) \frac{F(t) - F(Y_{i(v:m_1)j})}{1 - F(Y_{i(v:m_1)j})} \right) I(Y_{i(v:m_1)j} \leq t) + \left((v - 1) \frac{F(t)}{F(Y_{i(v:m_1)j})} \right) I(Y_{i(v:m_1)j} > t) \right).$$

Since $\sum_{j=1}^r I(Y_{i(v:m_1)j} \leq t) \xrightarrow{p} rB_{v,m_1-v+1}(F(t))$, as $r \rightarrow \infty$. Hence we get:

$$A \cong r \sum_{i=1}^k \left(vB_{v,m_1-v+1}(F(t)) + (m_1 - v) \int_{-\infty}^t \left(\frac{F(t) - F(u)}{1 - F(u)} \right) f_{v:m_1}(u) du + (v - 1) \int_t^{\infty} \left(\frac{F(t)}{F(u)} \right) f_{v:m_1}(u) du \right)$$

which is equivalent to:

$$A \cong rk \left(vB_{v,m_1-v+1}(F(t)) + (m_1 - v) \int_{-\infty}^t \left(\frac{F(t) - F(u)}{1 - F(u)} \right) f_{v:m_1}(u) du + (v - 1) \int_t^{\infty} \left(\frac{F(t)}{F(u)} \right) f_{v:m_1}(u) du \right)$$

$$\begin{aligned}
 &= rk \left(vB_{v,m_1-v+1}(F(t)) + (m_1 - v) \int_{-\infty}^t (F(t) - F(u)) \frac{(m_1)!}{(v-1)!(m_1-v)!} (F(u))^{v-1} (1-F(u))^{m_1-v} f(u) du \right. \\
 &\quad \left. + (v-1) \int_t^\infty (F(t)) \frac{(m_1)!}{(v-1)!(m_1-v)!} (F(u))^{v-2} (1-F(u))^{m_1-v} f(u) du \right) \\
 &= rk \left(vB_{v,m_1-v+1}(F(t)) + m_1 F(t) \int_{-\infty}^t \frac{(m_1-1)!}{(v-1)!(m_1-v-1)!} (F(u))^{v-1} (1-F(u))^{m_1-v-1} f(u) du \right. \\
 &\quad \left. - v \int_{-\infty}^t \frac{(m_1)!}{(v)!(m_1-v-1)!} (F(u))^v (1-F(u))^{m_1-v-1} f(u) du \right. \\
 &\quad \left. + m_1 F(t) \int_t^\infty \frac{(m_1-1)!}{(v-2)!(m_1-v)!} (F(u))^{v-2} (1-F(u))^{m_1-v} f(u) du \right) \\
 &= rk(vB_{v,m_1-v+1}(F(t)) + m_1(F(t))B_{v,m_1-v}(F(t)) - vB_{v+1,m_1-v}(F(t)) + m_1 F(t)(1 - B_{v-1,m_1-v+1}(F(t))))).
 \end{aligned}$$

By using the integration by parts, one can easily investigate that: $v(B_{v,m_1-v+1}(F(t)) - B_{v+1,m_1-v}(F(t))) = m_1(F(t))(B_{v-1,m_1-v+1}(F(t)) - B_{v,m_1-v}(F(t)))$ coming up with:

$$A \cong rkm_1(F(t)). \tag{5.3}$$

TABLE 1. The RE values of the CDF estimators using simulated normal distribution when $m = 3, k = 1$.

		$\rho = 1$											
		$m_1 = 4$					$m_1 = 10$						
F		\hat{F}_E	\hat{F}_{E1}	\hat{F}_M	\hat{F}_{M1}	\hat{F}_K	\hat{F}_{K1}	\hat{F}_E	\hat{F}_{E1}	\hat{F}_M	\hat{F}_{M1}	\hat{F}_K	\hat{F}_{K1}
$v = 1$	0.10	0.84	1.55	0.98	1.70	1.33	1.37	0.87	1.87	1.22	2.30	1.09	1.22
	0.25	1.08	1.54	1.10	1.44	1.60	1.55	1.43	2.56	1.10	1.21	1.99	2.26
	0.50	1.30	1.01	1.04	0.97	1.95	1.66	2.01	1.32	0.95	0.89	4.21	3.78
	0.75	1.05	0.87	1.05	0.87	1.77	1.53	1.16	1.01	1.19	0.80	1.66	1.78
	0.90	0.87	0.65	0.88	0.69	1.34	1.36	0.80	0.66	1.46	0.92	1.07	1.31
$v = 2$	0.10	0.87	2.01	1.56	1.89	2.60	2.76	0.99	1.18	1.16	2.02	1.98	2.15
	0.25	0.78	0.87	0.82	0.97	0.99	0.98	1.03	1.70	1.06	1.54	1.86	1.92
	0.50	0.87	0.87	0.97	0.76	0.97	0.94	1.28	1.36	1.40	1.24	2.96	3.01
	0.75	0.87	0.75	0.84	1.14	1.17	1.13	1.19	0.99	1.20	0.89	2.07	2.14
	0.90	1.78	1.37	1.55	2.29	2.45	2.19	1.11	0.99	1.31	0.92	2.23	2.36
$v = 3$	0.10	-	-	-	-	-	-	1.87	2.29	1.73	2.24	3.13	3.19
	0.25	-	-	-	-	-	-	0.98	1.21	0.99	1.27	1.57	1.68
	0.50	-	-	-	-	-	-	1.32	1.27	1.86	1.43	1.88	2.25
	0.75	-	-	-	-	-	-	0.99	0.87	1.06	0.98	1.68	1.75
	0.90	-	-	-	-	-	-	1.76	1.27	1.87	1.54	2.88	2.80
		$\rho = 0.9$											
$v = 1$	0.10	0.98	1.78	1.12	1.91	1.55	1.57	0.87	1.76	1.15	2.04	1.10	1.19
	0.25	1.03	1.43	1.01	1.32	1.51	1.40	1.06	1.71	0.88	0.96	1.63	1.71
	0.50	1.23	1.06	1.03	0.99	1.71	1.54	1.19	1.09	0.87	0.78	3.22	2.68
	0.75	1.01	0.87	0.99	0.87	1.53	1.37	1.18	0.78	0.97	0.87	1.76	1.82
	0.90	0.78	0.65	0.88	0.65	1.25	1.23	0.76	0.66	1.09	0.87	0.98	1.17
$v = 2$	0.10	1.54	1.77	1.27	1.65	2.21	1.85	1.06	1.19	1.25	2.01	1.92	1.98
	0.25	0.98	1.12	0.98	1.13	1.27	1.23	1.13	1.70	1.12	1.58	1.76	1.77
	0.50	0.78	0.87	0.93	1.05	0.99	1.07	1.45	1.17	1.16	1.05	2.10	2.08
	0.75	0.98	0.71	0.93	1.06	1.22	1.18	1.05	0.99	1.09	1.01	1.61	1.65
	0.90	1.65	1.12	1.36	1.47	2.4	2.01	1.01	0.98	1.26	0.81	1.85	1.90
$v = 3$	0.10	-	-	-	-	-	-	1.34	1.85	1.18	1.76	2.15	2.27
	0.25	-	-	-	-	-	-	1.01	1.32	1.01	1.42	1.49	1.54
	0.50	-	-	-	-	-	-	1.13	1.05	1.33	1.29	1.54	1.62
	0.75	-	-	-	-	-	-	1.03	0.98	1.05	0.87	1.49	1.50
	0.90	-	-	-	-	-	-	1.51	0.99	1.43	1.03	2.55	2.29
		$\rho = 0.5$											
$v = 1$	0.10	1.06	1.65	1.14	1.68	1.58	1.48	0.89	1.67	1.28	1.88	1.30	1.37
	0.25	0.98	1.23	0.97	1.20	1.27	1.21	1.02	1.38	0.87	1.01	1.36	1.29
	0.50	1.15	1.09	1.05	1.12	1.46	1.44	1.28	1.01	0.89	0.91	1.63	1.45
	0.75	1.05	0.78	1.01	0.89	1.35	1.28	1.04	0.78	0.88	0.72	1.45	1.34
	0.90	0.89	0.76	0.84	0.76	1.13	1.05	0.88	0.76	1.43	1.01	1.29	1.28
$v = 2$	0.10	1.08	1.32	0.89	1.20	1.54	1.23	1.03	1.61	1.16	1.72	1.41	1.34
	0.25	1.02	1.21	1.01	1.28	1.36	1.31	0.97	1.23	0.93	1.23	1.25	1.20
	0.50	0.98	0.97	1.04	1.21	1.13	1.23	1.06	1.01	0.97	1.02	1.29	1.30
	0.75	0.89	0.76	0.90	0.97	1.13	1.12	1.04	0.65	1.02	0.89	1.34	1.31
	0.90	0.98	0.79	0.84	0.72	1.38	1.14	1.01	0.99	1.22	0.87	1.45	1.32
$v = 3$	0.10	-	-	-	-	-	-	1.04	1.38	0.93	1.38	1.49	1.34
	0.25	-	-	-	-	-	-	1.01	1.25	0.99	1.31	1.28	1.27
	0.50	-	-	-	-	-	-	1.05	1.10	1.11	1.27	1.30	1.37
	0.75	-	-	-	-	-	-	1.09	0.79	1.07	0.97	1.46	1.41
	0.90	-	-	-	-	-	-	0.97	0.89	0.85	0.87	1.39	1.13

Following the same procedure, we will get:

$$B \cong rkm_1(F(t)) \text{ and } C \cong rm(m - 2k)(F(t)). \tag{5.4}$$

By substituting (5.3) and (5.4) in (5.2), this advocates the consistency of $\hat{F}^*(t)$ which completes the proof. \square

It is pertinent to mention that in the virtue of the preceding proved propositions, one can easy investigate that $\hat{F}^*(t)$ is asymptotically normally distributed. Further, by replacing $\hat{F}(t)$ in (5.1) with our proposed CDF estimator early mentioned, $\{\hat{F}_E(t), \hat{F}_M(t), \hat{F}_K(t)\}$, yields respectively to $\{\hat{F}_{E1}(t), \hat{F}_{M1}(t), \hat{F}_{K1}(t)\}$.

6. SIMULATION STUDY

To assess the performance of the proposed procedures, a simulation study is conducted for various values of $F(t), m, k, m_1$ and v when the underlying distribution for the data are the standard normal (symmetric distribution) as well as the standard log normal (asymmetric distribution). Aiming better analyzing the simulated results, the effect of the quality of ranking process is also taken into account by adopting Dell and Clutter [11]'s imperfect ranking model with correlation coefficient ρ . Three different configurations of ρ are considered: $\rho = 1$ for perfect ranking, $\rho = 0.9$ for imperfect ranking with reasonable good accuracy, and $\rho = 0.5$ for imperfect

TABLE 2. The RE values of the CDF estimators using simulated normal distribution when $m = 5, k = 1$.

		$\rho = 1$											
		$m_1 = 4$						$m_1 = 10$					
F		\hat{F}_E	\hat{F}_{E1}	\hat{F}_M	\hat{F}_{M1}	\hat{F}_K	\hat{F}_{K1}	\hat{F}_E	\hat{F}_{E1}	\hat{F}_M	\hat{F}_{M1}	\hat{F}_K	\hat{F}_{K1}
$v = 1$	0.10	1.04	1.39	1.02	1.35	1.58	1.56	0.87	1.67	1.29	1.99	1.42	1.54
	0.25	1.03	1.01	0.89	1.01	1.45	1.37	1.18	1.43	1.04	1.15	1.91	1.99
	0.50	1.15	0.86	0.98	0.82	1.32	1.29	1.01	0.98	0.87	0.71	1.56	1.42
	0.75	0.89	0.76	0.88	0.87	1.43	1.38	1.18	0.98	1.04	0.93	2.13	2.10
	0.90	1.01	0.76	1.01	0.79	1.53	1.50	0.87	0.81	1.43	0.92	1.21	1.43
$v = 2$	0.10	1.30	1.33	1.04	1.19	1.78	1.65	1.07	1.39	1.00	1.35	2.00	1.96
	0.25	0.65	0.76	0.68	0.75	0.99	0.98	1.01	1.23	1.06	1.32	1.82	1.89
	0.50	0.65	0.79	1.06	1.01	0.88	0.98	1.22	1.09	1.25	1.09	1.76	1.85
	0.75	0.68	0.61	0.74	0.79	0.97	0.99	1.04	0.87	1.13	0.97	1.87	1.93
	0.90	1.18	0.97	1.05	1.01	1.57	1.45	1.05	0.87	1.13	0.97	1.91	1.99
$v = 3$	0.10	-	-	-	-	-	-	1.26	1.32	1.06	1.22	1.87	1.80
	0.25	-	-	-	-	-	-	0.76	0.83	0.97	0.95	1.26	1.32
	0.50	-	-	-	-	-	-	0.87	0.98	1.32	1.23	1.14	1.32
	0.75	-	-	-	-	-	-	0.97	0.88	0.99	0.91	1.24	1.31
	0.90	-	-	-	-	-	-	1.12	1.02	1.15	1.48	1.98	1.85
		$\rho = 0.9$											
$v = 1$	0.10	0.97	1.25	1.10	1.26	1.47	1.49	0.89	1.45	1.33	1.87	1.57	1.88
	0.25	1.03	1.11	0.94	1.13	1.44	1.39	1.15	1.35	1.02	1.10	1.91	1.88
	0.50	1.11	0.98	1.03	1.05	1.21	1.26	1.17	1.01	0.89	0.88	1.56	1.45
	0.75	0.97	0.78	0.94	0.86	1.45	1.37	1.16	0.87	1.05	0.89	1.78	1.76
	0.90	0.98	0.65	1.05	0.75	1.55	1.41	0.96	0.72	1.15	0.87	1.26	1.37
$v = 2$	0.10	1.25	1.28	1.08	1.04	1.71	1.45	1.05	1.35	1.01	1.34	1.75	1.69
	0.25	0.76	0.87	0.73	0.88	1.04	1.03	0.89	1.01	0.88	1.03	1.37	1.40
	0.50	0.78	0.96	1.22	1.25	0.90	1.01	0.92	0.95	0.93	0.98	1.21	1.29
	0.75	0.86	0.76	0.88	0.89	1.14	1.16	0.97	0.77	1.05	0.87	1.48	1.50
	0.90	1.22	0.86	1.02	0.94	1.74	1.46	1.12	0.89	1.18	0.97	1.81	1.73
$v = 3$	0.10	-	-	-	-	-	-	1.27	1.40	0.99	1.23	1.99	1.76
	0.25	-	-	-	-	-	-	0.87	0.99	0.84	0.98	1.27	1.33
	0.50	-	-	-	-	-	-	0.88	0.98	1.23	1.21	1.23	1.32
	0.75	-	-	-	-	-	-	0.83	0.80	0.86	0.78	1.26	1.29
	0.90	-	-	-	-	-	-	0.96	0.84	0.88	0.79	1.25	1.29
		$\rho = 0.5$											
$v = 1$	0.10	0.98	1.09	0.86	0.98	1.41	1.12	0.97	1.25	1.05	1.32	1.37	1.25
	0.25	1.02	1.22	1.02	1.20	1.36	1.33	1.14	1.36	1.01	1.19	1.45	1.49
	0.50	1.03	1.15	1.18	1.30	1.23	1.43	1.17	0.97	0.76	0.83	1.21	1.24
	0.75	1.01	0.88	1.01	0.95	1.33	1.29	1.02	0.78	0.97	0.88	1.43	1.40
	0.90	0.98	0.65	0.83	0.76	1.46	1.16	0.93	0.75	1.28	0.98	1.38	1.30
$v = 2$	0.10	1.12	1.11	0.76	0.86	1.65	1.16	1.13	1.25	0.95	1.14	1.65	1.55
	0.25	0.95	1.13	0.94	1.15	1.29	1.28	0.96	1.15	0.98	1.14	1.32	1.28
	0.50	0.99	1.22	1.57	1.65	1.13	1.44	0.94	1.09	1.01	1.14	1.13	1.31
	0.75	0.92	0.78	0.90	0.89	1.25	1.22	1.02	0.87	1.04	0.93	1.39	1.35
	0.90	1.10	0.65	0.70	0.60	1.58	1.13	0.89	0.87	0.83	0.78	1.24	1.10
$v = 3$	0.10	-	-	-	-	-	-	1.04	1.07	0.87	0.98	1.46	1.07
	0.25	-	-	-	-	-	-	0.95	1.15	0.98	1.19	1.34	1.32
	0.50	-	-	-	-	-	-	0.99	1.14	1.34	1.45	1.17	1.27
	0.75	-	-	-	-	-	-	0.97	0.86	0.96	0.89	1.26	1.23
	0.90	-	-	-	-	-	-	1.03	0.87	0.78	0.76	1.45	1.11

ranking. The number of simulation runs is 10 000. The comparisons between these estimators are made in terms of relative efficiency (RE) defined as:

$$RE(t) = \frac{V(\hat{F}_S(t))}{MSE(\hat{F}_L(t))}, L \in \{E, E1, M, M1, K \text{ and } K1\},$$

where $\hat{F}_E(t)$ ($\hat{F}_{E1}(t)$), $\hat{F}_M(t)$ ($\hat{F}_{M1}(t)$) and $\hat{F}_K(t)$ ($\hat{F}_{K1}(t)$) are the CDF estimators under VLRSS using empirical function, MLE and kernel function based on measured (unmeasured) units, respectively. In addition, $\hat{F}_S(t)$ are the traditional CDF estimator proposed by Stokes and Sager [29] under RSS. With the above definition, an RE(t) larger than one implies that $\hat{F}_L(t)$ asymptotically outperforms $\hat{F}_S(t)$ at the point t and *vice versa*. Due to space considerations, we set $r = 1$ as we observed that RE is slightly affected by the number of cycles. Based on Tables 1–6, the following remarks can be highlighted.

- Firstly, one can clearly observe that $\hat{F}_K(t)$ and $\hat{F}_{K1}(t)$ are somewhat better than the other CDF estimators and usually have REs greater than 1 in most considered cases.
- It seems that the information supported by the unmeasured units are useful when t lies at the lower tail of the population distribution even when the quality of ranking is weak.

TABLE 3. The RE values of the CDF estimators using simulated normal distribution when $m = 5, k = 2$.

		$\rho = 1$											
		$m_1 = 4$						$m_1 = 10$					
v	F	\hat{F}_E	\hat{F}_{E1}	\hat{F}_M	\hat{F}_{M1}	\hat{F}_K	\hat{F}_{K1}	\hat{F}_E	\hat{F}_{E1}	\hat{F}_M	\hat{F}_{M1}	\hat{F}_K	\hat{F}_{K1}
$v = 1$	0.10	0.67	1.01	0.88	1.21	0.98	1.11	0.76	0.98	1.11	1.21	0.98	0.87
	0.25	0.89	1.30	0.96	1.18	1.55	1.57	0.86	2.01	1.02	1.14	1.12	1.43
	0.50	1.27	0.97	0.80	0.76	1.85	1.53	2.45	1.56	0.66	0.77	7.34	5.32
	0.75	0.90	0.76	0.92	0.87	1.48	1.44	0.76	0.65	0.73	0.55	1.10	1.23
	0.90	0.67	0.45	0.82	0.65	0.90	1.09	0.64	0.45	1.12	0.76	0.98	0.91
$v = 2$	0.10	1.25	1.23	0.99	1.11	1.61	1.47	0.87	1.19	1.07	1.45	1.53	1.80
	0.25	0.67	0.60	0.65	0.62	0.76	0.78	0.95	1.43	0.94	1.32	1.94	2.05
	0.50	0.54	0.64	0.87	0.90	0.64	0.80	1.94	1.24	0.97	0.86	3.55	3.43
	0.75	0.65	0.57	0.60	0.80	0.83	0.87	0.98	0.87	1.04	0.76	1.96	2.12
	0.90	1.19	0.98	0.93	1.09	1.65	1.48	0.87	0.76	1.32	1.04	1.74	2.06
$v = 3$	0.10	–	–	–	–	–	–	1.17	1.26	1.04	1.19	2.05	1.95
	0.25	–	–	–	–	–	–	0.87	0.98	0.88	1.04	1.65	1.68
	0.50	–	–	–	–	–	–	1.41	1.36	1.86	1.54	1.97	2.24
	0.75	–	–	–	–	–	–	0.86	0.76	0.96	0.76	1.66	1.72
	0.90	–	–	–	–	–	–	1.12	0.99	1.18	1.48	2.21	2.29
		$\rho = 0.9$											
$v = 1$	0.10	0.78	1.16	0.99	1.37	1.08	1.22	0.76	0.87	0.94	1.34	0.97	0.87
	0.25	0.98	1.23	0.90	1.09	1.41	1.37	0.87	1.65	0.87	0.88	1.27	1.51
	0.50	1.28	1.07	0.89	0.86	1.79	1.58	2.24	1.32	0.87	0.65	4.23	3.05
	0.75	0.92	0.65	0.88	0.67	1.42	1.45	0.87	0.97	0.76	0.57	1.34	1.58
	0.90	0.77	0.54	0.97	0.65	1.05	1.19	0.76	0.87	0.94	0.64	0.76	0.88
$v = 2$	0.10	1.45	1.40	0.96	1.14	1.95	1.66	0.98	1.29	1.18	1.59	1.43	1.61
	0.25	0.76	0.84	0.77	0.87	1.01	1.02	0.98	1.37	1.04	1.24	1.75	1.79
	0.50	0.69	0.87	1.15	1.19	0.88	1.05	1.69	1.39	0.99	0.96	2.38	2.20
	0.75	0.76	0.68	0.77	0.89	1.04	1.06	1.01	0.87	1.05	0.89	1.87	1.94
	0.90	1.25	0.89	0.91	1.13	1.70	1.45	0.90	0.87	1.45	1.04	1.56	1.76
$v = 3$	0.10	–	–	–	–	–	–	1.21	1.37	1.01	1.25	1.88	1.75
	0.25	–	–	–	–	–	–	0.87	1.03	0.98	1.08	1.37	1.43
	0.50	–	–	–	–	–	–	0.90	0.97	1.12	1.06	1.19	1.36
	0.75	–	–	–	–	–	–	0.89	0.80	0.99	0.91	1.40	1.45
	0.90	–	–	–	–	–	–	1.31	0.98	1.13	1.08	2.05	1.85
		$\rho = 0.5$											
$v = 1$	0.10	0.81	1.05	0.91	1.09	1.18	1.06	0.89	1.32	1.17	1.48	1.11	1.22
	0.25	1.07	1.27	0.99	1.13	1.49	1.42	1.02	1.33	0.86	0.79	1.41	1.38
	0.50	1.14	1.20	0.99	1.09	1.38	1.48	1.25	1.15	0.78	0.70	1.65	1.53
	0.75	0.98	0.78	0.93	0.87	1.33	1.26	1.20	0.98	0.84	0.81	1.75	1.70
	0.90	0.98	0.76	1.08	0.77	1.32	1.19	0.87	0.78	1.68	1.14	1.25	1.37
$v = 2$	0.10	1.01	1.01	0.63	0.78	1.40	1.01	0.97	1.20	1.10	1.31	1.35	1.17
	0.25	0.91	1.09	0.92	1.10	1.21	1.20	1.01	1.25	0.98	1.10	1.38	1.43
	0.50	0.90	1.16	1.54	1.61	1.06	1.37	0.98	1.05	0.79	0.89	1.21	1.37
	0.75	0.96	0.83	0.96	0.93	1.27	1.25	1.01	0.78	0.98	0.88	1.40	1.35
	0.90	1.25	0.71	0.71	0.65	1.70	1.20	0.87	0.76	1.10	0.99	1.27	1.12
$v = 3$	0.10	–	–	–	–	–	–	1.04	1.16	0.87	0.97	1.46	1.53
	0.25	–	–	–	–	–	–	0.98	1.21	0.94	1.25	1.34	1.37
	0.50	–	–	–	–	–	–	0.98	1.14	1.29	1.35	1.18	1.50
	0.75	–	–	–	–	–	–	1.05	0.87	1.01	0.98	1.39	1.36
	0.90	–	–	–	–	–	–	1.14	0.87	0.75	0.71	1.65	1.25

- Changing the underlying distribution does not substantially affect on the patterns of the studied estimators regardless a few cases. In several cases, the REs are higher when the parent distribution is log normal distribution rather than normal distribution.
- All the REs are affected strongly by the location of the value t . Generally speaking, $\hat{F}_E(t)$ outperforms its counterpart in RSS when t closes to the center of the population distribution and $v = 1$. Yet if $v > 1$, $\hat{F}_E(t)$ performs well when t tends to at least one of the boundaries. On the other hand, $\hat{F}_M(t)$ is becoming more efficient than $\hat{F}_S(t)$ if t is near to at least one of the boundaries with some exceptions. Interestingly, these results are generally valid regardless the quality ranking.
- The quality ranking has a positive effect on the performance of the studied estimators particularly for large m and m_1 . As expectedly, this effect becomes stronger on the performance of the estimators based on missing data approach and weaker for the others.
- With keeping all the factors are fixed, increasing either m_1 and v or m_1 and k can provide improvement for the behavior of the almost studied estimators in some cases.
- It is also observed that increasing m with keeping the other factors are fixed may not improve the REs even if the ranking process is perfectly done. Further, in several cases, the REs of almost studied estimators is higher when $m_1 \geq m$ than $m_1 < m$.

TABLE 4. The RE values of the CDF estimators using simulated log-normal distribution when $m = 3, k = 1$.

		$\rho = 1$											
		$m_1 = 4$						$m_1 = 10$					
F		\hat{F}_E	\hat{F}_{E1}	\hat{F}_M	\hat{F}_{M1}	\hat{F}_K	\hat{F}_{K1}	\hat{F}_E	\hat{F}_{E1}	\hat{F}_M	\hat{F}_{M1}	\hat{F}_K	\hat{F}_{K1}
$v = 1$	0.10	0.97	1.76	1.08	1.94	1.78	2.57	0.76	1.76	1.19	2.54	1.55	2.55
	0.25	1.07	1.54	1.08	1.45	3.11	2.56	1.15	1.97	0.87	0.98	4.34	6.44
	0.50	1.29	1.02	1.05	0.96	1.88	1.65	1.76	1.01	0.76	0.66	3.01	2.22
	0.75	1.12	0.87	1.16	0.78	1.17	1.20	1.18	0.77	0.91	0.86	0.98	0.99
	0.90	0.87	0.76	0.88	0.73	0.86	0.89	0.76	0.65	1.13	0.87	0.78	0.88
$v = 2$	0.10	1.76	2.04	1.54	1.99	3.45	2.99	0.98	1.65	1.12	1.96	1.99	3.45
	0.25	0.87	0.98	0.87	0.99	1.49	1.39	1.05	1.76	1.05	1.65	3.76	4.17
	0.50	0.87	0.68	0.87	0.95	0.94	0.87	1.55	1.22	1.22	1.05	2.34	2.38
	0.75	0.87	0.75	0.86	1.03	1.01	1.01	1.24	0.76	1.25	0.87	1.45	1.48
	0.90	1.65	1.35	1.56	2.15	2.10	1.97	1.02	0.87	1.28	0.87	1.35	1.39
$v = 3$	0.10	-	-	-	-	-	-	1.65	2.22	1.65	2.17	3.75	4.22
	0.25	-	-	-	-	-	-	1.05	1.44	1.04	1.47	2.97	3.02
	0.50	-	-	-	-	-	-	1.26	1.15	1.74	1.37	1.65	1.97
	0.75	-	-	-	-	-	-	1.09	0.78	1.11	0.87	1.46	1.43
	0.90	-	-	-	-	-	-	1.76	1.26	1.79	1.87	2.25	2.15
		$\rho = 0.9$											
$v = 1$	0.10	0.98	1.78	1.10	1.89	1.80	2.16	0.76	1.76	1.13	1.98	1.17	2.65
	0.25	1.14	1.63	1.14	1.55	3.01	2.45	1.13	1.78	0.87	0.99	4.32	4.70
	0.50	1.25	1.08	1.06	1.05	1.61	1.48	1.92	1.20	0.75	0.87	2.94	2.29
	0.75	1.02	0.65	1.01	0.76	1.06	1.04	1.21	0.67	0.98	0.89	1.13	1.10
	0.90	0.97	0.65	1.11	0.65	1.03	1.13	0.65	0.54	1.09	0.65	0.87	0.93
$v = 2$	0.10	1.54	1.76	1.29	1.64	2.98	2.45	1.07	2.01	1.29	2.23	2.17	3.25
	0.25	0.98	1.09	0.96	1.10	1.66	1.55	1.06	1.54	1.05	1.44	2.88	2.80
	0.50	0.78	0.76	0.88	0.98	0.91	0.98	1.56	1.26	1.23	1.13	2.19	2.10
	0.75	1.01	0.89	0.97	0.96	1.12	1.11	1.16	0.87	1.19	0.87	1.32	1.35
	0.90	1.54	1.07	1.27	1.41	1.88	1.65	1.08	0.87	1.43	0.87	1.34	1.35
$v = 3$	0.10	-	-	-	-	-	-	1.40	2.02	1.31	1.97	2.96	3.07
	0.25	-	-	-	-	-	-	0.97	1.19	0.95	1.24	2.08	2.04
	0.50	-	-	-	-	-	-	1.04	0.99	1.19	1.08	1.28	1.33
	0.75	-	-	-	-	-	-	1.07	0.78	1.09	0.88	1.29	1.34
	0.90	-	-	-	-	-	-	1.38	0.87	1.33	1.01	1.76	1.65
		$\rho = 0.5$											
$v = 1$	0.10	1.05	1.54	1.10	1.60	1.95	1.87	1.13	1.39	1.09	1.36	1.98	1.33
	0.25	0.99	1.25	0.98	1.22	1.95	1.64	1.01	1.35	0.78	0.99	2.15	1.89
	0.50	1.05	1.01	0.99	1.05	1.33	1.30	1.20	1.01	0.78	0.72	1.45	1.35
	0.75	0.99	0.76	0.97	0.79	1.06	1.04	1.08	0.65	0.88	0.78	1.12	1.08
	0.90	1.08	0.79	1.11	0.77	1.15	1.14	0.98	0.76	1.35	0.76	0.90	0.95
$v = 2$	0.10	1.26	1.44	1.01	1.33	2.23	1.65	1.07	1.32	1.03	1.34	1.91	1.21
	0.25	1.01	1.19	1.01	1.25	1.68	1.56	0.97	1.19	0.87	1.19	1.59	1.50
	0.50	0.97	1.01	1.08	1.25	1.13	1.25	1.01	0.97	0.88	0.98	1.24	1.27
	0.75	0.96	0.79	0.97	0.99	1.10	1.05	0.98	0.67	0.98	0.83	1.03	1.06
	0.90	1.06	0.76	0.89	0.88	1.20	1.09	1.05	0.87	1.25	0.65	1.14	1.11
$v = 3$	0.10	-	-	-	-	-	-	0.97	1.20	0.87	1.18	1.75	1.11
	0.25	-	-	-	-	-	-	1.09	1.40	1.04	1.45	1.97	1.81
	0.50	-	-	-	-	-	-	0.99	0.98	1.06	1.17	1.22	1.32
	0.75	-	-	-	-	-	-	1.05	0.85	1.03	0.90	1.14	1.10
	0.90	-	-	-	-	-	-	1.35	0.86	1.20	0.76	1.45	1.32

As overall, we can say that the conventional RSS-based CDF is outperformed by at least one of VLRSS-based CDF estimators in almost the considered cases, particularly when t lies at least one of the boundaries of the parent distribution (either symmetric or asymmetric). In addition, it is clearly observed from the superiority of $\hat{F}_K(t)$ and $\hat{F}_{K1}(t)$ over $\hat{F}_S(t)$ in most cases and they are mostly alternating in the first and second places among all the considered estimators.

7. AN EMPIRICAL STUDY

In what follows, we will illustrate the applicability of the proposed CDF by using real data set known as tree dataset found in Chen *et al.* [9]. This data set of size 396 observations and includes seven variables. Here, we will restrict our attention on two variables: “the entire height in feet” denoted by the response variable and the “diameter in centimeters at breast height” denoted by the concomitant variable. The linear correlation between “the entire height in feet” and “diameter in centimeters at breast height” equals 0.91 indicating that the quality ranking is nearly perfect.

We will consider the tree dataset as the hypothetical population. For the same values of $F(t), m, k, m_1$ and v shown by Tables 1–6, 10 000 samples with replacement are drawn using RSS and VLRSS schemes. Again for each the selected sample, the $\hat{F}_L(t)$ and $\hat{F}_s(t)$ are computed, then the ARE for all the considered CDF estimators are estimated and listed in Table 7.

TABLE 5. The RE values of the CDF estimators using simulated log-normal distribution when $m = 5, k = 1$.

		$\rho = 1$											
		$m_1 = 4$						$m_1 = 10$					
F		\hat{F}_E	\hat{F}_{E1}	\hat{F}_M	\hat{F}_{M1}	\hat{F}_K	\hat{F}_{K1}	\hat{F}_E	\hat{F}_{E1}	\hat{F}_M	\hat{F}_{M1}	\hat{F}_K	\hat{F}_{K1}
$v = 1$	0.10	1.12	1.48	1.12	1.50	2.18	2.55	0.98	1.93	1.43	2.35	1.67	2.90
	0.25	0.88	0.99	0.86	0.98	2.22	1.92	1.44	1.60	1.24	1.26	3.95	3.62
	0.50	1.08	0.83	0.90	0.85	1.07	1.05	1.20	0.86	0.75	0.70	1.36	1.23
	0.75	0.97	0.78	0.97	0.83	1.17	1.14	1.23	0.80	1.09	0.87	1.43	1.40
	0.90	0.95	0.65	1.14	0.78	1.11	1.13	0.92	0.76	1.54	0.98	1.05	1.23
$v = 2$	0.10	1.16	1.19	0.93	1.04	2.37	2.15	1.12	1.45	1.05	1.43	2.26	2.93
	0.25	0.77	0.67	0.87	0.82	1.23	1.26	1.05	1.21	1.09	1.28	2.87	2.76
	0.50	0.65	0.78	0.94	0.98	0.89	0.88	1.23	1.16	1.28	1.16	1.45	1.54
	0.75	0.78	0.66	0.67	0.85	0.86	0.89	0.87	0.67	0.94	0.75	1.25	1.24
	0.90	1.25	0.98	1.13	1.38	1.36	1.27	1.13	0.87	1.23	1.04	1.43	1.39
$v = 3$	0.10	–	–	–	–	–	–	1.23	1.30	0.98	1.17	2.97	2.99
	0.25	–	–	–	–	–	–	0.78	0.84	0.98	0.87	1.70	1.75
	0.50	–	–	–	–	–	–	0.87	0.93	1.27	1.15	1.05	1.28
	0.75	–	–	–	–	–	–	0.87	0.72	0.96	0.87	1.14	1.19
	0.90	–	–	–	–	–	–	1.17	0.91	1.08	0.93	1.39	1.32
		$\rho = 0.9$											
$v = 1$	0.10	1.12	1.45	1.07	1.43	2.22	2.25	0.87	1.45	1.16	1.65	1.43	2.23
	0.25	0.90	1.06	0.90	1.05	2.69	1.86	1.20	1.43	1.07	1.18	2.99	2.76
	0.50	1.07	0.93	0.98	0.99	1.08	1.15	1.11	0.98	0.83	0.80	1.28	1.21
	0.75	1.03	0.88	1.02	0.87	1.20	1.22	1.08	0.87	1.01	0.89	1.23	1.25
	0.90	1.09	0.83	1.04	0.87	1.23	1.20	0.91	0.58	1.48	0.97	1.09	1.19
$v = 2$	0.10	1.23	1.26	0.89	1.06	2.52	2.53	1.12	1.46	1.06	1.44	2.17	2.46
	0.25	0.76	0.84	0.75	0.88	1.31	1.33	0.92	1.10	0.97	1.14	2.11	2.06
	0.50	0.76	0.89	1.10	1.12	0.86	1.06	0.97	0.98	0.99	0.95	1.13	1.21
	0.75	0.79	0.70	0.78	0.85	0.95	0.97	1.01	0.79	1.09	0.91	1.20	1.23
	0.90	1.30	0.88	0.90	0.97	1.48	1.27	1.04	0.78	1.11	0.87	1.23	1.17
$v = 3$	0.10	–	–	–	–	–	–	1.01	1.32	0.98	1.25	2.16	2.40
	0.25	–	–	–	–	–	–	0.99	1.16	1.02	1.20	2.27	2.20
	0.50	–	–	–	–	–	–	0.99	1.01	1.01	1.01	1.20	1.27
	0.75	–	–	–	–	–	–	1.01	0.75	1.08	0.88	1.23	1.22
	0.90	–	–	–	–	–	–	1.16	0.78	1.22	0.99	1.44	1.37
		$\rho = 0.5$											
$v = 1$	0.10	0.94	1.05	0.78	0.95	1.71	1.81	1.16	1.54	1.33	1.64	2.03	2.01
	0.25	0.97	1.15	0.98	1.10	1.78	1.64	1.04	1.17	0.96	1.12	2.03	1.89
	0.50	0.98	1.08	1.13	1.20	1.08	1.30	1.04	1.28	0.87	0.92	1.18	1.23
	0.75	1.01	0.81	1.02	0.99	1.08	1.13	1.05	0.78	0.99	0.88	1.11	1.10
	0.90	1.08	0.78	1.13	0.74	1.20	1.10	0.97	0.57	1.30	0.80	1.06	0.99
$v = 2$	0.10	1.09	1.08	0.67	0.86	1.97	1.24	0.90	1.02	1.16	0.96	1.75	1.26
	0.25	1.01	1.21	1.01	1.22	1.80	1.74	0.94	1.12	0.91	1.10	1.66	1.55
	0.50	0.98	1.14	1.15	1.25	1.06	1.40	0.98	1.09	1.03	1.16	1.12	1.32
	0.75	0.95	0.77	0.91	0.87	1.01	1.02	1.04	0.87	1.05	0.97	1.15	1.18
	0.90	1.12	0.76	1.06	0.87	1.24	1.04	0.98	0.78	1.11	0.67	1.15	0.99
$v = 3$	0.10	–	–	–	–	–	–	1.09	1.15	0.76	0.98	2.08	1.38
	0.25	–	–	–	–	–	–	0.97	1.14	0.99	1.16	1.70	1.60
	0.50	–	–	–	–	–	–	0.98	1.11	1.29	1.37	1.10	1.34
	0.75	–	–	–	–	–	–	0.93	0.89	0.98	0.88	1.04	1.03
	0.90	–	–	–	–	–	–	1.12	0.65	0.77	0.76	1.23	0.99

TABLE 6. The RE values of the CDF estimators using simulated log-normal distribution when $m = 5, k = 2$.

		$\rho = 1$											
		$m_1 = 4$					$m_1 = 10$						
	F	\hat{F}_E	\hat{F}_{E1}	\hat{F}_M	\hat{F}_{M1}	\hat{F}_K	\hat{F}_{K1}	\hat{F}_E	\hat{F}_{E1}	\hat{F}_M	\hat{F}_{M1}	\hat{F}_K	\hat{F}_{K1}
$v = 1$	0.10	0.76	1.07	0.88	1.24	1.13	1.71	0.53	0.86	0.97	1.64	1.19	1.46
	0.25	0.98	1.33	0.97	1.17	3.24	2.65	1.35	1.58	1.19	1.26	3.88	3.68
	0.50	1.31	1.01	0.88	0.77	1.54	1.33	1.05	0.89	0.77	0.73	1.30	1.19
	0.75	0.86	0.65	0.87	0.60	0.88	0.92	1.21	0.84	1.07	0.88	1.45	1.40
	0.90	0.77	0.54	1.09	0.65	0.77	0.86	0.87	0.67	1.56	0.98	1.04	1.23
$v = 2$	0.10	1.25	1.34	0.99	1.10	2.92	1.97	0.78	1.22	1.08	1.45	1.32	2.44
	0.25	0.76	0.78	0.65	0.69	1.12	1.16	1.05	1.25	1.10	1.32	3.13	2.98
	0.50	0.55	0.74	0.97	1.01	0.65	0.88	1.13	1.06	1.17	1.07	1.35	1.44
	0.75	0.87	0.60	0.65	0.79	0.76	0.78	0.98	0.77	1.05	0.87	1.31	1.33
	0.90	1.29	1.01	1.18	1.43	1.40	1.27	1.12	0.78	1.20	0.98	1.37	1.33
$v = 3$	0.10	-	-	-	-	-	-	1.22	1.29	1.06	1.21	2.81	3.09
	0.25	-	-	-	-	-	-	0.78	0.86	0.88	0.92	1.71	1.85
	0.50	-	-	-	-	-	-	1.08	1.13	1.65	1.38	1.26	1.54
	0.75	-	-	-	-	-	-	0.83	0.69	0.88	0.79	1.05	1.09
	0.90	-	-	-	-	-	-	1.11	0.89	1.01	1.23	1.25	1.20
		$\rho = 0.9$											
$v = 1$	0.10	0.87	1.32	1.13	1.54	1.43	2.04	0.77	0.93	1.01	1.59	1.79	1.65
	0.25	0.98	1.18	0.87	1.05	2.55	2.06	1.07	1.31	0.98	1.09	2.75	2.65
	0.50	1.34	1.15	0.95	0.92	1.45	1.44	1.07	0.89	0.98	0.76	1.31	1.20
	0.75	0.94	0.76	0.89	0.69	0.99	1.01	1.18	0.87	1.07	0.89	1.35	1.32
	0.90	0.88	0.65	1.19	0.87	0.97	1.01	0.92	0.77	1.54	0.94	1.04	1.14
$v = 2$	0.10	1.25	1.26	0.88	1.04	2.56	1.96	0.94	1.43	1.32	1.76	1.68	2.48
	0.25	0.78	0.80	0.71	0.83	1.25	1.29	1.03	1.24	1.06	1.27	2.49	2.41
	0.50	0.62	0.77	1.01	1.04	0.76	0.91	1.95	0.99	0.93	0.97	1.09	1.17
	0.75	0.78	0.70	0.76	0.89	0.99	0.95	1.04	0.83	1.14	0.95	1.22	1.27
	0.90	1.26	0.87	1.08	1.11	1.40	1.21	1.10	0.76	1.18	0.97	1.25	1.21
$v = 3$	0.10	-	-	-	-	-	-	1.21	1.37	1.01	1.23	2.41	2.34
	0.25	-	-	-	-	-	-	0.85	0.97	0.88	0.99	1.75	1.78
	0.50	-	-	-	-	-	-	0.80	0.90	1.12	1.14	0.96	1.13
	0.75	-	-	-	-	-	-	0.83	0.70	0.87	0.79	1.03	1.06
	0.90	-	-	-	-	-	-	1.23	0.87	1.03	1.01	1.44	1.31
		$\rho = 0.5$											
$v = 1$	0.10	1.01	1.27	1.12	1.32	1.88	1.63	0.76	1.24	1.30	1.71	1.45	1.81
	0.25	0.97	1.25	0.90	1.09	1.85	1.66	0.98	1.18	0.87	1.02	1.87	1.76
	0.50	1.02	1.08	0.89	1.01	1.17	1.26	1.07	1.08	0.87	0.96	1.18	1.25
	0.75	1.08	0.77	0.89	0.87	1.04	1.06	1.06	0.80	1.01	0.90	1.13	1.12
	0.90	0.98	0.66	1.01	0.78	1.07	1.05	1.10	0.76	1.56	0.96	1.20	1.11
$v = 2$	0.10	1.03	1.01	0.64	0.77	1.80	1.20	1.04	1.29	1.18	1.38	1.86	1.54
	0.25	1.02	1.22	1.02	1.25	1.82	1.70	1.09	1.33	1.07	1.31	2.04	1.92
	0.50	1.01	1.15	1.53	1.60	1.31	1.34	0.94	0.99	0.96	1.05	1.02	1.19
	0.75	0.98	0.87	0.96	0.99	1.07	1.11	0.94	0.76	0.90	0.80	1.03	1.02
	0.90	1.12	0.67	0.66	0.59	1.19	0.99	1.02	0.77	0.98	0.70	1.16	1.01
$v = 3$	0.10	-	-	-	-	-	-	1.17	1.26	0.87	1.04	2.23	1.48
	0.25	-	-	-	-	-	-	0.99	1.19	0.99	1.22	1.84	1.80
	0.50	-	-	-	-	-	-	0.97	1.16	1.33	1.42	1.10	1.37
	0.75	-	-	-	-	-	-	0.99	0.77	0.98	0.85	1.03	1.07
	0.90	-	-	-	-	-	-	1.13	0.65	0.78	0.59	1.25	0.99

It is clear that the results presented in Table 7 are consistent with those shown in Tables 1–6. As $\hat{F}_S(t)$ is outperformed by at least one of VLRSS-based CDF estimators particularly when $F(t) \rightarrow 0$. Moreover, $\hat{F}_E(t)$ performs consistently well relative to its analog in RSS when $v = 1$ and the values of t around the center of the parent distribution. Also incorporating the information generated by the unmeasured items increases the efficiencies of the suggested CDF estimators provided that t closes to the lower tail of the population distribution. Further, increasing both m_1 and v provides sometimes sizeable benefits for the efficiencies $\hat{F}_K(t)$ and $\hat{F}_{K1}(t)$. For additional detailed results concerned to one special case is given by Figure 3, which presents the population CDF, $\hat{F}_K(t)$ and $\hat{F}_{K1}(t)$ based on tree dataset with $m = 5, k = 1, m_1 = 10, v = 3$ and $r = 10$. It is apparent that the lines of the $\hat{F}_K(t)$ and $\hat{F}_{K1}(t)$ are closer to the line of the true CDF when t becomes closer to the boundaries. Further at small (large) values of t , the performance of $\hat{F}_K(t)$ ($\hat{F}_{K1}(t)$) becomes better. Finally, one may be have to mention that all the numerical results displayed below are coded using R package and it is available upon request from the author.

8. CONCLUSION

This study is concerned with CDF estimation under VLRSS. Three novel estimators for CDF using VLRSS are suggested. It is theoretically shown that these estimators are consistent to the population CDF $F(t)$. By incorporating the information generated from the unmeasured sampling items, a general class of CDF estimators

TABLE 7. The RE values of the CDF estimators using Tree dataset.

		m = 3, k = 1											
		m ₁ = 4						m ₁ = 10					
F		\hat{F}_E	\hat{F}_{E1}	\hat{F}_M	\hat{F}_{M1}	\hat{F}_K	\hat{F}_{K1}	\hat{F}_E	\hat{F}_{E1}	\hat{F}_M	\hat{F}_{M1}	\hat{F}_K	\hat{F}_{K1}
v = 1	0.10	0.94	1.56	1.16	2.27	1.54	1.68	0.97	2.01	1.16	2.45	1.46	3.54
	0.25	1.03	1.34	1.05	1.23	2.65	1.64	1.21	2.30	0.98	1.11	4.54	5.87
	0.50	1.54	1.14	1.04	1.03	3.01	2.56	1.96	1.09	0.78	0.65	2.96	2.30
	0.75	1.19	0.94	1.18	0.87	1.86	1.80	1.20	0.87	0.98	0.76	1.43	1.44
	0.90	0.87	0.74	1.03	0.88	1.05	1.05	0.76	0.88	1.23	0.87	0.88	0.95
v = 2	0.10	1.01	2.01	1.46	1.91	2.28	2.49	0.97	2.01	1.17	2.13	2.27	3.97
	0.25	1.06	1.15	0.99	1.16	1.35	1.33	1.05	1.87	1.04	1.69	3.21	3.37
	0.50	0.87	0.90	0.98	0.96	1.21	1.27	1.65	1.16	1.27	1.07	2.15	2.08
	0.75	0.92	0.87	0.89	0.88	1.34	1.33	1.07	0.87	1.08	0.87	1.34	1.39
	0.90	1.34	1.15	1.54	1.43	2.34	2.13	0.87	0.56	1.18	0.66	0.98	1.01
v = 3	0.10	-	-	-	-	-	-	0.98	1.02	0.99	1.13	1.05	0.99
	0.25	-	-	-	-	-	-	0.98	1.02	0.99	1.16	1.06	0.99
	0.50	-	-	-	-	-	-	1.06	0.98	1.36	1.22	1.23	1.38
	0.75	-	-	-	-	-	-	0.78	0.67	0.81	0.70	0.94	0.96
	0.90	-	-	-	-	-	-	1.02	0.67	1.01	0.86	1.24	1.21
		m = 5, k = 1											
v = 1	0.10	0.98	1.85	1.14	2.11	1.43	1.53	0.87	1.82	1.26	2.12	1.66	2.95
	0.25	1.08	1.56	1.06	1.43	1.66	1.65	1.16	1.38	1.01	1.09	2.96	2.78
	0.50	1.43	1.10	1.02	0.99	2.15	1.85	1.04	0.89	0.77	0.76	1.18	1.08
	0.75	1.08	0.87	1.03	0.97	1.65	1.55	1.17	0.78	0.88	0.67	1.46	1.43
	0.90	0.98	0.87	1.23	0.76	1.43	1.54	0.87	0.76	0.99	0.76	0.81	0.91
v = 2	0.10	1.46	1.87	1.27	1.76	2.07	2.01	1.24	1.35	1.13	1.21	1.26	1.37
	0.25	0.97	1.25	0.99	1.28	1.36	1.39	0.98	1.13	0.97	1.18	2.28	2.16
	0.50	0.78	0.82	0.87	0.99	1.02	1.06	0.97	0.87	0.91	0.88	1.02	1.05
	0.75	0.99	0.78	0.94	0.96	1.34	1.22	0.92	0.54	0.97	0.76	1.18	1.21
	0.90	1.55	1.22	1.33	1.16	2.30	2.14	0.78	0.77	0.79	1.18	1.14	0.99
v = 3	0.10	-	-	-	-	-	-	1.17	1.44	1.14	1.47	1.18	1.14
	0.25	-	-	-	-	-	-	1.22	1.55	1.18	1.58	1.23	1.33
	0.50	-	-	-	-	-	-	1.31	1.44	1.26	1.65	1.41	1.47
	0.75	-	-	-	-	-	-	0.76	0.67	0.87	0.67	0.92	0.99
	0.90	-	-	-	-	-	-	1.23	0.98	1.32	1.05	1.54	1.55
		m = 5, k = 2											
v = 1	0.10	0.84	1.32	0.91	1.37	1.21	1.16	0.65	1.23	0.94	1.54	2.54	3.22
	0.25	0.98	1.25	0.99	1.17	1.29	1.21	0.98	1.65	0.87	0.98	4.34	7.84
	0.50	1.12	1.02	0.99	1.01	1.32	1.28	2.76	1.44	0.98	0.87	2.21	1.65
	0.75	1.13	0.87	1.06	0.89	1.43	1.40	0.98	0.89	0.86	0.76	0.93	1.02
	0.90	0.88	0.78	0.98	0.76	1.21	1.16	0.54	0.44	1.12	0.77	0.65	0.77
v = 2	0.10	1.14	1.45	0.98	1.39	1.16	1.44	0.87	1.43	1.24	1.87	1.44	2.65
	0.25	0.98	1.21	0.99	1.25	1.29	1.26	0.98	1.68	1.06	1.53	5.05	5.07
	0.50	1.02	1.06	1.15	1.33	1.23	1.32	1.98	1.35	1.07	0.98	1.76	1.67
	0.75	0.97	0.92	0.96	0.99	1.31	1.27	0.78	0.65	0.87	0.98	1.12	1.16
	0.90	1.10	0.72	0.95	0.73	1.65	1.22	0.87	0.68	1.23	0.95	1.19	1.07
v = 3	0.10	-	-	-	-	-	-	1.12	1.30	1.01	1.15	1.78	1.55
	0.25	-	-	-	-	-	-	0.96	1.14	0.95	1.24	2.25	2.32
	0.50	-	-	-	-	-	-	0.90	0.97	1.17	1.04	1.29	1.33
	0.75	-	-	-	-	-	-	0.89	0.80	0.99	0.91	1.40	1.45
	0.90	-	-	-	-	-	-	0.97	0.87	1.19	0.89	0.87	0.94

is also constructed which enables us to develop our proposed estimators. Additionally, the consistency of this class of CDF estimators is also analytically derived. Based on a small numerical experiments, we observe that the traditional RSS-based CDF is outperformed by at least one of VLRSS-based CDF estimators particularly when t lies at least one of the boundaries of the parent distribution even if the quality ranking is poor. It is also evident that kernel-based estimators are the best in almost considered cases. A considerable efficiency gain is obtained by incorporating the information generated from the unmeasured sampling items provided that t is near to the lower tail of the parent distribution. Thus we recommend to use $\hat{F}_{K1}(t)$ when $F(t) \rightarrow 0$ and the ranking quality is good enough. Otherwise, $\hat{F}_K(t)$ is the best choice. In a subsequent work, it may be of interest to compare our proposed estimators with the CDF estimator recently published by Eftekharian and Razmkhah [12], as they also used kernel approach under RSS. Additionally, constructing confidence interval for population CDF under VLRSS design is a much important future topic. The author plans to take these points in the near future.

Acknowledgements. The authors are very thankful to the referees and an editor for comments which led to a considerable improvement of an earlier version of this paper.

REFERENCES

- [1] A.I. Al-Omari, The efficiency of L ranked set sampling in estimating the distribution function. *Afrika Matematika* **26** (2015) 1457–1466.
- [2] A.I. Al-Omari, Quartile ranked set sampling for estimating the distribution function. *J. Egypt. Math. Soc.* **24** (2016) 303–308.
- [3] A.I. Al-Omari, Maximum likelihood estimation in location-scale families using varied L ranked set sampling. *RAIRO-Oper. Res.* **55** (2021) S2759–S2771.
- [4] A.I. Al-Omari and M.S. Abdallah, Estimation of the distribution function using moving extreme and MiniMax ranked set sampling. To appear in: *Commun. Stat. Simul. Comput.* (2021) 1–21. DOI: [10.1080/03610918.2021.1891433](https://doi.org/10.1080/03610918.2021.1891433).
- [5] M.F. Al-Saleh and D.M. Ahmad, Estimation of the distribution function using moving extreme ranked set sampling (MERSS). In: *Ranked Set Sampling: 65 years Improving the Accuracy in Data Gathering*. Academic Press (2019) 43–58.
- [6] S. Ashour and M. Abdallah, New distribution function estimators and tests of perfect ranking in concomitant-based ranked set sampling. To appear in: *Commun. Stat. Simul. Comput.* (2019) 1–26. DOI: [10.1080/03610918.2019.1659360](https://doi.org/10.1080/03610918.2019.1659360).
- [7] S. Ashour and M. Abdallah, Parametric estimation based on ranked set sampling: missing data approach. *J. Sci. Gazi Univ.* **32** (2019) 1356–1368.
- [8] S. Ashour and M. Abdallah, Estimation of distribution function based on ranked set sampling: missing data approach. *Thailand Stat.* **18** (2020) 27–42.
- [9] Z. Chen, Z. Bai and B.K. Sinha, *Ranked Set Sampling: Theory and Applications*. Springer, New York (2004).
- [10] W. Chen, C. Long, R. Yang and D. Yao, Maximum likelihood estimator of the location parameter under moving extremes ranked set sampling design. *Acta Math. App. Sin. Ser.* **37** (2021) 101–108.
- [11] T.R. Dell and J.L. Clutter, Ranked set sampling theory with order statistics background. *Biometrics* **28** (1972) 545–555.
- [12] A. Eftekharian and M. Razmkhah, On estimating the distribution function and odds using ranked set sampling. *Stat. Probab. Lett.* **122** (2017) 1–10.
- [13] J. Frey and L. Wang, EDF-based goodness-of-fit tests for ranked set sampling. *Can. J. Stat.* **42** (2014) 451–469.
- [14] J. Frey and Y. Zhang, Improved exact confidence intervals for a proportion using ranked-set sampling. *J. Korean Stat. Soc.* **48** (2019) 493–501.
- [15] J. Frey and Y. Zhang, Robust confidence intervals for a proportion using rankedset sampling. *J. Korean Stat. Soc.* **50** (2021) 1009–1028.
- [16] N. Gemayel, E. Stasny, J. Tackett and D. Wolfe, Ranked set sampling: an auditing application. *Rev. Quant. Finance Acc.* **39** (2012) 413–422.
- [17] A. Göçoğlu and N. Demirel, Estimating the population proportion in modified ranked set sampling methods. *J. Stat. Comput. Simul.* **89** (2019) 2694–2710.
- [18] A. Haq, J. Brown, E. Moltchanova and A.I. Al-Omari, Varied L ranked set sampling scheme. *J. Stat. Theory Pract.* **9** (2015) 741–767.
- [19] A. Hassan, A. Al-Omar and H. Nagy, Stress-strength reliability for the generalized inverted exponential distribution using MRSS. *Iran J. Sci. Technol. Trans. Sci.* **45** (2021) 641–659.
- [20] P.H. Kvam, Ranked set sampling based on binary water quality data with covariates. *J. Agric. Biol. Environ. Stat.* **8** (2003) 271–279.
- [21] P. Kvam and F. Samaniego, Nonparametric maximum likelihood estimation based on ranked set samples. *J. Am. Stat. Assoc.* **89** (1994) 526–537.
- [22] M. Mahdizadeh and E. Zamanzade, Kernel-based estimation of $P(x > y)$ in ranked set sampling. *SORT* **40** (2016) 243–266.
- [23] M. Mahdizadeh and E. Zamanzade, Efficient body fat estimation using multistage pair ranked set sampling. *Stat. Methods Med. Res.* **28** (2019) 223–234.
- [24] G.A. McIntyre, A method for unbiased selective sampling using ranked set sampling. *Aust. J. Agric. Res.* **3** (1952) 385–390.
- [25] H. Morabbi and M. Razmkhah, Quantile estimation based on modified ranked set sampling schemes using Pitman closeness. To appear in: *Commun. Stat. Simul. Comput.* (2020) DOI: [10.1080/03610918.2020.1811329](https://doi.org/10.1080/03610918.2020.1811329).
- [26] O. Ozturk, Nonparametric maximum-likelihood estimation of within-set ranking errors in ranked set sampling. *J. Nonparametric Stat.* **22** (2010) 823–840.
- [27] O. Ozturk, Parametric estimation of location and scale parameters in ranked set sampling. *J. Stat. Planning Inference* **141** (2011) 1616–1622.
- [28] A.M. Polansky and E.R. Baker, Multistage plugin bandwidth selection for kernel distribution function estimates. *J. Stat. Comput. Simul.* **65** (2000) 63–80.
- [29] S.L. Stokes and T.W. Sager, Characterization of a ranked-set sample with application to estimating distribution functions. *J. Am. Stat. Assoc.* **83** (1988) 374–381.
- [30] E. Zamanzade, EDF-based tests of exponentiality in pair ranked set sampling. *Stat. Papers* **60** (2019) 2141–2159.
- [31] E. Zamanzade and M. Mahdizadeh, A more efficient proportion estimator in ranked set sampling. *Stat. Probab. Lett.* **129** (2017) 28–33.
- [32] E. Zamanzade and M. Mahdizadeh, Estimating the population proportion in pair ranked set sampling with application to air quality monitoring. *J. Appl. Stat.* **45** (2018) 426–437.
- [33] E. Zamanzade and M. Mahdizadeh, Using ranked set sampling with extreme ranks in estimating the population proportion. *Stat. Methods Med. Res.* **29** (2020) 165–177.

- [34] E. Zamanzade and X. Wang, Estimation of population proportion for judgment post-stratification. *Comput. Stat. Data Anal.* **112** (2017) 257–269.
- [35] E. Zamanzade, A. Parvardeh and M. Asadi, Estimation of mean residual life based on ranked set sampling. *Comput. Stat. Data Anal.* **135** (2019) 35–55.
- [36] E. Zamanzade, M. Mahdizadeh and H. Samawi, Efficient estimation of cumulative distribution function using moving extreme ranked set sampling with application to reliability. *ASTA Adv. Stat. Anal.* **104** (2020) 485–502.

Subscribe to Open (S2O)

A fair and sustainable open access model



This journal is currently published in open access under a Subscribe-to-Open model (S2O). S2O is a transformative model that aims to move subscription journals to open access. Open access is the free, immediate, online availability of research articles combined with the rights to use these articles fully in the digital environment. We are thankful to our subscribers and sponsors for making it possible to publish this journal in open access, free of charge for authors.

Please help to maintain this journal in open access!

Check that your library subscribes to the journal, or make a personal donation to the S2O programme, by contacting subscribers@edpsciences.org

More information, including a list of sponsors and a financial transparency report, available at: <https://www.edpsciences.org/en/maths-s2o-programme>