DYNAMIC RELIABILITY ESTIMATION IN A RANK-BASED DESIGN

BY

M. MAHDIZADEH* (Sᴀʙᴢᴇᴠᴀʀ) ᴀɴᴅ **EHSAN ZAMANZADE** (Iꜱꜰᴀʜᴀɴ)

Abstract. Ranked set sampling (RSS) is a data collection method that allows us to direct attention toward measurements of more representative sample units. This article deals with estimating a time-dependent reliability measure under a generalization of the RSS. Some results concerning optimal properties of the proposed estimator are presented. Monte Carlo simulation is employed to assess performance of the estimator. A sport data set is finally analyzed.

2010 AMS Mathematics Subject Classification: Primary: 62G30, 62N05; Secondary: 62G05.

Key words and phrases: Covariate information, judgment ranking, stress-strength model.

1. INTRODUCTION

When the cost of identifying sampling units and approximately (judgment) ranking them according to the variable of interest is small compared to the cost of making formal quantifications, ranked set sampling (RSS) results in improved statistical inference over simple random sampling (SRS) of comparable size. This efficiency could be traced to the additional information provided by units that are ranked, but not actually quantified, and it holds across a wide variety of estimation and testing problems.

The RSS was first proposed by McIntyre [9] for use in estimating average yields in agriculture. While actually measuring a yield is costly because one must harvest the crops, an expert may be able to produce an accurate ranking of yields (in adjacent plots) based on a visual inspection. In general, the judgment ranking of the units in this scheme is usually done by using expert opinion, concomitant variable, or a combination of them, and need not be exact. In the decades since the original work by McIntyre [9] had appeared, the RSS was applied in a variety of fields including forestry, environmental science and medicine. The reader is

---

* Corresponding author.

referred to Chen [5] for some novel applications in areas such as clinical trials and genetic quantitative trait loci mappings. Also, a recent review of RSS can be found in Wolfe [13] and the references therein.

The RSS design can be summarized as follows:

1. Draw m random samples, each of size m, from the target population.

2. Apply judgment ordering, by any cheap method, on the elements of the ith ($i = 1, \ldots, m$) sample and identify the ith smallest unit.

3. Actually measure the m identified units in step 2.

4. Repeat steps 1–3, k times (cycles), if necessary, to obtain a ranked set sample of size mk.

Let X_{ij} be the ith judgment order statistic from the jth cycle. Then, the resulting ranked set sample is denoted by $\{X_{ij} : i = 1, \ldots, m\,; j = 1, \ldots, k\}$. The design parameter m is called a *set size*.

In practice, m should be kept small (e.g., 2–7) because judgment ranking with large set size is prone to ranking errors. For highly skewed distributions, small values of m do not lead to significant efficiency gain in estimating the population mean (see Takahasi and Wakimoto [12]). Multistage ranked set sampling (MSRSS) is a generalization of RSS that allows us to achieve higher efficiency, given a fixed set size.

The MSRSS scheme can be described as follows:

1. Randomly identify m^{r+1} units from the population of interest, where r is the number of stages.

2. Allot the m^{r+1} units randomly into m^{r-1} sets of m^2 units each.

3. For each set in step 2, apply 1–2 of the RSS procedure explained above, to get a (judgment) ranked set of size m. This step gives m^{r-1} (judgment) ranked sets, each of size m.

4. Without actual measuring of the ranked sets, apply step 3 on the m^{r-1} ranked sets to gain m^{r-2} second stage (judgment) ranked sets, each of size m.

5. Repeat step 3, without any actual measurement, until an rth stage (judgment) ranked set of size m is acquired.

6. Actually measure the m identified units in step 5.

7. Repeat steps 1–6, k times (cycles), if necessary, to obtain an rth stage ranked set sample of size mk.

Similar to our previous notation, $\{X_{ij}^{(r)} : i = 1, \ldots, m\,; j = 1, \ldots, k\}$ denotes the rth stage ranked set sample. Clearly, the especial case of MSRSS with $r = 1$ corresponds to RSS. Al-Saleh and Al-Kadiri [1] studied the case $r = 2$ which is known as double ranked set sampling. Al-Saleh and Al-Omari [2] utilized MSRSS in estimating the population mean. Al-Saleh and Samuh [3] investigated the distribution function and the median estimation based on MSRSS.

The so-called stress-strength model constitutes a useful tool for defining a reliability model. It is applicable in situations where a device, characterized by a random strength X, is subjected to a random stress Y during a certain time interval. Therefore, its reliability in the given interval may be evaluated as the probability

that X is greater than Y, i.e. $\theta = P(X > Y)$. Thanks to its potential as a general measure of the difference between two populations, θ has found applications in different fields such as economics, quality control, psychology, medicine and clinical trials. In medicine, for example, if X and Y are respectively the outcomes of experimental and control groups, then θ can be interpreted as the effectiveness of treatment. Kotz et al. [6] present the theoretical and practical results on the theory and applications of the stress-strength relationships in industrial and economic systems. In this article, we study estimating an extended concept of stress-strength reliability in MSRSS setting.

The new reliability estimator is presented in Section 2, and its properties are investigated in Section 3. Performance of the estimator is assessed by using Monte Carlo simulation whose results are reported in Section 4. A sport data set is analyzed in Section 5. Final conclusions are given in Section 6.

2. RELIABILITY ESTIMATION

Stress-strength models were introduced by Birnbaum [4] who proposed a non-parametric estimator of θ based on the Mann–Whitney statistic for the case where X and Y are independent. The density, distribution and survival functions of X are denoted by f, F and $\bar{F}$, respectively. The notation g, G and $\bar{G}$ will be used for similar functions associated with Y.

Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be independent random samples from f and g, respectively. The estimator of θ is given by

$$\hat{\theta} = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} I(X_i > Y_j),$$

where $I(\cdot)$ is the indicator function. Sengupta and Mukhuti [10] developed an unbiased estimator of θ in RSS. They showed that the proposed estimator is more efficient than $\hat{\theta}$, even in the presence of ranking errors.

There are several methods to compare lifetimes of two components in reliability theory. Comparing the survival functions, the failure rates and the mean residual lifetime functions are among the most popular approaches. Let the random variables X and Y be the lifetimes of two systems. Assume that both systems are operating at time $t > 0$. Then the residual lifetimes of them are $X_t = (X - t | X > t)$ and $Y_t = (Y - t | Y > t)$, respectively. Incorporating the age of systems, Zardasht and Asadi [14] introduced a time-dependent criterion to compare the two residual lifetimes. They considered the function $\theta(t) = P(X_t > Y_t)$. Note that $\theta(t)$ can be written as

$$\theta(t) = \frac{\theta_1(t)}{\theta_2(t)},$$

where $\theta_1(t) = P(X > Y > t)$ and $\theta_2(t) = P(X > t, Y > t)$. Using simple random samples $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ from f and g, an estimate of $\theta(t)$ can be

constructed as

$$\hat{\theta}(t) = \frac{\hat{\theta}_1(t)}{\hat{\theta}_2(t)},$$

where

$$\hat{\theta}_1(t) = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} I(X_i > Y_j > t)$$

and

$$\hat{\theta}_2(t) = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} I(X_i > t, Y_j > t).$$

Although $\hat{\theta}_1(t)$ and $\hat{\theta}_2(t)$ are unbiased estimators of $\theta_1(t)$ and $\theta_2(t)$, respectively, $\hat{\theta}(t)$ is only asymptotically unbiased. Mahdizadeh and Zamanzade [8] studied estimation of $\theta(t)$ from ranked set samples. This work deals with the same problem in MSRSS setup.

    Toward this end, we need two multistage ranked set samples from f and g. It is assumed that the samples are drawn using a single cycle. The results in the general setup are then easily followed. If $X_i^{(r)}$, $i = 1, \ldots, m$, and $Y_j^{(s)}$, $j = 1, \ldots, n$, are the two samples, then

$$\hat{\theta}^{r,s}(t) = \frac{\hat{\theta}_1^{r,s}(t)}{\hat{\theta}_2^{r,s}(t)}$$

is a natural estimator of $\theta(t)$, where

$$\hat{\theta}_1^{r,s}(t) = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} I(X_i^{(r)} > Y_j^{(s)} > t)$$

and

$$\hat{\theta}_2^{r,s}(t) = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} I(X_i^{(r)} > t, Y_j^{(s)} > t).$$

    Let $f_i^{(r)}$, $F_i^{(r)}$ and $\bar{F}_i^{(r)}$ be the density, distribution and survival functions of $X_i^{(r)}$, respectively. The notation $g_j^{(s)}$, $G_j^{(s)}$ and $\bar{G}_j^{(s)}$ will be used for similar functions associated with $Y_j^{(s)}$. Suppose the ith order statistic of an $(r-1)$st stage ranked set sample of size m from f, say $Z_1^{(r-1)}, \ldots, Z_m^{(r-1)}$, is denoted by $Z_{(i)}^{(r-1)}$. Under the assumption of no error in judgment ranking, we have $X_i^{(r)} \stackrel{d}{=} Z_{(i)}^{(r-1)}$.

    The two identities

(2.1) $$\frac{1}{m} \sum_{i=1}^{m} f_i^{(r)}(x) = f(x), \quad \frac{1}{n} \sum_{j=1}^{n} g_j^{(s)}(y) = g(y),$$

observed by Al-Saleh and Al-Omari [2], play a central role in the next section. These identities can be stated in terms of distribution functions as well.

3. THEORETICAL PROPERTIES

We begin with deriving the expectations and variances of $\hat{\theta}_1^{r,s}(t)$ and $\hat{\theta}_2^{r,s}(t)$, and comparing them with similar expressions in SRS.

PROPOSITION 3.1. *The estimators $\hat{\theta}_1^{r,s}(t)$ and $\hat{\theta}_2^{r,s}(t)$ are unbiased.*

P r o o f. It is verified by employing equations in (2.1). ∎

The following two results provide the variance expressions for components of $\hat{\theta}(t)$ and $\hat{\theta}^{r,s}(t)$.

PROPOSITION 3.2. *The variances of $\hat{\theta}_1(t)$ and $\hat{\theta}_1^{r,s}(t)$ are given by*

$$(3.1) \quad m^2 n^2 \operatorname{Var}\left(\hat{\theta}_1(t)\right)$$
$$= m(m-1)n(n-1)\theta_1^2(t) + nm(m-1)E\{\bar{F}(Y)I(Y > t)\}^2$$
$$+ mn(n-1)E\{G(X) - G(t)\}^2 + mn\theta_1(t) - m^2 n^2 \theta_1^2(t)$$

and

$$(3.2) \quad m^2 n^2 \operatorname{Var}\left(\hat{\theta}_1^{r,s}(t)\right)$$
$$= E\Big\{m^2\Big[\sum_{j=1}^{n} \bar{F}(Y_j^{(s)})I(Y_j^{(s)} > t)\Big]^2 - \sum_{i=1}^{m}\Big[\sum_{j=1}^{n} \bar{F}_i^{(r)}(Y_j^{(s)})I(Y_j^{(s)} > t)\Big]^2\Big\}$$
$$+ mE\Big\{n^2[G(X) - G(t)]^2 - \sum_{j=1}^{n}[G_j^{(s)}(X) - G_j^{(s)}(t)]^2\Big\}$$
$$+ mn\theta_1(t) - m^2 n^2 \theta_1^2(t).$$

P r o o f. It is readily seen that

$$(3.3) \qquad\qquad m^2 n^2 E\left(\hat{\theta}_1(t)\right)^2 = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4,$$

where

$$(3.4) \qquad \alpha_1 = E\Big\{\sum_{i \neq i'=1}^{m} \sum_{j \neq j'=1}^{n} I(X_i > Y_j > t)I(X_{i'} > Y_{j'} > t)\Big\}$$
$$= m(m-1)n(n-1)\theta_1^2(t),$$

$$(3.5) \quad \alpha_2 = E\Big\{\sum_{j=1}^{n} \sum_{i \neq i'=1}^{m} I(X_i > Y_j > t)I(X_{i'} > Y_j > t)\Big\}$$
$$= \sum_{j=1}^{n} \sum_{i \neq i'=1}^{m} EE\big\{I(X_i > Y_j > t)I(X_{i'} > Y_j > t)\big|Y_j\big\}$$
$$= \sum_{j=1}^{n} \sum_{i \neq i'=1}^{m} E\{\bar{F}(Y)I(Y > t)\}^2 = nm(m-1)E\{\bar{F}(Y)I(Y > t)\}^2,$$

$$(3.6) \quad \alpha_3 = E\{ \sum_{i=1}^{m} \sum_{j\neq j'=1}^{n} I(X_i > Y_j > t)I(X_i > Y_{j'} > t)\}$$

$$= \sum_{i=1}^{m} \sum_{j\neq j'=1}^{n} EE\{I(X_i > Y_j > t)I(X_i > Y_{j'} > t)\big|X_i\}$$

$$= \sum_{i=1}^{m} \sum_{j\neq j'=1}^{n} E\{G(X) - G(t)\}^2 = mn(n-1)E\{G(X) - G(t)\}^2$$

and

$$(3.7) \quad \alpha_4 = E\{ \sum_{i=1}^{m} \sum_{j=1}^{n} I(X_i > Y_j > t)\} = mn\theta_1(t).$$

By (3.3)–(3.7) and unbiasedness of $\hat{\theta}_1(t)$, the proof of the first part is complete. Similarly,

$$(3.8) \quad m^2 n^2 E\big(\hat{\theta}_1^{r,s}(t)\big)^2 = \beta_1 + \beta_2 + \beta_3,$$

where

(3.9)

$$\beta_1 = E\{ \sum_{i\neq i'=1}^{m} \sum_{j\neq j'=1}^{n} I(X_i^{(r)} > Y_j^{(s)} > t)I(X_{i'}^{(r)} > Y_{j'}^{(s)} > t)$$

$$+ \sum_{j=1}^{n} \sum_{i\neq i'=1}^{m} I(X_i^{(r)} > Y_j^{(s)} > t)I(X_{i'}^{(r)} > Y_j^{(s)} > t)\}$$

$$= \sum_{i\neq i'=1}^{m} \sum_{j\neq j'=1}^{n} EE\{I(X_i^{(r)} > Y_j^{(s)} > t)\big|Y_j^{(s)}\}EE\{I(X_{i'}^{(r)} > Y_{j'}^{(s)} > t)\big|Y_{j'}^{(s)}\}$$

$$+ \sum_{j=1}^{n} \sum_{i\neq i'=1}^{m} EE\{I(X_i^{(r)} > Y_j^{(s)} > t)I(X_{i'}^{(r)} > Y_j^{(s)} > t)\big|Y_j^{(s)}\}$$

$$= E\{ \sum_{i\neq i'=1}^{m} \sum_{j\neq j'=1}^{n} [\bar{F}_i^{(r)}(Y_j^{(s)})I(Y_j^{(s)} > t)][\bar{F}_{i'}^{(r)}(Y_{j'}^{(s)})I(Y_{j'}^{(s)} > t)]$$

$$+ \sum_{j=1}^{n} \sum_{i\neq i'=1}^{m} [\bar{F}_i^{(r)}(Y_j^{(s)})I(Y_j^{(s)} > t)][\bar{F}_{i'}^{(r)}(Y_j^{(s)})I(Y_j^{(s)} > t)]\}$$

$$= E\{ \big[\sum_{i=1}^{m} \sum_{j=1}^{n} \bar{F}_i^{(r)}(Y_j^{(s)})I(Y_j^{(s)} > t)\big]^2 - \sum_{i=1}^{m} \sum_{j=1}^{n} [\bar{F}_i^{(r)}(Y_j^{(s)})I(Y_j^{(s)} > t)]^2$$

$$- \sum_{i=1}^{m} \sum_{j\neq j'=1}^{n} [\bar{F}_i^{(r)}(Y_j^{(s)})I(Y_j^{(s)} > t)][\bar{F}_i^{(r)}(Y_{j'}^{(s)})I(Y_{j'}^{(s)} > t)]\}$$

$$= E\{m^2 \big[\sum_{j=1}^{n} \bar{F}(Y_j^{(s)})I(Y_j^{(s)} > t)\big]^2 - \sum_{i=1}^{m} \big[\sum_{j=1}^{n} \bar{F}_i^{(r)}(Y_j^{(s)})I(Y_j^{(s)} > t)\big]^2\},$$

(3.10) $\quad \beta_2 = E\{ \sum\limits_{i=1}^{m} \sum\limits_{j \neq j'=1}^{n} I(X_i^{(r)} > Y_j^{(s)} > t)I(X_i^{(r)} > Y_{j'}^{(s)} > t)\}$

$$= m \sum\limits_{j \neq j'=1}^{n} E\{I(X > Y_j^{(s)} > t)I(X > Y_{j'}^{(s)} > t)\}$$

$$= m \sum\limits_{j \neq j'=1}^{n} EE\{I(X > Y_j^{(s)} > t)I(X > Y_{j'}^{(s)} > t)\big|X\}$$

$$= m \sum\limits_{j \neq j'=1}^{n} E\{[G_j^{(s)}(X) - G_j^{(s)}(t)][G_{j'}^{(s)}(X) - G_{j'}^{(s)}(t)]\}$$

$$= mE\{n^2[G(X) - G(t)]^2 - \sum\limits_{j=1}^{n}[G_j^{(s)}(X) - G_j^{(s)}(t)]^2\}$$

and

(3.11) $\quad \beta_3 = E\{ \sum\limits_{i=1}^{m} \sum\limits_{j=1}^{n} I(X_i^{(r)} > Y_j^{(s)} > t)\} = mn\theta_1(t).$

Now, the second part follows from (3.8)–(3.11) and unbiasedness of $\hat{\theta}_1^{r,s}(t)$. ∎

PROPOSITION 3.3. *The variances of $\hat{\theta}_2(t)$ and $\hat{\theta}_2^{r,s}(t)$ are given by*

(3.12) $\quad m^2 n^2 \operatorname{Var}\big(\hat{\theta}_2(t)\big)$
$$= [m(m-1)\bar{F}^2(t)][n(n-1)\bar{G}^2(t)] + m\bar{F}(t)[n(n-1)\bar{G}^2(t)]$$
$$+ n\bar{G}(t)[m(m-1)\bar{F}^2(t)] + mn\theta_2(t) - m^2 n^2 \theta_2^2(t)$$

and

(3.13) $\quad m^2 n^2 \operatorname{Var}\big(\hat{\theta}_2^{r,s}(t)\big)$
$$= \big[m^2 \bar{F}^2(t) - \sum\limits_{i=1}^{m}\big(\bar{F}_i^{(r)}(t)\big)^2\big]\big[n^2 \bar{G}^2(t) - \sum\limits_{j=1}^{n}\big(\bar{G}_j^{(s)}(t)\big)^2\big]$$

$$+ m\bar{F}(t)\big[n^2 \bar{G}^2(t) - \sum\limits_{j=1}^{n}\big(\bar{G}_j^{(s)}(t)\big)^2\big]$$

$$+ n\bar{G}(t)\big[m^2 \bar{F}^2(t) - \sum\limits_{i=1}^{m}\big(\bar{F}_i^{(r)}(t)\big)^2\big] + mn\theta_2(t) - m^2 n^2 \theta_2^2(t).$$

P r o o f. It is similar to the proof of the previous result. ∎

It is possible to find approximations for variances of $\hat{\theta}(t)$ and $\hat{\theta}^{r,s}(t)$ using the first-order Taylor series expansion. These expressions, however, are not helpful in establishing superiority of the RSS-based estimator over the SRS estimator. In the

sequel, we show that the numerator and denominator of $\hat{\theta}^{r,s}(t)$ are more efficient than similar components in $\hat{\theta}(t)$.

The variances of $\hat{\theta}_1(t)$ and $\hat{\theta}_1^{r,s}(t)$ are compared in the next proposition.

PROPOSITION 3.4. *For any $m, n \geqslant 2$ and $r, s \geqslant 1$,*

$$\mathrm{Var}\left(\hat{\theta}_1^{r,s}(t)\right) \leqslant \mathrm{Var}\left(\hat{\theta}_1(t)\right).$$

P r o o f. Using equations (3.1) and (3.2), it can be seen that

$$m^2 n^2 \left[\mathrm{Var}\left(\hat{\theta}_1(t)\right) - \mathrm{Var}\left(\hat{\theta}_1^{r,s}(t)\right)\right] = \zeta_1 + \zeta_2 + \zeta_3,$$

where

$$\zeta_1 = E\Big\{\sum_{i=1}^m \big[\sum_{j=1}^n \bar{F}_i^{(r)}(Y_j^{(s)})I(Y_j^{(s)} > t)\big]^2 - m\big[\sum_{j=1}^n \bar{F}(Y_j^{(s)})I(Y_j^{(s)} > t)\big]^2\Big\}$$

$$= E\Big\{\sum_{i=1}^m \big(\sum_{j=1}^n [\bar{F}_i^{(r)}(Y_j^{(s)})I(Y_j^{(s)} > t) - \bar{F}(Y_j^{(s)})I(Y_j^{(s)} > t)]\big)^2\Big\},$$

$$\zeta_2 = mn(n-1)E\{G(X) - G(t)\}^2$$

$$\quad - mE\Big\{n^2[G(X) - G(t)]^2 - \sum_{j=1}^n [G_j^{(s)}(X) - G_j^{(s)}(t)]^2\Big\}$$

$$= mE\Big\{\sum_{j=1}^n [G_j^{(s)}(X) - G_j^{(s)}(t)]^2 - n[G(X) - G(t)]^2\Big\}$$

$$= mE\Big\{\sum_{j=1}^n \big([G_j^{(s)}(X) - G_j^{(s)}(t)] - [G(X) - G(t)]\big)^2\Big\},$$

$$\zeta_3 = m(m-1)n(n-1)\theta_1^2(t) + nm(m-1)E\{\bar{F}(Y)I(Y > t)\}^2$$

$$\quad - m(m-1)E\Big\{\big[\sum_{j=1}^n \bar{F}(Y_j^{(s)})I(Y_j^{(s)} > t)\big]^2\Big\}$$

$$= m(m-1)\Big[\Big(1 - \frac{1}{n}\Big)\big(\sum_{j=1}^n E\{\bar{F}(Y_j^{(s)})I(Y_j^{(s)} > t)\}\big)^2$$

$$\quad - \sum_{j \neq j'=1}^n E\{\bar{F}(Y_j^{(s)})I(Y_j^{(s)} > t)\}E\{\bar{F}(Y_{j'}^{(s)})I(Y_{j'}^{(s)} > t)\}\Big]$$

$$= m(m-1)\Big[\sum_{j=1}^n E^2\{\bar{F}(Y_j^{(s)})I(Y_j^{(s)} > t)\}$$

$$\quad - \frac{1}{n}\big(\sum_{j=1}^n E\{\bar{F}(Y_j^{(s)})I(Y_j^{(s)} > t)\}\big)^2\Big]$$

$$= m(m-1)\sum_{j=1}^n E^2\{\bar{F}(Y_j^{(s)})I(Y_j^{(s)} > t) - \bar{F}(Y)I(Y > t)\}.$$

Since ζ_i's are non-negative, we have shown the result. ∎

The next proposition compares the variances of $\hat{\theta}_2(t)$ and $\hat{\theta}_2^{r,s}(t)$.

PROPOSITION 3.5. *For any $m, n \geqslant 2$ and $r, s \geqslant 1$,*

$$\mathrm{Var}\left(\hat{\theta}_2^{r,s}(t)\right) \leqslant \mathrm{Var}\left(\hat{\theta}_2(t)\right).$$

P r o o f. Using equations (3.12) and (3.13), it is easy to see that

$$m^2 n^2 \left[\,\mathrm{Var}\left(\hat{\theta}_2(t)\right) - \mathrm{Var}\left(\hat{\theta}_2^{r,s}(t)\right)\right] = \eta_1 + \eta_2 + \eta_3,$$

where

$$\eta_1 = m\bar{F}(t)\Big[\sum_{j=1}^n \big(\bar{G}_j^{(s)}(t)\big)^2 - n\bar{G}^2(t)\Big] + n\bar{G}(t)\Big[\sum_{i=1}^m \big(\bar{F}_i^{(r)}(t)\big)^2 - m\bar{F}^2(t)\Big],$$

$$\eta_2 = m^2\bar{F}^2(t)\Big[\sum_{j=1}^n \big(\bar{G}_j^{(s)}(t)\big)^2 - n\bar{G}^2(t)\Big] + n^2\bar{G}^2(t)\Big[\sum_{i=1}^m \big(\bar{F}_i^{(r)}(t)\big)^2 - m\bar{F}^2(t)\Big]$$

and

$$\eta_3 = mn\bar{F}^2(t)\bar{G}^2(t) - \Big[\sum_{i=1}^m \big(\bar{F}_i^{(r)}(t)\big)^2\Big]\Big[\sum_{j=1}^n \big(\bar{G}_j^{(s)}(t)\big)^2\Big].$$

Let $a_i = \bar{F}_i^{(r)}(t)$ $(i = 1, \ldots, m)$ and $b_j = \bar{G}_j^{(s)}(t)$ $(j = 1, \ldots, n)$ with the corresponding moments $\mu_{k,a} = \sum_{i=1}^m a_i^k/m$ and $\mu_{k,b} = \sum_{j=1}^n b_j^k/n$ for $k = 1, 2$. Also, assume that $\sigma_a^2 = \mu_{2,a} - \mu_{1,a}^2$ and $\sigma_b^2 = \mu_{2,b} - \mu_{1,b}^2$. Then, we have

$$\eta_1 = mn[\mu_{1,a}\sigma_b^2 + \mu_{1,b}\sigma_a^2],$$

$$\eta_2 = mn[m\mu_{1,a}^2\sigma_b^2 + n\mu_{1,b}^2\sigma_a^2]$$

and

$$\eta_3 = -mn[\mu_{1,a}^2\sigma_b^2 + \mu_{1,b}^2\sigma_a^2 + \sigma_a^2\sigma_b^2].$$

The proof is complete due to the inequality

$$\frac{\eta_1 + \eta_2 + \eta_3}{mn}$$
$$= \mu_{1,a}\sigma_b^2 + \mu_{1,b}\sigma_a^2 + m\mu_{1,a}^2\sigma_b^2 + n\mu_{1,b}^2\sigma_a^2 - \mu_{1,a}^2\sigma_b^2 - \mu_{1,b}^2\sigma_a^2 - \sigma_a^2\sigma_b^2$$
$$\geqslant \mu_{2,a}\sigma_b^2 + \mu_{2,b}\sigma_a^2 + m\mu_{1,a}^2\sigma_b^2 + n\mu_{1,b}^2\sigma_a^2 - \mu_{1,a}^2\sigma_b^2 - \mu_{1,b}^2\sigma_a^2 - \sigma_a^2\sigma_b^2$$
$$= (\mu_{2,a} - \mu_{1,a}^2)\sigma_b^2 + (\mu_{2,b} - \mu_{1,b}^2)\sigma_a^2 + m\mu_{1,a}^2\sigma_b^2 + n\mu_{1,b}^2\sigma_a^2 - \sigma_a^2\sigma_b^2$$
$$= m\mu_{1,a}^2\sigma_b^2 + n\mu_{1,b}^2\sigma_a^2 + \sigma_a^2\sigma_b^2 \geqslant 0,$$

where the inequality follows from the fact that $\mu_{1,a} \geqslant \mu_{2,a}$ and $\mu_{1,b} \geqslant \mu_{2,b}$. ∎

According to the RSS literature, more efficient estimators of the population mean, distribution function and median are obtained by increasing the number of stages in MSRSS, given a fixed set size. The next two results verify a similar property in estimating $\theta_1(t)$ and $\theta_2(t)$.

PROPOSITION 3.6. *For fixed m and n, $\mathrm{Var}(\hat{\theta}_1^{r,s})$ is decreasing in r and s.*

P r o o f. It suffices to show that $\mathrm{Var}(\hat{\theta}_1^{r,s}) \leqslant \mathrm{Var}(\hat{\theta}_1^{r-1,s})$ and $\mathrm{Var}(\hat{\theta}_1^{r,s}) \leqslant \mathrm{Var}(\hat{\theta}_1^{r,s-1})$. Using the reasoning from the beginning of the proof for the second part of Proposition 3.2, one can write

(3.14)

$$
m^2 n^2 E(\hat{\theta}_1^{r,s})^2 = E\Big\{ \sum_{i \neq i'=1}^{m} \sum_{j \neq j'=1}^{n} I(X_i^{(r)} > Y_j^{(s)} > t)I(X_{i'}^{(r)} > Y_{j'}^{(s)} > t)
$$

$$
+ \sum_{i=1}^{m} \sum_{j \neq j'=1}^{n} I(X_i^{(r)} > Y_j^{(s)} > t)I(X_i^{(r)} > Y_{j'}^{(s)} > t)
$$

$$
+ \sum_{j=1}^{n} \sum_{i \neq i'=1}^{m} I(X_i^{(r)} > Y_j^{(s)} > t)I(X_{i'}^{(r)} > Y_j^{(s)} > t)
$$

$$
+ \sum_{i=1}^{m} \sum_{j=1}^{n} I(X_i^{(r)} > Y_j^{(s)} > t)\Big\}.
$$

To prove the result, we need some equalities and inequalities regarding the four expectation terms on the right-hand side of equation (3.14). Suppose $W_{(i)}^{(r-1)}$ is the ith order statistic of an $(r-1)$st stage ranked set sample of size m from f. As to the first term, we have

$$
\begin{aligned}
(3.15) \quad & E\{I(X_i^{(r)} > Y_j^{(s)} > t)I(X_{i'}^{(r)} > Y_{j'}^{(s)} > t)\} \\
&= EE\{I(X_i^{(r)} > Y_j^{(s)} > t)I(X_{i'}^{(r)} > Y_{j'}^{(s)} > t)\big|Y_j^{(s)}, Y_{j'}^{(s)}\} \\
&= E\big[E\{I(X_i^{(r)} > Y_j^{(s)} > t)\big|Y_j^{(s)}, Y_{j'}^{(s)}\}E\{I(X_{i'}^{(r)} > Y_{j'}^{(s)} > t)\big|Y_j^{(s)}, Y_{j'}^{(s)}\}\big] \\
&= E\big[E\{I(W_{(i)}^{(r-1)} > Y_j^{(s)} > t)\big|Y_j^{(s)}, Y_{j'}^{(s)}\} \\
&\qquad \times E\{I(W_{(i')}^{(r-1)} > Y_{j'}^{(s)} > t)\big|Y_j^{(s)}, Y_{j'}^{(s)}\}\big] \\
&\leqslant EE\{I(W_{(i)}^{(r-1)} > Y_j^{(s)} > t)I(W_{(i')}^{(r-1)} > Y_{j'}^{(s)} > t)\big|Y_j^{(s)}, Y_{j'}^{(s)}\} \\
&= E\{I(W_{(i)}^{(r-1)} > Y_j^{(s)} > t)I(W_{(i')}^{(r-1)} > Y_{j'}^{(s)} > t)\},
\end{aligned}
$$

where the inequality holds owing to the positive covariance between any pair of order statistics in a sample (see Lehmann [7]).

Similarly, it follows that

$$(3.16) \quad E\{I(X_i^{(r)} > Y_j^{(s)} > t)I(X_{i'}^{(r)} > Y_j^{(s)} > t)\}$$
$$= EE\{I(X_i^{(r)} > Y_j^{(s)} > t)I(X_{i'}^{(r)} > Y_j^{(s)} > t)|Y_j^{(s)}\}$$
$$= E\big[E\{I(X_i^{(r)} > Y_j^{(s)} > t)|Y_j^{(s)}\}E\{I(X_{i'}^{(r)} > Y_j^{(s)} > t)|Y_j^{(s)}\}\big]$$
$$= E\big[E\{I(W_{(i)}^{(r-1)} > Y_j^{(s)} > t)|Y_j^{(s)}\}E\{I(W_{(i')}^{(r-1)} > Y_j^{(s)} > t)|Y_j^{(s)}\}\big]$$
$$\leqslant EE\{I(W_{(i)}^{(r-1)} > Y_j^{(s)} > t)I(W_{(i')}^{(r-1)} > Y_j^{(s)} > t)|Y_j^{(s)}\}$$
$$= E\{I(W_{(i)}^{(r-1)} > Y_j^{(s)} > t)I(W_{(i')}^{(r-1)} > Y_j^{(s)} > t)\}.$$

Additionally,

$$(3.17) \quad E\{I(X_i^{(r)} > Y_j^{(s)} > t)I(X_i^{(r)} > Y_{j'}^{(s)} > t)\}$$
$$= EE\big\{I(X_i^{(r)} > Y_j^{(s)} > t)I(X_i^{(r)} > Y_{j'}^{(s)} > t)|Y_j^{(s)}, Y_{j'}^{(s)}\big\}$$
$$= EE\big\{I(W_{(i)}^{(r-1)} > Y_j^{(s)} > t)I(W_{(i)}^{(r-1)} > Y_{j'}^{(s)} > t)|Y_j^{(s)}, Y_{j'}^{(s)}\big\}$$
$$= E\{I(W_{(i)}^{(r-1)} > Y_j^{(s)} > t)I(W_{(i)}^{(r-1)} > Y_{j'}^{(s)} > t)\}$$

and

$$(3.18) \qquad E\{I(X_i^{(r)} > Y_j^{(s)} > t)\} = EE\big\{I(X_i^{(r)} > Y_j^{(s)} > t)|Y_j^{(s)}\big\}$$
$$= EE\big\{I(W_{(i)}^{(r-1)} > Y_j^{(s)} > t)|Y_j^{(s)}\big\}$$
$$= E\{I(W_{(i)}^{(r-1)} > Y_j^{(s)} > t)\}.$$

From (3.14)–(3.18) we get

$$m^2 n^2 E(\hat{\theta}_1^{r,s})^2 \leqslant E\Big\{ \sum_{i \neq i'=1}^{m} \sum_{j \neq j'=1}^{n} I(W_{(i)}^{(r-1)} > Y_j^{(s)} > t)I(W_{(i')}^{(r-1)} > Y_{j'}^{(s)} > t)$$
$$+ \sum_{i=1}^{m} \sum_{j \neq j'=1}^{n} I(W_{(i)}^{(r-1)} > Y_j^{(s)} > t)I(W_{(i)}^{(r-1)} > Y_{j'}^{(s)} > t)$$
$$+ \sum_{j=1}^{n} \sum_{i \neq i'=1}^{m} I(W_{(i)}^{(r-1)} > Y_j^{(s)} > t)I(W_{(i')}^{(r-1)} > Y_j^{(s)} > t)$$
$$+ \sum_{i=1}^{m} \sum_{j=1}^{n} I(W_{(i)}^{(r-1)} > Y_j^{(s)} > t)\Big\} = m^2 n^2 E(\hat{\theta}_1^{r-1,s})^2.$$

This and unbiasedness of $\hat{\theta}_1^{r,s}$ imply that $\mathrm{Var}(\hat{\theta}_1^{r,s}) \leqslant \mathrm{Var}(\hat{\theta}_1^{r-1,s})$. A similar argument proves the second part. ∎

We close this section by the analog of Proposition 3.6 for $\hat{\theta}_2^{r,s}$. The proof is done similarly, and thus it is not detailed.

PROPOSITION 3.7. *For fixed m and n, $\mathrm{Var}(\hat{\theta}_2^{r,s})$ is decreasing in r and s.*

4. NUMERICAL STUDIES

We carried out a simulation study to examine the performance of the dynamic reliability estimator. To this end, it was assumed that the distribution functions of X and Y are related as $G(y) = F(y/\sigma)$ for some $\sigma > 0$. In particular, we considered the following choices of F:

- The standard uniform distribution with

$$F(x) = x, \quad 0 < x < 1,$$

which gives

$$\theta(t) = \begin{cases} (\sigma - t)/\big(2(1 - t)\big) & \text{if } 0 < \sigma < 1, \\ (1 - t)/\big(2(\sigma - t)\big) & \text{if } \sigma \geqslant 1, \end{cases} \quad 0 < t < \min\{1, \sigma\}.$$
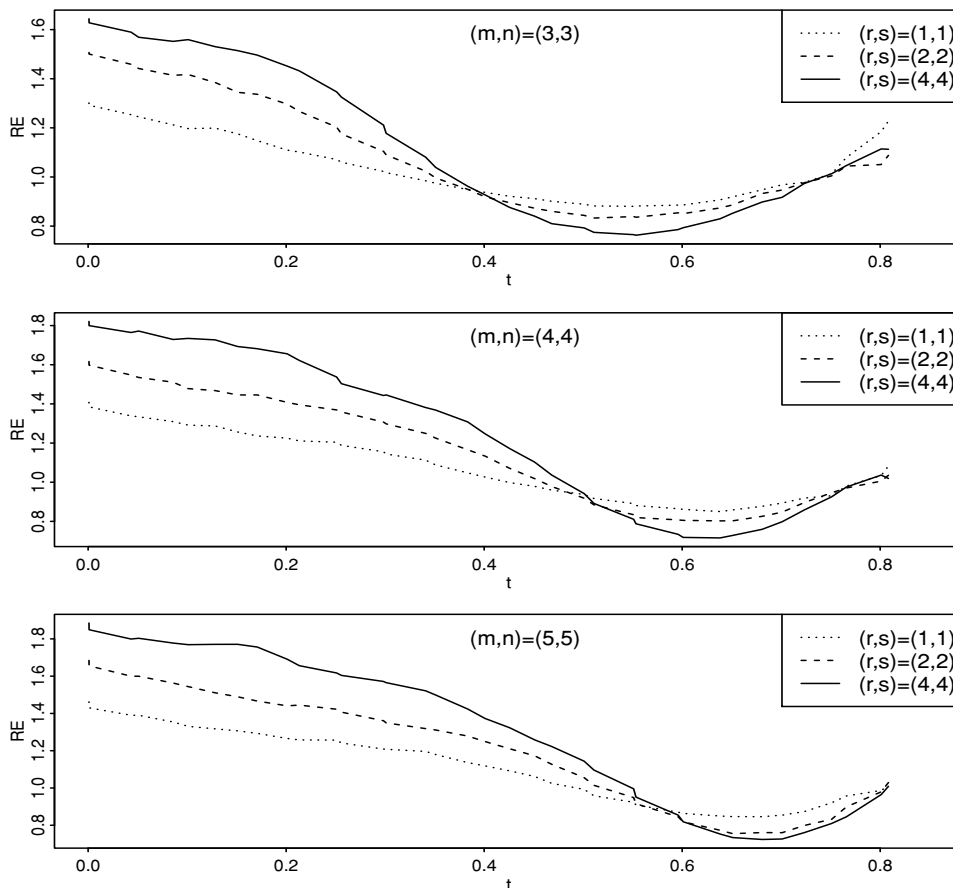

FIGURE 1. Estimated REs for $F(x) = x, 0 < x < 1$, and $\sigma = 0.85$.

- The standard exponential distribution with

$$F(x) = 1 - e^{-x}, \quad x > 0,$$

which gives

$$\theta(t) = \frac{1}{\sigma + 1}, \quad t > 0.$$

- The standard normal distribution with

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}\, \mathrm{d}z = \Phi(x), \quad x \in \mathbb{R},$$

which gives

$$\theta(t) = \left[\int_{t}^{\infty} \Phi\left(\frac{x}{\sigma}\right) \frac{1}{\sqrt{2\pi}\sigma}\, e^{-x^2/2}\, \mathrm{d}x - \Phi\left(\frac{t}{\sigma}\right)\left(1 - \Phi(t)\right) \right]$$
$$\times \left[\left(1 - \Phi\left(\frac{t}{\sigma}\right)\right)\left(1 - \Phi(t)\right) \right]^{-1}, \quad t \in \mathbb{R}.$$
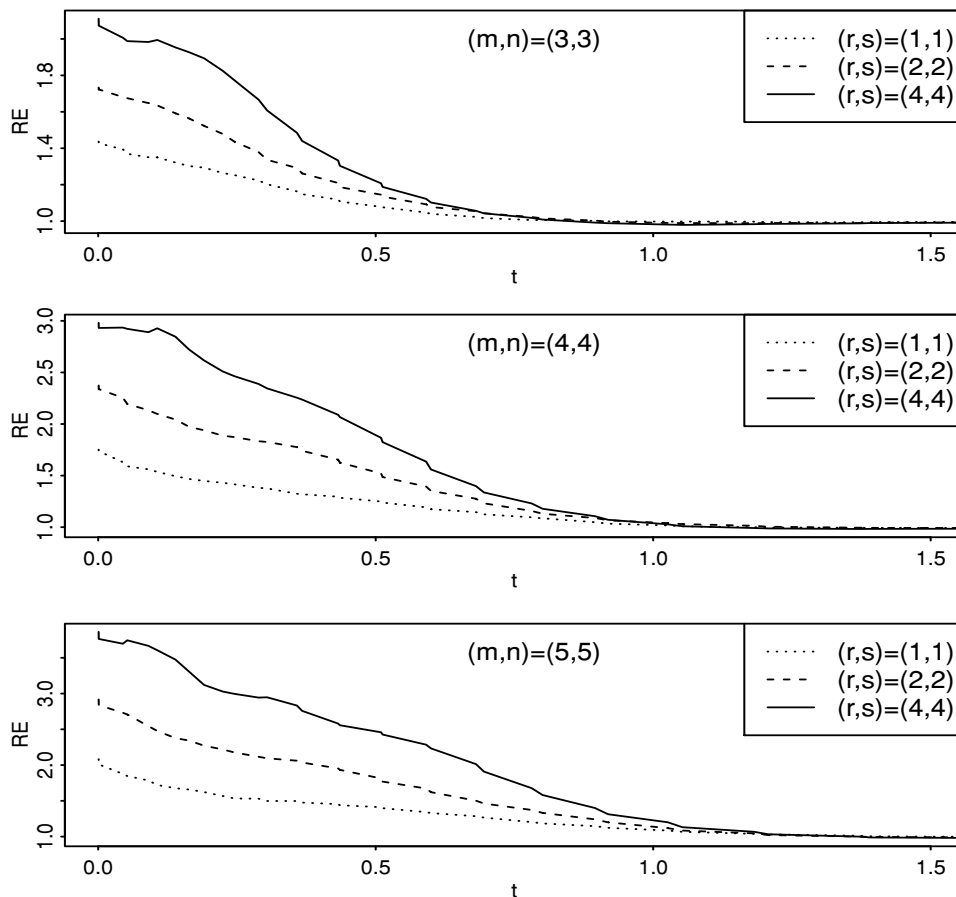

FIGURE 2. Estimated REs for $F(x) = 1 - e^{-x}$, $x > 0$, and $\sigma = 0.85$.

- The standard logistic distribution with

$$F(x) = \frac{1}{1 + e^{-x}}, \quad x \in \mathbb{R},$$

which gives

$$\theta(t) = \left[\int_t^\infty \frac{1}{1 + e^{-x/\sigma}} \frac{e^{-x}}{(1 + e^{-x})^2} \, dx - \frac{1}{1 + e^{-t/\sigma}} \left(1 - \frac{1}{1 + e^{-t}} \right) \right]$$

$$\times \left[\left(1 - \frac{1}{1 + e^{-t/\sigma}} \right) \left(1 - \frac{1}{1 + e^{-t}} \right) \right]^{-1}, \quad t \in \mathbb{R}.$$

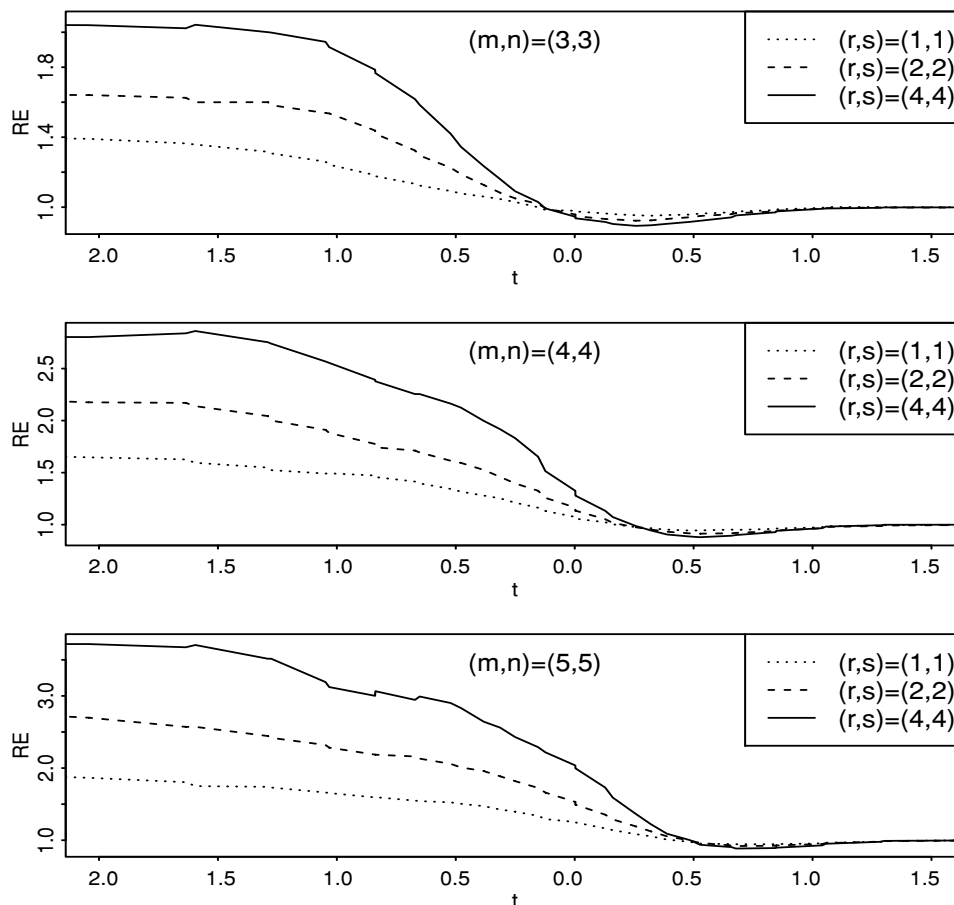Also, sample sizes

$$(m, n) \in \{(3, 3), (4, 4), (5, 5)\}$$

FIGURE 3. Estimated REs for $F(x) = \Phi(x), x \in \mathbb{R}$, and $\sigma = 1.25$.

and stage numbers

$$(r, s) \in \{(1, 1), (2, 2), (4, 4)\}$$

were selected.

The efficiency of $\hat{\theta}^{r,s}(t)$ relative to $\hat{\theta}(t)$ was estimated as follows. For each combination of distributions and sample sizes, 10,000 pairs of samples were generated in SRS and MSRSS (with the aforesaid stage numbers). The two estimators were computed from each pair of samples in the corresponding designs, and their mean squared errors (MSEs) were determined. The relative efficiency (RE) is defined as

$$\mathrm{RE}(t) = \frac{\widehat{\mathrm{MSE}}\big(\hat{\theta}(t)\big)}{\widehat{\mathrm{MSE}}\big(\hat{\theta}^{r,s}(t)\big)}.$$
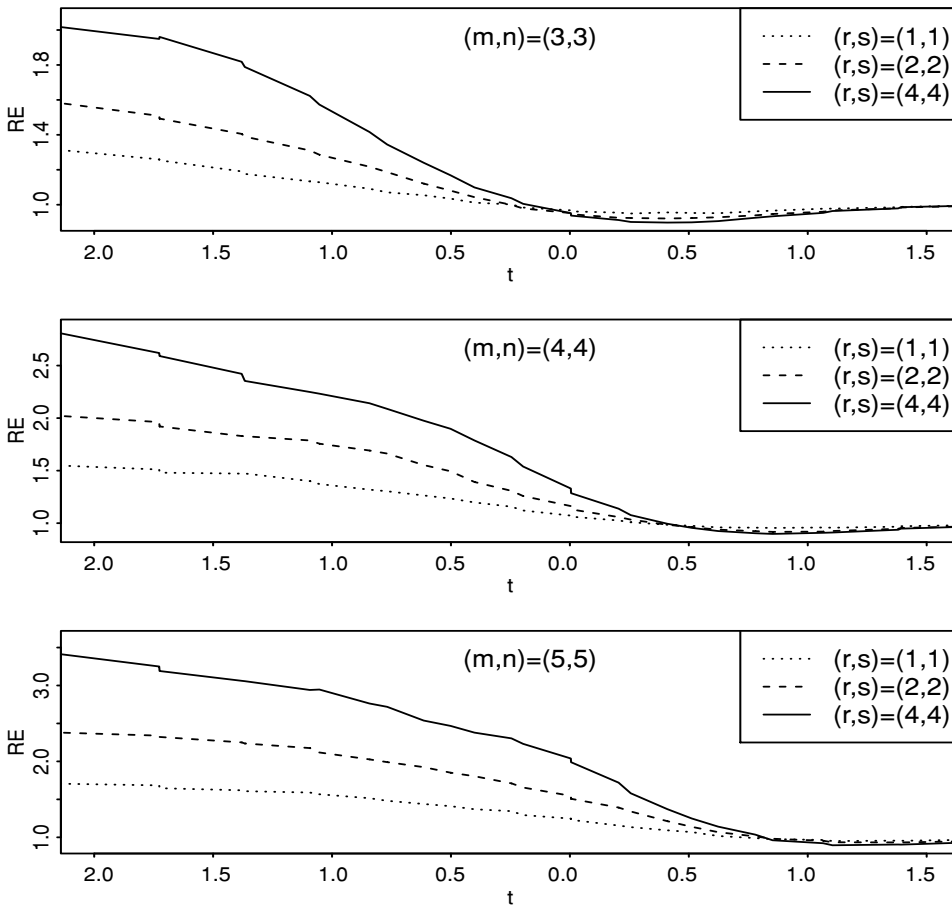

FIGURE 4. Estimated REs for $F(x) = 1/(1 + e^{-x})$, $x \in \mathbb{R}$, and $\sigma = 1.25$.

The RE(t) values larger than one indicate that $\hat{\theta}^{r,s}(t)$ is more efficient than $\hat{\theta}(t)$. Figures 1–4 display some of the results for $\sigma = 0.85, 1.25$. Full output figures are available on request from the first author.

It is observed that the SRS estimator is dominated by its MSRSS analog. For fixed sample sizes, the larger stage numbers, the larger REs. These trends are sometimes violated for a limited range of t values. For example, see Figure 1 in this respect. It is worth noting that the efficiency gain is relatively high even with small sample size $(m, n) = (3, 3)$. In general, the RE tends to be decreasing as a function of t, as seen in most figures.

5. ILLUSTRATION

The MSRSS can be very efficient if the variable of interest is highly correlated with a concomitant variable. In this case, if the second variable can be measured with negligible cost, then we may use it in judgment ranking process (see Stokes [11] for more details). In doing so, in step 2 of the RSS procedure, the elements of the ith sample are ordered according to the concomitant variable, and then study variable is actually measured for unit ranked ith smallest. The MSRSS with concomitant variable is implemented similarly.

In this section, we illustrate the proposed procedure using a data set collected at the Australian Institute of Sport. It is made up of thirteen measured variables on 102 male and 100 female athletes.[1] We will consider lean body mass (LBM) and body mass index (BMI) for each athlete. The LBM is a component of body composition, calculated by subtracting body fat weight from total body weight. Exact measurement of the LBM is done using various technologies such as dual energy X-ray absorptiometry (DEXA) which is costly. On the other hand, the BMI is a well-accepted measure of obesity which is easy to calculate and readily accessible. A BMI value is simply weight (in kg) divided by square of height (in m). The correlation coefficient between the two variables is 0.71. So, the BMI can serve as a concomitant variable.

Let X and Y be the LBM variables for the male and female populations, respectively. It is of interest to estimate $\theta(t) = P(X_t > Y_t)$, as defined in Section 2. The threshold t may be interpreted as a lower bound on the LBM values, which is easily available from previous studies or experts' opinions. For $(m, n) = (4, 4)$, 100,000 samples were drawn from the two hypothetical populations based on SRS and MSRSS designs, where $(r, s) \in \{(1, 1), (2, 2), (3, 3)\}$. The sampling is done with replacement to ensure that the measured units are independent of each other. From each sample, the appropriate estimator ($\hat{\theta}(t)$ or $\hat{\theta}^{r,s}(t)$) was computed, and its MSE was determined. Finally, RE(t) was evaluated as stated in the previous section. The results are depicted in Figure 5. As expected, the REs always exceed the

---

[1] The data set can be found at http://www.statsci.org/data/oz/ais.html

unity, and higher precision is achieved by increasing the number of stages. These are consistent with our theoretical results and simulation studies.
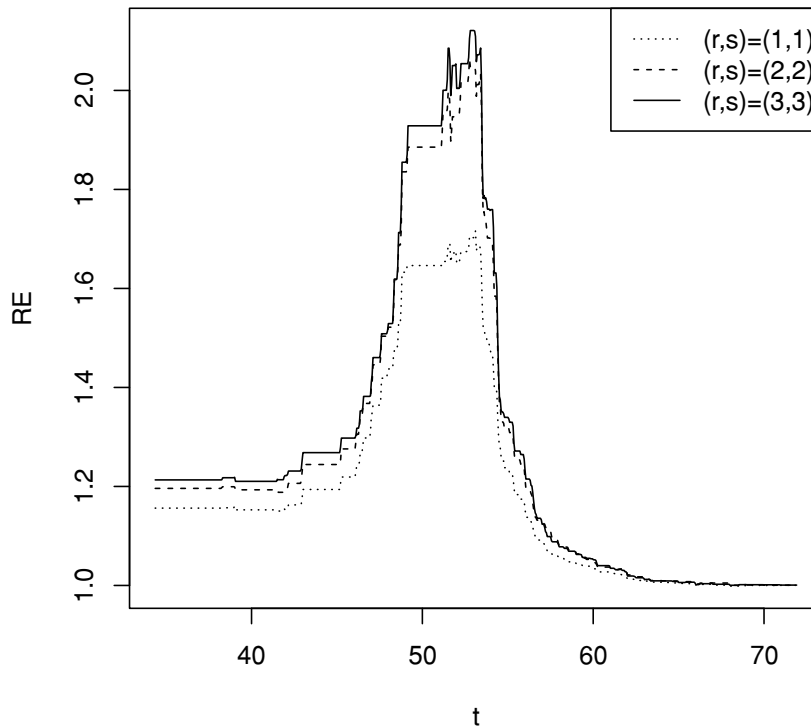

FIGURE 5. Estimated RE as a function of t for the sport data set.

6. CONCLUSION

The problem of estimating the reliability parameter θ, when X and Y are independent random variables, has received considerable attention. This index has been recently extended to a dynamic form $\theta(t)$. This article concerns estimation of the latter measure based on MSRSS. A nonparametric estimator is developed, and its optimal properties are studied. Results of simulation studies reveal that the proposed estimator often outperforms its SRS analog. Also, the estimator becomes more efficient by increasing the stage numbers. A real data set is utilized to illustrate application of the proposed index.

Acknowledgments. The authors are thankful to the referees for careful reading of the paper and insightful comments.

REFERENCES

[1] M. F. Al-Saleh and M. A. Al-Kadiri, *Double-ranked set sampling*, Statist. Probab. Lett. 48 (2) (2000), pp. 205–212.

[2] M. F. Al-Saleh and A. I. Al-Omari, *Multistage ranked set sampling*, J. Statist. Plann. Inference 102 (2) (2002), pp. 273–286.

[3] M. F. Al-Saleh and M. H. Samuh, *On multistage ranked set sampling for distribution and median estimation*, Comput. Statist. Data Anal. 52 (4) (2008), pp. 2066–2078.

[4] Z. M. Birnbaum, *On a use of the Mann–Whitney statistic*, in: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1: *Contributions to the Theory of Statistics*, J. Neyman (Ed.), University of California Press, Berkeley 1956, pp. 13–17.

[5] Z. Chen, *Ranked set sampling: Its essence and some new applications*, Environ. Ecol. Stat. 14 (4) (2007), pp. 355–363.

[6] S. Kotz, Y. Lumelskii, and M. Pensky, *The Stress-Strength Model and Its Generalizations: Theory and Applications*, World Scientific, River Edge, NJ, 2003.

[7] E. L. Lehmann, *Some concepts of dependence*, Ann. Math. Statist. 37 (1966), pp. 1137–1153.

[8] M. Mahdizadeh and E. Zamanzade, *A new reliability measure in ranked set sampling*, Statist. Papers 59 (3) (2018), pp. 861–891.

[9] G. A. McIntyre, *A method for unbiased selective sampling using ranked sets*, Aust. J. Agric. Res. 3 (1952), pp. 385–390.

[10] S. Sengupta and S. Mukhuti, *Unbiased estimation of $P(X > Y)$ using ranked set sample data*, Statistics 42 (3) (2008), pp. 223–230.

[11] S. L. Stokes, *Ranked set sampling with concomitant variables*, Comm. Statist. Theory Methods 6 (1977), pp. 1207–1211.

[12] K. Takahasi and K. Wakimoto, *On unbiased estimates of the population mean based on the sampling stratified by means of ordering*, Ann. Inst. Statist. Math. 20 (1968), pp. 1–31.

[13] D. A. Wolfe, *Ranked set sampling: An approach to more efficient data collection*, Statist. Sci. 19 (4) (2004), pp. 636–643.

[14] V. Zardasht and M. Asadi, *Evaluation of $P(X_t > Y_t)$ when both X_t and Y_t are residual lifetimes of two systems*, Stat. Neerl. 64 (4) (2010), pp. 460–481.

M. Mahdizadeh
Department of Statistics
Hakim Sabzevari University
P.O. Box 397, Sabzevar, Iran
E-mail: mahdizadeh.m@live.com

Ehsan Zamanzade
Department of Statistics
University of Isfahan
P.O. Box 81746-73441, Isfahan, Iran
E-mail: e.zamanzade@sci.ui.ac.ir;
ehsanzamanzadeh@yahoo.com