VOLUME/CİLT 13 NUMBER/SAYI 2 JULY/TEMMUZ 2021

İSTATİSTİK

JOURNAL OF THE TURKISH STATISTICAL ASSOCIATION

TÜRK İSTATİSTİK DERNEĞİ DERGİSİ

ISSN: 1300-4077

SUPPORTED AND PRINTED BY

İZMİR UNIVERSITY OF ECONOMICS

İSTATİSTİK
JOURNAL OF THE TURKISH STATISTICAL ASSOCIATION

ISSN: 1300-4077
https://dergipark.org.tr/tr/pub/ijtsa

SUPPORTED AND PRINTED BY

İZMİR UNIVERSITY OF ECONOMICS

Izmir University of Economics Sakarya Cad. No:156 Balcova - Izmir/TURKEY

Phone: +90 (232) 488-8139 Fax: + 90 (232) 279-2626 www.ieu.edu.tr

AIM AND SCOPE

İSTATİSTİK, Journal of the Turkish Statistical Association is a refereed journal which publishes papers containing original contributions to probability, statistics and the interface of them with other disciplines where prosperity of scientific thought emerges through innovation, vitality and communication on interdisciplinary grounds. The Journal is published four-monthly. The media of Journal is English. The main areas of publication are probability and stochastic processes, theory of statistics, applied statistics, statistical computing and simulation, and interdisciplinary applications in social, demographic, physical, medical, biological, agricultural studies, engineering, computer science, management science, econometrics, etc.

In addition, the journal contains original research reports, authoritative review papers, discussed papers and occasional special issues or relevant conference proceedings.

ABSTRACTED/INDEXED: İSTATİSTİK, Journal of the Turkish Statistical Association is indexed in ULAKBİM TR Dizin Database, MathSciNet, Zentralblatt MATH and EBSCO.

SUBMISSION OF MANUSCRIPTS:

Authors should submit their papers electronically by using online submission system (https://dergipark.org.tr/tr/pub/ijtsa).

As part of the submission process, authors are required to check their submission's compliance with all of the following items. Submissions that do not adhere to these guidelines may be returned to authors.

1. The submission has not been previously published, or is being considered for publication in another journal.

2. The submission file is in Portable Document Format (PDF) prepared in LaTeX (TEX) document file format.

3. Where available, URLs for the references have been provided.

4. The text is single-spaced; uses a 12-point font; employs italics, rather than underlining (except with URL addresses); and all illustrations, figures, and tables are placed within the text at the appropriate points, rather than at the end.

5. References should be listed in alphabetical order and should be in the following formats:

Cai, T. and Low, M. (2005). Non-quadratic estimators of a quadratic functional. *The Annals of Statistics*, 33, 2930-2956.

Meyer, Y. (1992). *Wavelets and Operators.* Cambridge University Press, Cambridge.

Cox, D. (1969). Some sampling problems in technology. *In New Developments in Survey Sampling* (N.L. Johnson and H. Smith, Jr., eds.). Wiley, New York, 506-527.

İSTATİSTİK: JOURNAL OF THE TURKISH STATISTICAL ASSOCIATION
Vol. 13, No. 2, July 2021, pp. 39–51
ISSN 1300-4077 | 21 | 2 | 39 | 51

İSTATİSTİK

IMPROVED ESTIMATORS FOR ESTIMATING THE POPULATION MEAN IN TWO OCCASION SUCCESSIVE SAMPLING

Vishwantra Sharma *
Department of Statistics,
Jammu University,
J&K, India

Sunil Kumar
Department of Statistics,
Jammu University,
J&K, India

Abstract: This paper addresses the problem of estimating the population mean of the study variable in two occasions successive sampling. Based on the available information from the first and second occasions, class of estimators produced under two situations, i) when the information on a positively correlated auxiliary variable with the study variable is available on both the occasions and ii) when the information on the auxiliary variable which is negatively correlated with the study variable is available on both the occasions. Properties of the suggested class of estimators have been studied and compared with the sample mean estimator with no matching from the previous occasion and traditional successive sampling linear estimator. The study is supported by an optimal replacement policy. Empirical study also has been illustrated to show the performance of the recommended estimators theoretically.

Key words: Study variable, Auxiliary variable, Bias, Mean squared error, Successive sampling.

1. Introduction

In most surveys, the interest is on the current average despite looking at it from one occasion to the next occasion and all occasions. In successive (rotation) sampling, it is common to use the entire information gathered on the previous occasions to improve the precision of the estimator on the current occasion. The main objective of the sampling on two successive occasions is to estimate the population parameters viz. population total, mean, ratio, product, etc. for the most recent occasion as well as changes in the parameters from one occasion to the next occasion, see Okafor and Arnab [6]. Jessen [4] was the first who pioneered the procedure of utilizing the information obtained on the first occasion in improving the estimates of the current occasion. Patterson [7] extended the work of Jessen from two occasions to more. Further, Eckler [2], Rao and Graham [8], Singh et al. [10], Feng and Zou [3], Biradar and Singh [1], Singh and Vishwakarma [12], Singh and Vishwakarma [13], Singh and Pal [15] among others have suggested several estimators by using the auxiliary information for estimating the population mean on the current occasion successive (rotation) sampling.

In this paper, we extend a procedure of utilizing the information of the auxiliary variable readily available on both the equations under two different situations, by suggesting the estimator of the population mean $\bar{Y}$ of the study variable y:

Situation I: When the auxiliary variable z_1 is positively correlated with the study variable y.

Situation II: Readily available auxiliary variable z_2 is negatively correlated with the study variable y.

* Corresponding author. E-mail address:vishwantrasharma07@gmail.com

Keeping in view the situation I and II, we have suggested two estimators and studied the properties of the suggested estimators. The behaviour of the suggested estimator is explained through empirical study. We found that the proposed study is more efficient than the other considered estimators when there is close association between auxiliary and study variables.

2. Notations used and the proposed estimator

Consider a finite population $U = (U_1, U_2, ..., U_N)$ of N distinct identifiable units. Let the variable under study on the first (second) occasion be denoted by $x(y)$ respectively. It is assumed that the information on the auxiliary variable z_1 and z_2 are known and have positive and negative correlation with x and y respectively readily available on both the occasions. A simple random sample (without replacement) of n units is taken on the first occasion from population U. A random sub-sample of $m(= n\lambda)$ units is retained (matched) for use on the second occasion. Now, at the current occasion, we again withdraw a simple random sample (without replacement) of size $u = n - m = n\mu$ units from the remaining $(N - n)$ units of the population so that the sample size on second occasion is also n. λ and μ are the fractions of matched and fresh samples respectively at the second (current) occasion such that $(\lambda + \mu = 1)$. We shall use the following notations:

- $\bar{X}, \bar{Y}, \bar{Z}_1, \bar{Z}_2$: The population means of variables x, y, z_1 and z_2 respectively.
- $S_{xy}, S_{yz_1}, S_{yz_2}, S_{xz_1}, S_{xz_2}$: The population covariance between variables in suffixes.
- $\rho_{xy}, \rho_{yz_1}, \rho_{yz_2}, \rho_{xz_1}, \rho_{xz_2}$: The population correlation coefficients between variables in suffixes.
- $S_x^2, S_y^2, S_{z_1}^2, S_{z_2}^2$: The population variances of x, y, z_1 and z_2 respectively.

For obtaining the expression of bias and mean squared error of the proposed estimator, we assume that

$$y_u = \bar{Y}(1+e_{0u}), y_m = \bar{Y}(1+e_{0m}), x_m = \bar{X}(1+e_{1m}), x_n = \bar{X}(1+e_{1n}), z_{1u} = \bar{Z}_1(1+e_{1u}),$$

$$z_{2u} = \bar{Z}_2(1+e_{2u}), z_{1n} = \bar{Z}_1(1+e_{2n}), z_{2n} = \bar{Z}_2(1+e_{2n'}), b_{yz_1u} = \frac{s_{yz_1(u)}}{s^2_{z_1(u)}}, b_{yx(m)} = \frac{s_{yx(m)}}{s^2_{x(m)}},$$

$$s_{yz_1(u)} = S_{yz_1(u)}(1+e_{3u}), s^2_{z_1(u)} = S^2_{z_1(u)}(1+e_{4u}), s_{yx(m)} = S_{yx(m)}(1+e_{3m}), s^2_{x(m)} = S^2_{x(m)}(1+e_{4m})$$

such that

$$E(e_{0u}) = E(e_{0m}) = E(e_{1m}) = E(e_{1n}) = E(e_{1u}) = E(e_{2u}) = E(e_{2n}) = E(e_{2n'}) = E(e_{3u})$$

$$= E(e_{4u}) = E(e_{3m}) = E(e_{4m}) = 0 \text{ and}$$

$$E(e^2_{0u}) = (\tfrac{1}{u} - \tfrac{1}{N})C^2_y, E(e^2_{0m}) = (\tfrac{1}{m} - \tfrac{1}{N})C^2_y,\ E(e^2_{1m}) = (\tfrac{1}{m} - \tfrac{1}{N})C^2_x, E(e^2_{1n}) = (\tfrac{1}{n} - \tfrac{1}{N})C^2_x,$$

$$E(e^2_{1u}) = (\tfrac{1}{u} - \tfrac{1}{N})C^2_{z_1}, E(e^2_{2u}) = (\tfrac{1}{u} - \tfrac{1}{N})C^2_{z_2}, E(e^2_{2n}) = (\tfrac{1}{n} - \tfrac{1}{N})C^2_{z_1}, E(e'^2_{2n}) = (\tfrac{1}{n} - \tfrac{1}{N})C^2_{z_2},$$

$$E(e_{0u}e_{0m}) = -(\tfrac{1}{N})C^2_y, E(e_{0u}e_{1m}) = -(\tfrac{1}{N})\rho_{xy}C_yC_x,\ E(e_{0u}e_{1n}) = -(\tfrac{1}{N})\rho_{xy}C_yC_x,$$

$$E(e_{0u}e_{1u}) = (\tfrac{1}{u} - \tfrac{1}{N})\rho_{yz_1}C_yC_{z_1},\ E(e_{0u}e_{2u}) = (\tfrac{1}{u} - \tfrac{1}{N})\rho_{yz_2}C_yC_{z_2}, E(e_{0u}e_{2n}) = -(\tfrac{1}{N})\rho_{xz_1}C_yC_{z_1},$$

$$E(e_{0u}e_{2n'}) = -(\tfrac{1}{N})\rho_{xz_1}C_yC_{z_1}, E(e_{0m}e_{1m}) = (\tfrac{1}{m} - \tfrac{1}{N})\rho_{yx}C_yC_x, E(e_{0m}e_{1n}) = (\tfrac{1}{n} - \tfrac{1}{N})\rho_{yx}C_yC_x,$$

$$E(e_{0m}e_{1u}) = -(\tfrac{1}{N})\rho_{yz_1}C_yC_{z_1}, E(e_{0m}e_{2u}) = -(\tfrac{1}{N})\rho_{yz_2}C_yC_{z_2}, E(e_{0m}e_{2n}) = (\tfrac{1}{n} - \tfrac{1}{N})\rho_{yz_1}C_yC_{z_1},$$

$$E(e_{0m}e_{2n'}) = (\tfrac{1}{n} - \tfrac{1}{N})\rho_{yz_2}C_yC_{z_2}, E(e_{0m}e_{2n}) = (\tfrac{1}{n} - \tfrac{1}{N})C^2_x, E(e_{1m}e_{1u}) = -(\tfrac{1}{N})\rho_{xz_1}C_xC_{z_1},$$

$E(e_{1m}e_{2u}) = -(\frac{1}{N})\rho_{xz_2}C_xC_{z_2}, E(e_{1n}e_{1u}) = -(\frac{1}{N})\rho_{xz_1}C_xC_{z_1}, E(e_{1n}e_{2u}) = -(\frac{1}{N})\rho_{xz_2}C_xC_{z_2},$

$E(e_{1n}e_{2n}) = (\frac{1}{n} - \frac{1}{N})\rho_{xz_1}C_xC_{z_1}, E(e_{1n}e_{2n'}) = (\frac{1}{n} - \frac{1}{N})\rho_{xz_2}C_xC_{z_2}, E(e_{1u}e_{2u}) = (\frac{1}{u} - \frac{1}{N})\rho_{z_1z_2}C_{z_1}C_{z_2},$

$E(e_{1u}e_{2n}) = -(\frac{1}{N})C_{z_1}^2, E(e_{1u}e_{2n'}) = -(\frac{1}{N})C_{z_2}^2, E(e_{2n}e_{2n'}) = (\frac{1}{n} - \frac{1}{N})\rho_{z_1z_2}C_{z_1}C_{z_2},$

$E(e_{1u}e_{3u}) = \frac{(N-u)}{u(N-2)}\frac{\mu_{012}}{\bar{Z}\mu_{011}}, E(e_{1u}e_{4u}) = \frac{(N-u)}{u(N-2)}\frac{\mu_{003}}{\bar{Z}\mu_{002}}, E(e_{2u}e_{3u}) = \frac{(N-u)}{u(N-2)}\frac{\mu_{012}}{\bar{Z}\mu_{011}},$

$E(e_{2u}e_{4u}) = \frac{(N-u)}{u(N-2)}\frac{\mu_{003}}{\bar{Z}\mu_{002}}, E(e_{1m}e_{3m}) = \frac{(N-m)}{m(N-2)}\frac{\mu_{210}}{\bar{X}\mu_{110}}, E(e_{1n}e_{3m}) = \frac{(N-n)}{n(N-2)}\frac{\mu_{210}}{\bar{X}\mu_{110}},$

$E(e_{1m}e_{4m}) = \frac{(N-m)}{m(N-2)}\frac{\mu_{300}}{\bar{X}\mu_{200}}, E(e_{1n}e_{4m}) = \frac{(N-n)}{n(N-2)}\frac{\mu_{300}}{\bar{X}\mu_{200}},$

$\mu_{pqr} = E((x_i - \bar{X})^p(y_i - \bar{Y})^q(z_i - \bar{Z})^r), \ (p,q,r) = 0,1,2., C_x^2 = S_x^2/\bar{X}^2, C_y^2 = S_y^2/\bar{Y}^2,$

$C_{z_1}^2 = S_{z_1}^2/\bar{Z}_1^2, C_{z_2}^2 = S_{z_2}^2/\bar{Z}_2^2.$

We assume that information on the auxiliary character is readily available on both occasions under two different situations, one can define the estimator when

Situation I: Estimation of the population mean $\bar{Y}$ of the study variable y when the auxiliary variable 'z_1' is positively correlated with the study variable.
In a situation, when the regression of Y on X is a straight line that does not pass through the origin then regression estimators are used. Replacing regression estimator in place of a sample mean and using in exponential-type estimators of Singh and Pal [15]. We have suggested two independent estimators for estimating the population mean $\bar{Y}$ of the study variable y on the second occasion. One is based on the sample of size $u(=n\mu)$ drawn afresh on the second occasion defined by

$$T_u = t_{lregu} exp\left(\frac{\bar{Z}_1 - \bar{z}_{1u}}{\bar{Z}_1 + \bar{z}_{1u}}\right) \tag{2.1}$$

where $t_{lregu} = [\bar{y}_u + b_{yz_1(u)}(\bar{Z}_1 - \bar{z}_{1u})]$ and $b_{yz_1(u)}$ is the sample regression coefficient of y and z_1 based on the sample size u.
Second estimator is based on the sample of size $m(=n\lambda)$ common for both the occasions is defined by

$$T_m = t_{lregm} exp\left(\frac{\bar{x}_n - \bar{x}_m}{\bar{x}_n + \bar{x}_m}\right) exp\left(\frac{\bar{Z}_1 - \bar{z}_{1n}}{\bar{Z}_1 + \bar{z}_{1n}}\right) \tag{2.2}$$

where $t_{lregm} = [\bar{y}_m + b_{yx(m)}(\bar{x}_n - \bar{x}_m)]$ and $b_{yx(m)}$ is the sample regression coefficient of y and x based on the matched sample of size m.

The estimator T_u may be used to estimate the population mean on each occasion, while the estimator T_m is suitable to estimate the change over occasions. To device suitable estimation procedures for both the problems simultaneously, a convex linear combination of T_u and T_m is considered as a final estimator of the population mean $\bar{Y}$ and is given by

$$T = \phi T_u + (1-\phi)T_m \tag{2.3}$$

where $\phi(0 \leq \phi \leq 1)$ is an unknown scalar to be defined such that the mean squared error (MSE) of T is minimum.

Remark 1 : For estimating the mean on each occasion the estimator T_u is suitable, which implies that for ϕ close to 1 while for estimating the change from one occasion to next occasion, the estimator T_m is more suitable so that the value of ϕ might be close to 0. For asserting both the problems simultaneously, the optimum (minimized) choice of ϕ is required.

3. Bias and Mean Square Error of T

Since T_u and T_m are biased estimators of $\bar{Y}$, therefore the estimator T is also a biased estimator of $\bar{Y}$. For bias, express the estimator T_u and T_m in terms of $\epsilon^{'}$s, we have

$$T_u = \{\bar{Y}(1+e_{0u}) - b_{yz_1}(u)\bar{Z}_1 e_{1u}\} exp\left[\frac{-e_{1u}}{2}\left(1+\frac{e_{1u}}{2}\right)^{-1}\right] \tag{3.1}$$

$$T_m = \bar{Y}\{(1+e_{0m}) + k_{yx}(e_{1n} - e_{1m} + e_{1n}e_{3m} - e_{1n}e_{4m} - e_{1m}e_{3m} + e_{1m}e_{4m})\} \\ exp\left[\frac{e_{1n}-e_{1m}}{2}\left(1+\frac{e_{1n}+e_{1m}}{2}\right)^{-1}\right] exp\left[\frac{-e_{2n}}{2}\left(1+\frac{e_{2n}}{2}\right)^{-1}\right] \tag{3.2}$$

Expanding the right-hand side of equation (4) and (5) in terms of e's and neglecting the terms having power greater than two, we get

$$T_u \cong \bar{Y}\left[1 + e_{0u} - \left(\frac{1}{2}\right)e_{1u} - \left(\frac{1}{2}\right)e_{0u}e_{1u} + \left(\frac{3}{8}\right)e_{1u}^2 - k_{yz_1}(e_{1u} - \left(\frac{1}{2}\right)e_{1u}^2 - e_{1u}e_{4u} + e_{1u}e_{3u}\right]$$

$$(T_u - \bar{Y}) \cong \bar{Y}\left[e_{0u} - \left(\frac{1}{2}\right)e_{1u} - \left(\frac{1}{2}\right)e_{0u}e_{1u} + \left(\frac{3}{8}\right)e_{1u}^2 - k_{yz_1}(e_{1u} - \left(\frac{1}{2}\right)e_{1u}^2 - e_{1u}e_{4u} + e_{1u}e_{3u}\right] \tag{3.3}$$

$$T_m \cong \bar{Y}\Big[1 + e_{0m} - \left(\frac{1}{2}\right)(e_{1n} - e_{1m} - e_{2n}) + \left(\frac{3}{8}\right)e_{2n}^2 + \left(\frac{1}{2}\right)(e_{0m}e_{1n} - e_{0m}e_{1m} - \\ e_{0m}e_{2n}) - \left(\frac{1}{4}\right)(e_{1n}^2 - e_{1m}^2 + e_{1n}e_{2n} - e_{1m}e_{2n}) + \left(\frac{1}{8}\right)(e_{1n}^2 + e_{1m}^2 - 2e_{1n}e_{1m}) \\ + k_{yx}\{e_{1n} - e_{1m} + \left(\frac{1}{2}\right)(e_{1n}^2 + e_{1m}^2) - \left(\frac{1}{2}\right)(e_{1n}e_{2n} + e_{1n}e_{1m} - e_{1m}e_{2n} + \\ e_{1n}e_{1m}) + e_{1n}e_{3m} - e_{1n}e_{4m} - e_{1m}e_{3m} + e_{1m}e_{4m}\}\Big]$$

$$(T_m - \bar{Y}) \cong \bar{Y}\Big[e_{0m} - \left(\frac{1}{2}\right)(e_{1n} - e_{1m} - e_{2n}) + \left(\frac{3}{8}\right)e_{2n}^2 + \left(\frac{1}{2}\right)(e_{0m}e_{1n} - e_{0m}e_{1m} - e_{0m}e_{2n}) - \\ \left(\frac{1}{4}\right)(e_{1n}^2 - e_{1m}^2 + e_{1n}e_{2n} - e_{1m}e_{2n}) + \left(\frac{1}{8}\right)(e_{1n}^2 + e_{1m}^2 - \\ 2e_{1n}e_{1m}) + k_{yx}(e_{1n} - e_{1m} + \left(\frac{1}{2}\right)(e_{1n}^2 + e_{1m}^2) - \left(\frac{1}{2}\right)(e_{1n}e_{2n} + e_{1n}e_{1m} \\ - e_{1m}e_{2n} + e_{1n}e_{1m}) + e_{1n}e_{3m} - e_{1n}e_{4m} - e_{1m}e_{3m} + e_{1m}e_{4m})\Big] \tag{3.4}$$

where, $k_{yz_1} = \rho_{yz_1}\frac{C_y}{C_{z_1}}$ and $k_{yx} = \rho_{yx}\frac{C_y}{C_x}$.
Taking expectation on both sides of equation (6) and (7), one can obtain the bias of T_u and T_m to

the first degree of approximation as

$$B(T_u) = \bar{Y}\left[\left(\frac{1}{u} - \frac{1}{N}\right)\left[\left(\frac{3}{8}\right) + k_{yz_1}\right]C_{Z_1}^2 - \left(\frac{1}{2}\right)\left(\frac{1}{u} - \frac{1}{N}\right)\rho_{yz_1}C_yC_{z_1} - k_{yz_1}\left(\frac{N-u}{u(N-2)\bar{Z}}\right)\left[\frac{\mu_{012}}{\mu_{011}} - \frac{\mu_{003}}{\mu_{002}}\right]\right] \tag{3.5}$$

$$B(T_m) = \bar{Y}\left[\left(\frac{3}{8}\right)\left(\frac{1}{n} - \frac{1}{N}\right)C_{z_1}^2 + \left(\frac{1}{2}\right)\left[\left(\frac{1}{n} - \frac{1}{m}\right)\rho_{yx}C_yC_x - \left(\frac{1}{n} - \frac{1}{N}\right)\rho_{yz_1}C_yC_{z_1}\right] - \left(\frac{7}{8}\right)\left(\frac{1}{n} - \frac{1}{m}\right)C_x^2 + \frac{(N-n)}{n(N-2)\bar{X}}\left[\left(\frac{1}{n}\right)\left(\frac{\mu_2 10}{\mu_1 10} - \frac{\mu_0 12}{\mu_0 11}\right) - \left(\frac{1}{m}\right)\left(\frac{\mu_{210}}{\mu_{110}} - \frac{\mu_{012}}{\mu_{011}}\right)\right]\right] \tag{3.6}$$

For MSE, squaring both side of equation (6) and (7), and neglecting terms of e's having power greater than two, we have

$$(T_u - \bar{Y})^2 \cong \bar{Y}^2\left[e_{0u} - \left(\frac{1}{2}\right)e_{1u} - k_{yz_1}e_{1u}\right]^2 \tag{3.7}$$

$$(T_m - \bar{Y})^2 \cong \bar{Y}^2\left[e_{0m} + \left(\frac{1}{2}\right)(e_{1n} - e_{1m} - e_{2n}) - k_{yx}(e_{1n} - e_{1m})\right]^2 \tag{3.8}$$

Taking expectations to both sides of equation (10) and (11) we get the MSE of T_u and T_m respectively, as

$$MSE(T_u) = \bar{Y}^2\left(\frac{1}{u} - \frac{1}{N}\right)\left[C_{y^2} + C_{z_1}^2\left\{\left(\frac{1}{4}\right) + k_{yz_1}^2 + k_{yz_1}\right\} - (1 + 2k_{yz_1})\rho_{yz_1}C_yC_{z_1}\right] \tag{3.9}$$

$$MSE(T_m) = \bar{Y}^2\left[\left(\frac{1}{m} - \frac{1}{N}\right)\left\{C_y^2 + \left(\frac{1}{4}\right)C_x^2 + k_{yx}^2C_x^2 - \rho_{yx}C_yC_x + k_{yx}C_x^2 - 2k_{yx}\rho_{yx}C_yC_x\right\} + \left(\frac{1}{n} - \frac{1}{N}\right)\left\{\left(\frac{1}{4}\right)C_{z_1}^2 - \left(\frac{1}{4}\right)C_x^2 + k_{yx}^2C_x^2 + \rho_{yx}C_yC_x - k_{yx}C_x^2 - \rho_{yz_1}C_yC_{z_1} - 2k_{yx}\rho_{yx}C_yC_x\right\}\right] \tag{3.10}$$

The covariance between the two estimators T_u and T_m to the first degree of approximation is obtained as follows:

$$\begin{aligned}
Cov(T_u, T_m) &= E[(T_u - \bar{Y})(T_m - \bar{Y})] \\
&= \bar{Y}^2E\big[\big(e_{0u} - (1/2)e_{1u} - k_{yz_1}e_{1u}\big)\big(e_{0m} - (1/2)(e_{1n} - e_{1m} - e_{2n}) + k_{yx}(e_{1n} - e_{1m})\big)\big] \\
&= \bar{Y}^2E\big[e_{0u}e_{0m} - (1/2)e_{0m}e_{1u} - k_{yz_1}e_{0m}e_{1u} + (1/2)\big(e_{0u}e_{1n} - (1/2)e_{1u}e_{1n} - k_{yz_1}e_{1u}e_{1n} - e_{0u}e_{1m} + \\
&\quad (1/2)e_{1u}e_{1m} + k_{yz_1}e_{1u}e_{1m} - e_{0u}e_{2n} + (1/2)e_{1u}e_{2n} + k_{yz_1}e_{1u}e_{2n}\big) + k_{yx}\big(e_{0u}e_{1n} - (1/2)e_{1u}e_{1n} - \\
&\quad k_{yz_1}e_{1n}e_{1u} - e_{0u}e_{1m} + (1/2)e_{1u}e_{1m} + k_{yz_1}e_{1u}e_{1m}\big)\big] \\
&= -(\bar{Y}^2/N)\big[C_y^2 - \rho_{yz_1}C_yC_{z_1} + (1/4)C_{z_1}^2 - k_{yz_1}\rho_{yz_1}C_yC_{z_1} + (1/2)k_{yz_1}C_{z_1}^2\big]
\end{aligned} \tag{3.11}$$

Assumption 1. Considering the stability nature of the variables, the coefficient of variation of x, y, z_1, z_2 are assumed to be approximately equal ($C_y \cong C_x \cong C_{z_1} \cong C_{z_2}$), see Murthy [5], Reddy [9], Singh and Ruiz-Espejo [11]. Under Assumption 1, we state the following theorems without proof.

THEOREM 1. *The bias of the proposed estimator 'T' to the first degree of approximation is given by*

PROOF.

$$B(T) = \phi B(T_u) + (1-\phi)B(T_m) \tag{3.12}$$

where

$$B(T_u) = \bar{Y}\left[\left(\frac{1}{u}-\frac{1}{N}\right)\left(\frac{3}{8}+\frac{\rho_{yz_1}}{2}\right)C_y^2 - \rho_{yz_1}\left(\frac{N-u}{u(N-2)\bar{Z}_1}\right)\left(\frac{\mu_{012}}{\mu_{011}}-\frac{\mu_{003}}{\mu_{002}}\right)\right] \tag{3.13}$$

and

$$\begin{aligned} B(T_m) = \bar{Y}\Bigg[\Bigg\{&\left(\frac{3}{8}\right)\left(\frac{1}{n}-\frac{1}{N}\right)+\left(\frac{1}{2}\right)\left\{\left(\frac{1}{n}-\frac{1}{m}\right)\rho_{yx}-\left(\frac{1}{n}-\frac{1}{N}\right)\rho_{yz_1}\right\}C_y^2 \\ &-\left(\frac{7}{8}\right)\left(\frac{1}{n}-\frac{1}{m}\right)\Bigg\}+\frac{N-n}{n(N-2)\bar{X}}\Bigg\{\left(\frac{1}{n}\right)\left(\frac{\mu_{210}}{\mu_{110}}-\frac{\mu_{012}}{\mu_{011}}\right)-\left(\frac{1}{m}\right) \\ &\left(\frac{\mu_{210}}{\mu_{110}}-\frac{\mu_{300}}{\mu_{200}}\right)\Bigg\}\Bigg] \end{aligned} \tag{3.14}$$

THEOREM 2. *The MSE of 'T' to the first degree of approximation is obtained by*

PROOF.

$$MSE(T) = \phi^2 MSE(T_u) + (1-\phi)^2 MSE(T_m) + 2\phi(1-\phi)Cov(T_u,T_m), \tag{3.15}$$

where

$$MSE(T_u) = \left(\frac{1}{u}-\frac{1}{N}\right)\left[\frac{5}{4}-\rho_{yz_1}^2\right]S_y^2, \tag{3.16}$$

$$MSE(T_m) = \left[\frac{1}{m}\left(\frac{5}{4}+\rho_{yx}^2\right)+\frac{1}{n}\left(\rho_{yz_1}+\rho_{yx}^2\right)-\frac{1}{N}\left(\frac{5}{4}-\rho_{yz_1}\right)\right]S_y^2 \tag{3.17}$$

and

$$Cov(T_u,T_m) = -\frac{S_y^2}{N}\left[\frac{5}{4}-\frac{\rho_{yz_1}}{2}-\rho_{yz_1}^2\right] \tag{3.18}$$

4. Minimum mean squared error of the estimator 'T'

Since MSE(T) in equation (18) is a function of unknown constant ϕ, therefore, it can be minimized with respect to ϕ and equating it to zero, we get the optimum value of ϕ as

$$\phi_{opt} = \frac{[MSE(T_m) - Cov(T_u,T_m)]}{[MSE(T_u)+MSE(T_m)-2Cov(T_u,T_m)]} \tag{4.1}$$

By substituting the value of optimum 'ϕ' from equation (22) in equation (18) we will have the minimum MSE of 'T' as

$$min.MSE(T) = \frac{[MSE(T_u)MSE(T_m) - (Cov(T_u,T_m))^2]}{[MSE(T_u)+MSE(T_m)-2Cov(T_u,T_m)]} \tag{4.2}$$

Substituting the values of $MSE(T_u)$, $MSE(T_m)$ and $Cov(T_u,T_m)$ in equations (22) and (23), we will have the value of ϕ_{opt} and min.MSE(T), respectively.

For simplification, further we use the following notations,

$\delta_1 = N\alpha_2 - n\alpha_5,\ \delta_2 = N\alpha_1 - N\alpha_2 - n\alpha_5 - N\alpha_4,\ \delta_3 = n^2\alpha_8^2 - nN\alpha_2\alpha_4 - n^2\alpha_4\alpha_7,$

$\delta_4 = N^2\alpha_2\alpha_4 + n^2(\alpha_4\alpha_7 - \alpha_8 2), \delta_5 = N^2\alpha_1\alpha_4 - N^2\alpha_2\alpha_4 - nN\alpha_4\alpha_7,$

$\alpha_1 = (5/4) - \rho_{yx}^2, \alpha_2 = \rho_{yz_1} - \rho_{yx}^2, \alpha_3 = \rho_{yz_1}^2 - (\rho_{yz_1}/2),\ \alpha_4 = (5/4) - \rho_{yz_1}^2,$

$\alpha_5 = \rho_{yz_1}^2,\ \alpha_7 = (5/4) - \rho_{yz_1}, \alpha_8 = (5/4) - (\rho_{yz_1}/2) - \rho_{yz_1}^2.$

Now, we have the reduced form of ϕ_{opt} and min.MSE(T) from equation (22) and (23) as

$$\phi_{opt} = \frac{[\mu N\alpha_1 - \mu(1-\mu)(N\alpha_2 + n\alpha_3)]}{[\mu N\alpha_1 - (1-\mu)(\mu N\alpha_2 + n\mu\alpha_5 - N\alpha_4)]} \tag{4.3}$$

and

$$min.MSE(T) = \left(\frac{S_y^2}{nN}\right)\frac{\mu^2\delta_3 + \mu\delta_4 + \delta_5}{\mu^2\delta_1 + \mu\delta_2 + N\alpha_4} \tag{4.4}$$

5. Optimum replacement policy

For obtaining the optimum value of μ (fraction of a sample to be taken afresh at the second occasion) so that the population mean $\bar{Y}$ may be estimated with maximum precision, we minimize MSE of T in equation (25) by differentiating it with respect to 'μ' and hence we get the optimum value of 'μ' as

$$\mu^2\lambda_1 + \mu\lambda_2 + \lambda_3 = 0 \tag{5.1}$$

where $\lambda_1 = (\delta_2\delta_3 - \delta_1\delta_4);\ \lambda_2 = (2N\alpha_4\delta_3 - 2\delta_1\delta_5);\ \lambda_3 = (N\alpha_4\delta_4 - \delta_2\delta_5).$
Solving equation (26) for 'μ', we get

$$\hat{\mu} = \frac{-\lambda_2 \pm \sqrt{\lambda_2^2 - 4\lambda_1\lambda_3}}{2\lambda_1} \tag{5.2}$$

The value of $\hat{\mu}$ exists, if $\lambda_2^2 \geq 4\lambda_1\lambda_3$. For any combinations of correlations (ρ_{yx}, ρ_{yz_1}) that satisfy the condition of solution, two values of $\hat{\mu}$ are possible. If both the two values $\hat{\mu}$ are admissible, then the lowest one is best. Substituting the admissible values of $\hat{\mu}$, say μ_0, from equation (27) into (25), we get the optimum value of the mean squared error of 'T', which is given by

$$min.MSE(T)_{opt} = \left(\frac{S_y^2}{nN}\right)\left[\frac{\mu_0^2\delta_3 + \mu_0\delta_4 + \delta_5}{\mu_0^2\delta_1 + \mu_0\delta_2 + N\alpha_4}\right] \tag{5.3}$$

6. Efficiency comparison

The percent relative efficiencies of the estimators T with respect to (i) $\bar{y}_n$, when there is no matching, (ii) usual successive sampling estimator, $\hat{\hat{Y}} = \psi\bar{y}_u + (1-\psi)\bar{y}_{d'}$, when no auxiliary information is used at any occasion, where $[\bar{y}_{d'} = \bar{y}_m + b^m_{yx}(\bar{x}_n - \bar{x}_m)]$ have been obtained for different choices of ρ_{yx}, ρ_{yz_1} and ρ_{yz_2}. Since $\bar{y}_n$ and $\hat{\hat{Y}}$ are unbiased estimators of the population mean $\bar{Y}$, the variance of $\bar{y}_n$ and the minimum variance of $\hat{\hat{Y}}$ [as given in Sukhatme et al.[13]] are given by

$$V(\bar{y}_n) = \frac{1-f}{n} S_y^2 \tag{6.1}$$

$$V(\hat{\hat{Y}}) = \left[\left(\frac{1}{2}\right)\left\{1 + \sqrt{(1-\rho_{yx}^2)}\right\} - f\right]\frac{S_y^2}{n} \tag{6.2}$$

From (27), (28), (29), and (30) the percent relative efficiencies of the estimators 'T' with respect to $\bar{Y}_n$ are given by

$$\begin{aligned} E_1 = PRE(T, \bar{y}_n) &= \frac{V(\bar{y}_n)}{min.MSE(T)_{opt}} \times 100 \\ &= \frac{N(1-f)(\mu_0^2\delta_1 + \mu_0\delta_2 + N\alpha_4)}{\mu_0^2\delta_3 + \mu_0\delta_4 + \delta_5} \times 100 \end{aligned} \tag{6.3}$$

$$\begin{aligned} E_2 = PRE(T, \hat{\hat{Y}}) &= \frac{V(\hat{\hat{Y}})}{min.MSE(T)_{opt}} \times 100 \\ &= \frac{N\left[\left\{1 + \sqrt{(1-\rho_{yx}^2)}\right\} - 2f\right](\mu_0^2\delta_1 + \mu_0\delta_2 + N\alpha_4)}{2(\mu_0^2\delta_3 + \mu_0\delta_4 + \delta_5)} \times 100 \end{aligned} \tag{6.4}$$

For $N = 2000,\ n = 200$ and various choices of correlations (ρ_{yx}, ρ_{yz_1}) and using the formulae from equations (27), (31) and (32) we have computed the optimum values of μ_0 and percent relative efficiencies E_1 and E_2. The findings are displayed in Table 1.

TABLE 1. Optimum values μ_0 and percent relative efficiency of T with respect to $\bar{y}_n$ and $\hat{\hat{Y}}$.

ρ_{yx}	0.2			0.3			0.4			0.5		
ρ_{yz_1}	μ_0	E_1	E_2	μ_0	E_1	E_2	μ_0	E_1	E_2	μ_0	E_1	E_2
0.6	0.72	101.46	100.32	0.75	102.23	99.61	0.80	103.91	-	0.90	103.91	-
0.7	0.65	112.20	110.94	0.67	113.18	110.28	0.71	114.49	109.18	0.78	115.92	107.30
0.8	0.58	128.32	126.88	0.59	129.57	126.25	0.62	131.31	125.22	0.67	133.45	123.52
0.9	0.48	155.05	153.31	0.50	156.70	152.69	0.52	159.05	151.67	0.55	162.10	150.03
ρ_{yx}	0.6			0.7			0.8			0.9		
ρ_{yz_1}	μ_0	E_1	E_2	μ_0	E_1	E_2	μ_0	E_1	E_2	μ_0	E_1	E_2
0.6	1.13	102.77	91.35	2.16	-	-	-1.33	178.14	138.56	0.07	154.09	105.80
0.7	0.91	116.70	103.74	1.25	112.55	-	*	-	-	-0.27	184.80	126.89
0.8	0.75	135.61	120.54	0.9	135.96	114.37	1.59	119.04	-	-1.54	281.87	193.53
0.9	0.60	165.71	147.30	0.70	169.04	142.19	0.94	166.72	129.67	*	-	-

Note : * denotes μ_0 does not exist and − implies very low efficiency.

It is envisaged from Table 1 that the proposed estimator 'T' is more efficient than the estimators $\bar{y}_n$ and $\hat{\hat{Y}}$ for different levels of correlation between the variables (y and x) and (y and z_1). The following point have been noted from the Table 1 as

1. For moderate to high correlation between y and z_1, efficiency increases with respect to $\bar{y}_n$ and $\hat{\hat{Y}}$.
2. When the correlation between y and x is very high i.e, $\rho_{yx} = 0.9$ corresponding to the different levels of correlation between y and z_1 i.e, $(\rho_{yz_1} = 0.6 \text{ to } 0.9)$, the proposed estimator 'T' performs efficiently among $\bar{y}_n$ and $\hat{\hat{Y}}$ respectively.
3. With different levels of correlation between y and z_1 i.e, $(\rho_{yz_1} = 0.6 \text{ to } 0.9)$ and for different correlation between y and x i.e, $(\rho_{yx} = 0.2 \text{ to } 0.9)$, the PRE of the proposed estimator T increases except the case when $\rho_{yx} = 0.7\, and\, 0.8$ and $\rho_{yz_1} = 0.6\, to\, 0.9$ where the PRE of the proposed estimator first decreases then increases because the value of μ_0 first increases then decreases respectively.

Situation II : Estimation of the population mean $\bar{Y}$ of the study variable 'y' when the auxiliary variable z_2 is negatively correlated with the study variable 'y'.

This section deals with case II of our problem, where the correlation between study variable 'y' and the auxiliary variable z_2 is negative. In this case, for estimating the population mean $\bar{Y}$ at the current (second) occasion with negatively correlated auxiliary variable z_2 at the first (second) occasion, we suggest the following estimators as

$$T_u^* = \{\bar{y}_u + b_{yz_2(u)}(\bar{Z}_2 - \bar{z}_{2u})\} exp\left(\frac{\bar{z}_{2u} - \bar{Z}_2}{\bar{z}_{2u} + \bar{Z}_2}\right) \tag{6.5}$$

where $b_{yz_2}(u)$ is the sample regression coefficient of y and z_2 based on the sample size u.

$$T_m^* = \{\bar{y}_m + b_{yx(m)}(\bar{x}_n - \bar{x}_m)\} exp\left(\frac{\bar{x}_n - \bar{x}_m}{\bar{x}_n + \bar{x}_m}\right) exp\left(\frac{\bar{z}_{2n} - \bar{Z}_2}{\bar{z}_{2n} + \bar{Z}_2}\right) \tag{6.6}$$

where $b_{yx(m)}$ is the sample regression coefficient of y and x based on the matched sample of size m. Consider the linear combination of $T_u^*\, and\, T_m^*$, we define the following estimator as

$$T^* = \phi^* T_u^* + (1 - \phi^*) T_m^* \tag{6.7}$$

where ϕ^* is any suitably chosen scalar.
Using the result from section '2', one can obtain the bias and mean square error of $T_u^*\, and\, T_m^*$ respectively, results of which are mentioned in the form of theorems.

THEOREM 3. *The bias of the proposed estimator T^* to the first degree of approximation is*

PROOF.

$$B(T^*) = \phi^* B(T_u^*) + (1 - \phi^*) B(T_m^*) \tag{6.8}$$

where

$$B(T_u^*) = \bar{Y}\left[\left(\frac{1}{u} - \frac{1}{N}\right)\left[\frac{\rho_{yz_2} C_y C_{z_2}}{2} - \frac{C_{z_2}^2}{8}\right] - k_{yz_2}\left(\frac{1}{u} - \frac{1}{N}\right)\frac{C_{z_2}^2}{2} + k_{yz_2}\left(\frac{N-u}{u(N-2)\bar{Z}_2}\right)\left(\frac{\mu_{003}}{\mu_{002}} - \frac{\mu_{012}}{\mu_{011}}\right)\right] \tag{6.9}$$

and

$$B(T_m^*) = \bar{Y}\left[\left(\frac{1}{2}\right)\left(\frac{1}{n} - \frac{1}{N}\right)\rho_{yz_2} C_y C_{z_2} - \left(\frac{C_{z_2}^2}{8}\right)\left(\frac{1}{n} - \frac{1}{N}\right) - \left(\left(\frac{7}{8}\right)\frac{1}{n} - \frac{1}{m}\right) C_x^2 + \left(\frac{1}{2}\right)\left(\frac{1}{n} - \frac{1}{m}\right)\rho_{yx} C_y C_x + \left(\frac{1}{4}\right)\left(\frac{1}{n} - \frac{1}{N}\right)\rho_{xz_2} C_x C_{z_2} + k_{yx}\left\{\frac{1}{(N-2)\bar{X}}\left(\frac{\mu_{210}}{\mu_{110}} - \frac{\mu_{300}}{\mu_{200}}\right)\left(\frac{N-n}{n} - \frac{N-m}{m}\right)\right\}\right] \tag{6.10}$$

where $k_{yx} = \rho_{yx}\frac{C_y}{C_x}$, $k_{yz_2} = \rho_{yz_2}\frac{C_y}{C_{z_2}}$.

THEOREM 4. *To the first degree of approximation, the MSE of 'T*' is given by*

PROOF.

$$MSE(T^*) = \phi^{*2}MSE(T_u^*) + (1-\phi^*)^2 MSE(T_m^*) + 2\phi(1-\phi)Cov(T_u^*, T_m^*) \tag{6.11}$$

where,

$$MSE(T_u^*) = \bar{Y}^2\left(\frac{1}{u} - \frac{1}{N}\right)\left[C_y^2 + C_{z_2}^2((1/4) + k_{yz_2}^2 - \rho_{yz_2}) + (1-2k_{yz_2})\rho_{yz_2}C_yC_{z_2}\right] \tag{6.12}$$

$$\begin{aligned}MSE(T_m^*) = \bar{Y}^2\Bigg[&\left(\frac{1}{m} - \frac{1}{N}\right)\left\{C_y^2 + \left(\frac{1}{4}\right)C_x^2 + k_{yx}^2C_x^2 - \rho_{yx}C_yC_x + k_{yx}C_x^2 - 2k_{yx}\rho_{yx}C_yC_x\right\} + \\ &\left(\frac{1}{n} - \frac{1}{N}\right)\left\{\left(\frac{1}{4}\right)C_{z_2}^2 - \left(\frac{1}{4}\right)C_x^2 - k_{yx}^2C_x^2 + \rho_{yx}C_yC_x + \rho_{yz_2}C_yC_{z_2} - k_{yx}C_x^2 + 2k_{yx}\rho_{yx}C_yC_x\right\}\Bigg]\end{aligned} \tag{6.13}$$

and

$$Cov(T_u^*, T_m^*) = -(\bar{Y}^2/N)(C_y^2 + \rho_{yz_2}C_yC_{z_2} - k_{yz_2}\rho_{yz_2}C_yC_{z_2} + (1/4)C_{z_2}^2 - (1/2)\rho_{yz_2}C_{z_2}^2) \tag{6.14}$$

THEOREM 5. *Considering Assumption 1, the bias of the proposed estimator 'T*' reduces to*

PROOF.

$$B(T^*) = \phi^* B(T_u^*) + (1-\phi^*)B(T_m^*) \tag{6.15}$$

where

$$B(T_u^*) = \bar{Y}\left[\rho_{yz_2}\left\{\left(\frac{N-u}{u(N-2)\bar{Z}_2}\right)\left(\frac{\mu_{003}}{\mu_{002}} - \frac{\mu_{012}}{\mu_{011}}\right)\right\} - \left(\frac{1}{u} - \frac{1}{N}\right)\left(\frac{1}{8}\right)\right] \tag{6.16}$$

and

$$\begin{aligned}B(T_m^*) = \bar{Y}\Bigg[&\left\{\left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{\rho_{yz_2}}{2} - \frac{1}{8}\right) - \left(\frac{1}{n} - \frac{1}{m}\right)\left(\left(\frac{7}{8}\right) - \frac{\rho_{yx}}{2} - \frac{\rho_{xz_2}}{4}\right)\right\}C_y^2 + \\ &\rho_{yx}\left\{\frac{1}{(N-2)\bar{X}}\left(\frac{\mu_{210}}{\mu_{110}} - \frac{\mu_{300}}{\mu_{200}}\right)\left(\frac{N-n}{n} - \frac{N-m}{m}\right)\right\}\Bigg]\end{aligned} \tag{6.17}$$

THEOREM 6. *Under Assumption 1, the MSE of T^* to the first degree of approximation reduces to*

PROOF.

$$MSE(T^*) = \phi^{*2}MSE(T_u^*) + (1-\phi^*)^2 MSE(T_m^*) + 2\phi(1-\phi)Cov(T_u^*, T_m^*) \tag{6.18}$$

where,

$$MSE(T_u^*) = (1/u - 1/N)[(5/4) - \rho_{yz_2}^2]S_y^2, \tag{6.19}$$

$$MSE(T_m^*) = [(1/m)(5/4 - \rho_{yx}^2) + (1/n)(\rho_{yx}^2 + \rho_{yz_2}) - (1/N)(5/4 + \rho_{yz_2})]S_y^2 \tag{6.20}$$

and

$$Cov(T_u^*, T_m^*) = -(S_y^2/N)((5/4) + (1/2)\rho_{yz_2} - \rho_{yz_2}^2) \tag{6.21}$$

7. Minimum mean squared error of the estimator T^*

For minimum MSE of T^*, we partially differentiate equation (46) with respect to the unknown constant ϕ^* and equating it to zero, we get the optimum value of ϕ^* as

$$\phi^*_{opt} = \frac{[MSE(T^*_m) - Cov(T^*_u, T^*_m)]}{[MSE(T^*_u) + MSE(T^*_m) - 2Cov(T^*_u, T^*_m)]}$$

$$= \frac{[\mu n N\alpha_1 - \mu(1-\mu)(N\alpha_2^{'} + n\alpha_3^{'})]}{[\mu N\alpha_1 - (1-\mu)(\mu N\alpha_2^{'} + n\mu\alpha_5^{'} - N\alpha_4^{'})]} \tag{7.1}$$

Putting the value of ϕ^*_{opt} from equation (50) in equation (46) we get the minimized MSE of T^* as

$$min.MSE(T^*) = \frac{[MSE(T^*_u)MSE(T^*_m) - Cov(T^*_u, T^*_m)^2]}{[MSE(T^*_u) + MSE(T^*_m) - 2Cov(T^*_u, T^*_m)]}$$

$$= \left(\frac{S_y^2}{nN}\right)\left[\frac{\mu^2\delta_3^{'} + \mu\delta_4^{'} + \delta_5^{'}}{\mu^2\delta_1^{'} + \mu\delta_2^{'} + N\alpha_4^{'}}\right] \tag{7.2}$$

where
$\delta_1^{'} = n\alpha_2^{'} - N\alpha_5^{'}$, $\delta_2^{'} = N\alpha_1 + N\alpha_2^{'} - n\alpha_5^{'} - N\alpha_4^{'}$, $\delta_3^{'} = n^2\alpha_8^{'2} - nN\alpha_2^{'}\alpha_4^{'} - n^2\alpha_4^{'}\alpha_7^{'}$,

$\delta_4^{'} = N^2\alpha_2^{'}\alpha_4^{'} + n^2(\alpha_4^{'}\alpha_7^{'} - \alpha_8^{'2}) + nN\alpha_4^{'}(\alpha_7^{'} + \alpha_2^{'} - \alpha_1)$,

$\delta_5^{'} = N^2\alpha_1\alpha_4^{'} - N^2\alpha_2^{'}\alpha_4^{'} - nN\alpha_4^{'}\alpha_7^{'}$,
$\alpha_1 = (5/4) - \rho_{yx}^2$, $\alpha_2^{'} = \rho_{yz_2} + \rho_{yx}^2$, $\alpha_3^{'} = \rho_{yz_2}^2 + \rho_{yz_2}/2$, $\alpha_4^{'} = (5/4) - \rho_{yz_2}^2$,

$\alpha_5^{'} = \rho_{yz_2}^2$, $\alpha_7^{'} = (5/4) + \rho_{yz_2}\alpha_8^{'} = (5/4) + (\rho_{yz_2}/2) - \rho_{yz_2}^2$

8. Optimum replacement policy in case of negative correlation between study and auxiliary variables.

In this section, we will obtain the optimum value of μ (fraction of sample to be drawn afresh at the second occasion) so that the population mean $\bar{Y}$ may be estimated with maximum precision. Differentiating the min.MSE(T^*) given by equation (52) with respect to μ and equating to zero we get

$$\mu^2(\delta_2^{'}\delta_3^{'} - \delta_1^{'}\delta_4^{'}) + \mu(2N\alpha_4^{'}\delta_3^{'} - 2\delta_1^{'}\delta_5^{'}) + (N\alpha_4^{'}\delta_4^{'} - \delta_2^{'}\delta_5^{'}) = 0$$

$$\mu^2\lambda_1^{'} + \mu\lambda_2^{'} + \lambda_3^{'} = 0 \tag{8.1}$$

where $\lambda_1^{'} = \delta_2^{'}\delta_3^{'} - \delta_1^{'}\delta_4^{'}$, $\lambda_2^{'} = 2N\alpha_4^{'}\delta_3^{'} - 2\delta_1^{'}\delta_5^{'}$, $\lambda_3^{'} = N\alpha_4^{'}\delta_4^{'} - \delta_2^{'}\delta_5^{'}$
Solving equation (52) for μ, we get

$$\hat{\mu} = \frac{-\lambda_2^{'} \pm \sqrt{(\lambda_2^{'2} - 4\lambda_1^{'}\lambda_3^{'})}}{2\lambda_1^{'}} \tag{8.2}$$

The value of $\hat{\mu}$ exists, if $(\lambda_2^{'2} - 4\lambda_1^{'}\lambda_3^{'}) \geq 0$. For any combinations of correlations $(\rho_{yx}, \rho_{yz_2}))$ that satisfy the solution, two values of $\hat{\mu}$ are possible. Substituting the admissible values of $\hat{\mu}$, say μ_0, from equation (53) into (51), we get the optimum value of mean squared error of T^*, which is given by

$$min.MSE(T^*)_{opt} = \left(\frac{S_y^2}{nN}\right)\frac{(\mu_0^2\delta_3^{'} + \mu_0\delta_4^{'} + \delta_5^{'})}{(\mu_0^2\delta_1^{'} + \mu_0\delta_2^{'} + N\alpha_4^{'})} \tag{8.3}$$

TABLE 2. Optimum values μ_0 and percent relative efficiency of T^* with respect to $\bar{y}_n$ and $\hat{\hat{Y}}$.

ρ_{yx}	0.8			0.9		
ρ_{yz_2}	μ_0	E_1^*	E_2^*	μ_0	E_1^*	E_2^*
-0.70	0.086	131.14	102.00	*	-	-
-0.72	0.606	132.27	102.88	*	-	-
-0.74	0.431	131.20	102.04	2.161	97.40	60.77
-0.76	0.909	128.97	100.31	1.626	130.71	81.55
-0.78	0.221	126.15	98.12	3.690	68.37	46.94
-0.80	0.156	123.06	95.72	2.041	140.28	96.32
-0.82	0.109	119.89	93.25	1.281	170.62	117.15
-0.84	0.073	116.73	90.79	0.834	180.53	125.32
-0.86	0.047	113.63	88.38	0.551	184.95	126.98
-0.88	0.027	110.64	86.05	0.365	182.54	125.33
-0.90	0.012	107.77	83.82	0.241	177.82	122.09
-0.92	0.002	105.01	81.67	0.155	172.10	118.16
-0.94	-0.005	102.38	79.63	0.097	166.06	114.01
-0.96	-0.010	99.87	77.68	0.056	160.05	109.89
-0.98	-0.014	97.48	75.81	0.029	154.25	105.91

Note: $*$ denotes μ_0 does not exist and $-$ implies very low efficiency.

9. Efficiency comparison

The percent relative efficiencies of the estimators T with respect to (i) $\bar{y}_n$, when there is no matching, (ii) usual successive sampling estimator, $\hat{\hat{Y}} = \psi\bar{y}_u + (1-\psi)\bar{y}_{d'}$, when no auxiliary information is used at any occasion, where $[\bar{y}_{d'} = \bar{y}_m + b_{yx}^m(\bar{x}_n - \bar{x}_m)]$ have been obtained for different choices of ρ_{yx}, ρ_{yz_1} and ρ_{yz_2}. Since $\bar{y}_n$ and $\hat{\hat{Y}}$ are unbiased estimators of the population mean $\bar{Y}$, the variance of $\bar{y}_n$ and the minimum variance of $\hat{\hat{Y}}$ [as given in Sukhatme et al.[16]] are given by equation (29) and (30) in section 6.

From (29), (30) and (54), the percent relative efficiencies of the estimators T^* with respect to $\bar{y}_n$ and $\hat{\hat{Y}}$ are given by

$$E_1^* = PRE(T^*, \bar{y}_n) = \frac{V(\bar{y}_n)}{min.MSE(T^*)_{opt}} \times 100$$

$$= \frac{N(1-f)[\mu_0^2\delta_1' + \mu_0\delta_2' + N\alpha_4']}{\mu_0^2\delta_3' + \mu_0\delta_4' + \delta_5'} \times 100 \tag{9.1}$$

$$E_2^* = PRE(T^*, \hat{\hat{Y}}) = \frac{V(\hat{\hat{Y}})}{min.MSE(T^*)_{opt}} \times 100$$

$$= \frac{N\left[\{1+\sqrt{(1-\rho_{yx}^2)}\} - 2f\right](\mu_0^2\delta_1' + \mu_0\delta_2' + N\alpha_4')}{2(\mu_0^2\delta_3' + \mu_0\delta_4' + \delta_5')} \times 100 \tag{9.2}$$

For $N = 2000, n = 200$, and various choices of correlations$(\rho_{yx}, \rho_{yz_2})$and using the formulae from equations (53), (55) and (56) we have computed the optimum values of μ_0 and percent relative efficiencies E_1^* and E_2^*. The findings are displayed in Table 2.

It is noticed from Table 2 that for $\rho_{yx} = 0.8$ and $\rho_{yz_2} = -0.70\,to\,-0.94$, the performance of the proposed estimator T^* is efficient than $\bar{y}_n$ while T^* is efficient than $\hat{\hat{Y}}$ for different values of ρ_{yz_2} from $-0.70\,to\,-0.76$. For $\rho_{yx} = 0.8$ and $\rho_{yz_2} = -0.72$, the efficiency of the proposed estimator T^* over $\bar{y}_n$ is maximum, after that the efficiency decreases with increase in the value of ρ_{yz_2}.

Further, it is noticed that for $\rho_{yx} = 0.9$, the efficiency of T^* over $\bar{y}_n$ and $\hat{\hat{Y}}$ behaves in the following manner

- Efficiency increases with the increase in the value of ρ_{yz_2} i.e $\rho_{yz_2} = -0.80$ to -0.84,
- Efficiency is maximum when $\rho_{yz_2} = -0.86$, and
- Efficiency decreases with the increase in the value of ρ_{yz_2} i.e. $\rho_{yz_2} = -0.86$ to -0.98.

10. Conclusions

This article deals with the problem of estimating the population mean of the study variable on current (second) occasion in two-occasion successive sampling under two situations i) when the auxiliary variable is positively correlated with the study variable and ii) when the auxiliary variable is negatively correlated with the study variable. Properties of the suggested estimators have been discussed and the conditions where the suggested estimators are optimum are also obtained. It is found that the suggested estimator in both cases has shown efficient results when there is high correlation between study and auxiliary variables. From the empirical results, it can be concluded that the proposed estimator is more rewarding in the estimation of the population mean of the study variable at the current occasion in two occasion successive sampling. Finally, our recommendation is to use the proposed estimator by the survey practitioners in practice.

References

[1] Biradar R. S. and Singh, H. P. (2001). Successive sampling using auxiliary information on both occasions. Calcutta Statistical Association Bulletin, 51, 243-251.

[2] Eckler, A. R.(1955). Rotation sampling. Ann. Math. Stat. 26: 664-685.

[3] Feng, S, and Zou, G. (1997). Sampling rotation method with auxiliary variable. Communication in Statistics – Theory and Methods, 26(6), 1497-1509.

[4] Jessen, R. (1942). Statistical investigation of a sample survey for obtaining form facts. Iowa Agricultural Experiment Station Road Bulletin no. 304, Ames, USA.

[5] Murthy, M. (1967). Sampling Theory and Methods. Statistical Publishing Society, Kolkata, India.

[6] Okafor, F. C., Arnab, R.(1987). Some strategies of two-stage sampling for estimating population ratio over two occasions. Austrian Journal of Statistics 29(2): 128-142.

[7] Patterson, H. D. (1950). Sampling on successive occasions with partial replacement of units", Journal of the Royal Statistical Society B, 12, 241-255.

[8] Rao, J. N. K., and Graham, J.E. (1964). Rotation design for sampling on repeated occasions", Journal of the American Statistical Association, 59, 492-502.

[9] Reddy, V. (1978). A general class of estimators in successive sampling. Sankhya C, 40, pp. 29–37.

[10] Singh, Hari P., Singh, Hausila P., and Singh, V.P. (1992). A generalized efficient class of estimators of population mean in two-phase and successive sampling. International Journal of Management Systems, 8(2), 173-183.

[11] Singh, H. P., Ruiz-Espejo, M.R. (2003). On linear regression and ratio–product estimation of a finite population mean. Journal of the Royal Statistical Society: Series D (The Statistician), 52, no. 1, pp. 59–67.

[12] Singh, H. P., and Vishwakarma, G. K. (2007a). Modified exponential ratio estimators for finite population mean in double sampling. Austrian Journal of Statistics, 36(3), 217-225.

[13] Singh, H. P., and Vishwakarma, G. K. (2007b). A general class of estimators in successive sampling", Metron, 65(2), 201-227.

[14] Singh, H. P., and Vishwakarma, G. K. (2009). A general procedure for estimating the population mean in successive sampling. Communication in Statistics – Theory and Methods, 38, 293-308.

[15] Singh, H. P., and Pal, S. (2017). A Modified Procedure for Estimating the Population Mean in Two-occasion Successive Samplings. Afr. Stat. 12 (3), 1347-1365.

[16] Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S. and Ashok C.(1984). Sampling theory of University Press.

İSTATİSTİK: JOURNAL OF THE TURKISH STATISTICAL ASSOCIATION

İSTATİSTİK

Vol. 13, No. 2, July 2021, pp. 52–73

ISSN 1300-4077 | 21 | 2 | 52 | 73

TESTS OF NORMALITY BASED ON EDF STATISTICS USING PARTIALLY RANK ORDERED SET SAMPLING DESIGNS

Yusuf Can Sevil

The Graduate School of Natural and Applied Sciences,
Dokuz Eylül University,
35160, İzmir, Turkey

Tuğba Özkal Yıldız *

Department of Statistics,
Dokuz Eylül University,
35160, İzmir, Turkey

Abstract: In this study, we considered five goodness-of-fit (GOF) tests based on empirical distribution function (EDF) which are Kolmogorov-Smirnov (D), Kuiper (V), Cramér-von Mises (W^2), Watson (U^2), Anderson-Darling (A^2) tests to assess normality. Thus, we first suggested the EDFs based partially rank ordered set (PROS) sampling designs which are known as PROS Level-0, Level-1 and Level-2 sampling designs. Then, we discussed the relative efficiencies of the suggested EDFs w.r.t their counterparts of simple random sampling (SRS) and ranked set sampling (RSS). The main idea of this study is to compare the performances of the five different GOF tests based on PROS sampling designs with the GOF tests based on SRS and RSS. For this purpose, we investigated the power of the suggested GOF tests based on PROS sampling designs by performing simulations. In addition to the simulations, a real data set is considered to illustrate the GOF tests based on PROS sampling designs. According to the results, it can be seen that the EDFs based on PROS sampling designs are more efficient than the EDFs based on SRS and RSS. Also, it is clearly appeared that the GOF tests, Kolmogorov-Smirnov (D), Cramér-von Mises (W^2) and Anderson-Darling (A^2), based on PROS sampling designs has the best power performance.

Key words: Ranked set sampling; Partially rank ordered set; Sampling designs; Empirical distribution function; Goodness-of-fit tests; Type I error; Power of test

1. Introduction

In scientific researches, basic statistical principles play vital roles and one of these principles is to ensure experimental data for making valid judgements on the question(s) of interest under investigation. To obtain the experimental data, sampling methods are used in researches across all of the sciences-agricultural, biological, ecological, engineering, medical, physical, and social. The most fundamental of these sampling methods is simple random sampling (SRS). Via SRS, a single random sample of size n, $X_1, \cdots, X_n$, is selected from a population of interest. To make valid statistical inference, the sample should be representative of the population characteristic, say mean, median, etc., of interest. However, in practice there is no guarantee that the single random sample is truly representative of the entire population. In this case, sample size is usually increased by researcher. However, if sample size is increased, it may not be appropriate in terms of cost or time.

To deal with the problem, McIntyre [1] introduced ranked set sampling (RSS) as an advantageous alternative to SRS. McIntyre [1] benefited from RSS for seeking to estimate mean of the yield of pasture in Australia, effectively. McIntyre [1] described RSS procedure as follows: First, a set of size k is drawn by using SRS from population and the sample observations are ranked by visual

* Corresponding author. E-mail address: tugba.ozkal@deu.edu.tr

inspection. Then, the first smallest observation is identified and taken for full measurement. The other observations are discarded. Next, another set of size k is drawn by using SRS. The second smallest observation is measured and the other observations are discarded. This process is repeated until the kth smallest observation is measured in the kth set, so a cycle is completed. Then, the cycle repeats l times and ranked set sample of size $n = lk$ is obtained. When the ranking is perfect, population is divided into k homogeneous groups by RSS. Thus, sample units can be obtained in each groups and the ranked set sample is to be more representative of the population characteristic than a simple random sample. McIntyre [1] showed that mean of the measured sample observations is an unbiased estimator of the population mean regardless of any error in ranking process. Takahasi and Wakimoto [2] established the first theoretical result about RSS. It is showed that mean of ranked set sample is unbiased estimator. Also, they showed that the variance of the estimator is always smaller than the variance of the mean of a simple random sample under perfect ranking. Dell and Clutter [3] evaluated the effect of ranking errors on RSS. For the other basic studies on RSS, see [4], [5] and [6].

Deshpande et al. [7] developed three sampling designs for RSS. By using the sampling designs, ranked set sampling can be obtained in different ways depending on the replacement policy. The sampling designs have similar behaviours for infinite population, but they perform differently for finite population. To obtain Level-0 sampling design, units in the set are selected without replacement, but all units in the set are replaced back into the population before the following set is selected. In the sampling design, a population unit may be selected more than once both in the ranking process and in the final sample. Also, we need $k \leq N$, where k is the set size and N is the population size, for Level-0. If the measured unit in the set is not replaced back into the population, Level-1 sampling design is obtained. In the Level-1 sampling design, a population unit may be appeared in the ranking process, but may not be appeared in the final sample. To obtain Level-1 sampling design, we need $N - lk \leq N$, where l is the number of cycles. On the other hand, Level-2 sampling design is obtained if none of the units in the set are replaced back into the population before the following set is selected. In the Level-2 sampling design, a unit in the population is not appeared more than once neither in the ranking process nor in the final sample. Also, we need $N - lk^2 \leq N$ in the Level-2 sampling design. In the literature, research in RSS draw considerable attention in finite population setting as well, e.g. [8]-[14].

Having knowledge about the population distribution is required to apply accurate tests in statistics. Goodness-of-fit (GOF) tests have been used in scientific researches to check distributional assumptions. In literature, the estimation of cumulative distribution function (CDF) with various settings of the RSS has been studied by many authors. Stokes and Sager [15] suggested an unbiased estimator for the population distribution function based on the EDF of RSS. Under the assumption of perfect ranking, they considered the performance of Kolmogorov-Smirnov statistic by using the EDF. It is seen that the RSS can result in a substantial decrease in the width of the simultaneous confidence band for the CDF in this study. Frey and Wang [16] suggested alternative GOF tests that are sensitive both to imperfect rankings and to departures from parametric family by using the RSS. Nazari et al. [17] studied empirical density and distribution function estimators based on PROS. Sevil and Yildiz [18] examined the power of Kolmogorov-Smirnov test for standard normal and inverse Gaussian distribution. In the RSS process, they benefited from auxiliary informations, Level-2 sampling design and PROS. Yildiz and Sevil [19] proposed GOF tests based on EDFs for Level-0, Level-1 and Level-2 in RSS. Also, Yildiz and Sevil [20] and Sevil and Yildiz [21] investigated relative efficiencies of EDFs based on sampling designs in RSS w.r.t the EDF based on SRS.

GOF tests based on the EDFs indicate whether the sample data is appropriate or not to any specific distribution function. This is vital for parametric assumption. Even if Yildiz and Sevil [19] have proposed GOF tests based on Level-0, Level-1 and Level-2, there is still a gap in estimating the distribution function in finite population. Therefore, we suggested EDFs based on Level-0, Level-1 and Level-2 PROS sampling designs as better alternatives against the EDF estimators of Yildiz and Sevil [19, 20]. Then, we considered GOF tests based on EDFs by using the PROS sampling designs. As the main purpose of this study, we compared the GOF tests based on the suggested EDFs with counterparts of SRS and RSS. This study is organized as follows. In Section two, we introduce Level-0, Level-1 and Level-2 PROS sampling designs. Then, the EDFs estimators based on PROS sampling designs are suggested in Section three. Also, the relative efficiencies of the suggested EDFs w.r.t EDFs based on SRS and RSS are investigated by setting a simulation. On the other hand, GOF tests, Kolmogorov-Smirnov (D), Kuiper (V), Cramér-von Mises (W^2), Watson (U^2), Anderson-Darling (A^2), based on PROS sampling designs are given in Section four.We investigate the proposed GOF tests based on EDFs in terms of their type I and powers. The powers of the proposed GOF tests are compared with the powers of the GOF tests based on SRS and RSS in this section. In Section five, the proposed GOF tests are applied to a percentage of body fat data. We test the sample data which is obtained by using PROS Level-2 sampling design is appropriate or not to normal distribution with mean μ and variance σ^2. Some concluding remarks are given in Section six.

2. PROS Sampling Designs

In RSS, rankers aim to rank the all units in the sets accurately even with low confidence. However, in practice, the units in the set are ranked inaccurately if the rankers have low confidence. Also, if there are two or more tied units in selected set, this case makes it difficult to rank the units in the set. These situations reduce the efficiency of RSS. PROS is suggested by Ozturk [22] as a solution to these situations. Nonparametric inference is developed for one and two sample problems in PROS by Ozturk [23, 24]. Ozturk [25] used PROS in a data including multiple auxiliary variables. For finite population, PROS sampling designs are proposed by Ozturk and Jozani [26]. In this section, we introduce the PROS sampling designs which are known as PROS Level-0, PROS Level-1 and PROS Level-2. Note that, we give balanced PROS sampling designs procedures in the section.

First, $X_1, \cdots, X_k$ are selected without replacement from a finite population. Then, these units are assigned into H mutually exclusive subsets and these subsets are denoted by d_v, $v = 1, \cdots, H$. So each subsets includes s units where $s = k/H$. If ranking procedure is performed perfectly, then it is assumed that all units in the subset d_v have smaller ranks than all units in the subset $d_{v'}$, $v < v'$. After that, a unit is selected at random from d_1 for full measurement, $X_{[d_1]1}$. If the all k units in the set are replaced back into the population before the following set is selected, the PROS Level-0 sampling design is obtained. For PROS Level-1 sampling design, the measured unit is not replaced back into the population, but $k-1$ units are replaced back into the population before the following set is selected. If none of the units in the sets are replaced back into the population, PROS Level-2 sampling design is obtained. For each PROS sampling designs, the procedure is repeated H times and one cycle is completed. Then, this procedure is repeated l cycles to obtain PROS sampling designs as following matrix.

$$\begin{pmatrix} X_{[d_1]1} & X_{[d_1]2} & \cdots & X_{[d_1]l} \\ X_{[d_2]1} & X_{[d_2]2} & \cdots & X_{[d_2]l} \\ \vdots & \vdots & & \vdots \\ X_{[d_H]1} & X_{[d_H]2} & \cdots & X_{[d_H]l} \end{pmatrix}$$

In this matrix, $X_{[d_v]i}$ is obtained from the vth subset of vth set in ith cycle, $v = 1, \cdots, H$ and $i = 1, \cdots, l$. Let us illustrate the procedures of PROS sampling designs when $k = 9$, $H = 3$ and $l = 2$ in the following table. In this table, sets are denoted by S_v, $v = 1, 2, 3$. In each row, a unit is selected from the bold faced subset, d_v. Also, it can be said that we need $9 < N$, $N - 2*3 < N$ and $N - 2*3*9 < N$ for PROS Level-0, Level-1 and Level-2 sampling designs, respectively.

TABLE 1. PROS sampling designs when $k=9$, $H=3$ and $l=2$, Hatefi et al. [27]

l	Set	Subsets	Observations
1	S_1	$\{\boldsymbol{d_1},d_2,d_3\}=\{\{\mathbf{1},\mathbf{2},\mathbf{3}\},\{4,5,6\},\{7,8,9\}\}$	$X_{[d_1]1}$
	S_2	$\{d_1,\boldsymbol{d_2},d_3\}=\{\{1,2,3\},\{\mathbf{4},\mathbf{5},\mathbf{6}\},\{7,8,9\}\}$	$X_{[d_2]1}$
	S_3	$\{d_1,d_2,\boldsymbol{d_3}\}=\{\{1,2,3\},\{4,5,6\},\{\mathbf{7},\mathbf{8},\mathbf{9}\}\}$	$X_{[d_3]1}$
2	S_1	$\{\boldsymbol{d_1},d_2,d_3\}=\{\{\mathbf{1},\mathbf{2},\mathbf{3}\},\{4,5,6\},\{7,8,9\}\}$	$X_{[d_2]2}$
	S_2	$\{d_1,\boldsymbol{d_2},d_3\}=\{\{1,2,3\},\{\mathbf{4},\mathbf{5},\mathbf{6}\},\{7,8,9\}\}$	$X_{[d_2]2}$
	S_3	$\{d_1,d_2,\boldsymbol{d_3}\}=\{\{1,2,3\},\{4,5,6\},\{\mathbf{7},\mathbf{8},\mathbf{9}\}\}$	$X_{[d_3]2}$

To show the connection between $f_{[d_v]}(x)$ and $f_{(r:k)}(x)$ where $f_{[d_v]}(x)$ and $f_{(r:k)}(x)$ are the density function of $X_{[d_v]}$ and the *rth* order statistics, following lemma is given by Nazari et al. [17] for PROS. This lemma is valid for PROS sampling designs as well. When the $\alpha_{d_v,d_{v'}}=1$ for $v=v'$ and otherwise 0, it means that the units in the set are assigned into the subsets, perfectly.

LEMMA 1. *When the units in the set are assigned into the subsets, imperfectly, it is assumed that all units in subset d_v may not be smaller than all units in subset $d_{v'}$, $v<v'$. Let $\alpha_{d_v,d_{v'}}$ is the misplacement probability of a unit from subset d_v into subset $d_{v'}$ with $\sum_{v'=1}^{H}\alpha_{d_v,d_{v'}}=\sum_{v=1}^{H}\alpha_{d_v,d_{v'}}=1$. Then, we have*

$$
\begin{aligned}
f_{[d_v]}^{(t)}(x)=&Hf(x)\sum_{v'=1}^{H}\sum_{u\in d_{v'}}\alpha_{d_v,d_{v'}}\binom{k-1}{u-1}F(x)^{u-1}\left(1-F(x)\right)^{k-u}\\
=&\frac{1}{s}\sum_{v'=1}^{H}\sum_{u\in d_{v'}}\alpha_{d_v,d_{v'}}f_{(u:k)}^{(t)}(x)
\end{aligned}
$$

and consequently

$f(x)=\frac{1}{H}\sum_{v=1}^{H}f_{[d_v]}^{(t)}(x)$ and $F(x)=\frac{1}{H}\sum_{v=1}^{H}F_{[d_v]}^{(t)}(x)$ where $F_{[d_v]}^{(t)}(x)$ is the CDF of $X_{[d_v]i}^{(t)}$, $i=1\cdots,l$ and $t=0,1,2$ for Level-0, Level-1 and Level-2 PROS sampling designs, $F_{[d_v]}^{(t)}(x)=\frac{1}{s}\sum_{v'=1}^{H}\sum_{u\in d_{v'}}\alpha_{d_v,d_{v'}}F_{(u:k)}^{(t)}$.

Then, the lemma reduces the following remark that is given by Nazari et al. [17].

REMARK 1. When the units in the set are assigned into the subsets, perfectly, it is assumed that all units in subset d_v have smaller ranks than all units in subset $d_{v'}$, $v<v'$. Then, we have

$$
\begin{aligned}
f_{[d_v]}^{(t)}(x)=&Hf(x)\sum_{u\in d_v}\binom{k-1}{u-1}F(x)^{u-1}\left(1-F(x)\right)^{k-u}\\
=&\frac{1}{s}\sum_{u\in d_v}f_{(u:k)}^{(t)}(x)=\frac{1}{s}\sum_{r=(v-1)s+1}^{vs}f_{(r:k)}^{(t)}(x)
\end{aligned}
$$

and

$$
F_{[d_v]}^{(t)}(x)=\frac{1}{s}\sum_{r=(v-1)s+1}^{vs}F_{(r:k)}^{(t)}.
$$

3. Emprical Distribution Functions

EDF is basically a cumulative distribution function (CDF). However, EDF models empirical data while CDF is a hypothetical model of a distribution. That means, EDF is used for making inference about entire distribution function. Let us give theoretical definition of EDF.

Definition 1. Let $x_1, x_2, \cdots, x_n$ be random sample and $\partial_n(B)$ be a number of observations $x_1, x_2, \cdots, x_n$ falling into B and $B = (-\infty, x]$, where $B \in$B, B is Borel σ-algebra. Then,

$$F_n(x) = \frac{\partial_n(B)}{n}, \, x \in \mathbb{R}$$

is called EDF of the sample $x_1, x_2, \cdots, x_n$.

In this section, we give EDFs based on PROS sampling designs. Before that, let us assume that a simple random sample of size n, $X_1, \cdots, X_n$ is selected from a specific population having CDF $F(x)$, then the EDF estimator ($\hat{F}(x)$) is defined as follows:

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x) \tag{3.1}$$

where $I(.)$ is indicator function. The EDF based on SRS is unbiased estimator of $F(x)$ for given x, with variance $V(\hat{F}(x)) = \frac{1}{n} F(x)(1 - F(x))$. Similarly, let us describe EDF estimators using sampling designs which are suggested by Yildiz and Sevil [19, 20]. If the RSS sample, $X^{(t)}_{[1:k]i}, \cdots, X^{(t)}_{[k:k]i}$ and $i = 1, \cdots, l$, is selected by using Level-t from $F(x)$, then the EDF estimator ($\hat{F}_{RSS_{L-t}}(x)$) is given in below

$$\hat{F}_{RSS_{L-t}}(x) = \frac{1}{lk} \sum_{i=1}^{l} \sum_{r=1}^{k} I(X^{(t)}_{[r:k]i} \leq x). \tag{3.2}$$

where $t = 0, 1, 2$ for Level-0, Level-1 and Level-2, respectively. Yildiz and Sevil [20] showed that $\hat{F}_{RSS_{L-t}}(x)$ is unbiased estimator for $F(x)$ with variance $V\left(\hat{F}_{RSS_{L-t}}(x)\right) = \frac{1}{lk^2} \sum_{r=1}^{k} F_{[r:k]}\left(1 - F_{[r:k]}\right)$. Also, Yildiz and Sevil [20] and Sevil and Yildiz [21] proved that $V\left(\hat{F}_{RSS_{L-t}}(x)\right) \leq V(\hat{F}(x))$ even ranking is imperfect.

Now, we describe the new EDF estimators based on PROS sampling designs. It is assumed that $\left\{X^{(t)}_{[d_1]1}, \cdots, X^{(t)}_{[d_H]l}\right\}$ is Level-t PROS sampling design, $t = 0, 1, 2$. Then, the EDF based on Level-t is given as follows:

$$\hat{F}_{PROS_{L-t}}(x) = \sum_{i=1}^{l} \sum_{v=1}^{H} I\left(X^{(t)}_{[d_v]i} \leq x\right). \tag{3.3}$$

The following theorem includes the basic and large sample properties of $\hat{F}_{PROS_{L-t}}(x)$.

Theorem 1. *Let $\hat{F}_{PROS_{L-t}}(x)$ be the EDF estimator for each sampling designs, where $t = 0, 1, 2$ and for a fixed $x \in \mathbb{R}$,*

i. $E[\hat{F}_{PROS_{L-t}}(x)] = F(x)$.

ii. $V\left(\hat{F}_{PROS_{L-t}}(x)\right) \leq V\left(\hat{F}(x)\right)$.

iii. $\hat{F}_{PROS_{L-t}}(x)$ *is a strong consistent estimator of* $F(x)$ *as* $l \to \infty$.

Proof. i. It is a result of Lemma 1. ii. According to Corollary 2 in Ozturk [22], we have

$$V(\hat{F}_{PROS_{L-t}}(x)) = V\left(\hat{F}(x)\right) - \frac{1}{l^2 H} \sum_{v=1}^{H} \left(F^{(t)}_{[d_v]} - F(x)\right)^2$$

Thus, Part ii. is proved. Note that $F^{(t)}_{[d_v]} = F(x)$ if the k units in the sets are assigned into the subsets, randomly. iii. It is can be proved by Strong Law of Large Numbers.

Note that it is proved that $V(\hat{F}_{PROS_{L-t}}(x)) \leq V\left(\hat{F}_{RSS_{L-t}}(x)\right)$ by setting simulation in this section. Another large sample property of $\hat{F}_{PROS_{L-t}}(x)$ is given by the following theorem. This theorem says that $\hat{F}_{PROS_{L-t}}(x)$ converges almost sure (a.s.) for $F(x)$.

THEOREM 2. *It is assumed that $X^{(t)}_{[d_v]i}$; $v=1\cdots,H$ $i=1\cdots,l$ and $t=0,1,2$ are selected from a population with its CDF $F(x)$. Then, it is described as a distance measure as follows:*

$$D_l = \sup_{x\in\mathbb{R}} |\hat{F}_{PROS_{L-t}}(x) - F(x)|.$$

It can be said that $D_l \overset{a.s.}{\to} 0$ as $l \to \infty$.

PROOF. We can write the following inequality by using the Lemma 1. Also, we know that $X^{(t)}_{[d_v]i}$ is independent and identically distributed.

$$D_l \le \frac{1}{H}\sum_{v=1}^{H} \sup_{x\in\mathbb{R}} |\hat{F}^{(t)}_{[d_v]}(x) - F^{(t)}_{[d_v]}(x)|$$

where the right-hand side of the inequality goes to zero by the Gilvenko-Cantelli for $\hat{F}^{(t)}_{[d_v]}(x) = \frac{1}{l}\sum_{i=1}^{l} I\left(X^{(t)}_{[d_v]i} \le x\right)$ as $l \to \infty$.

After the theoretical properties of the proposed EDF estimators, we investigate the relative efficiencies of $\hat{F}_{PROS_{L-t}}(x)$ w.r.t $\hat{F}(x)$ and $\hat{F}_{RSS_{L-t}}(x)$ by the simulation study. In this simulation, populations are generated by using g-and-h distribution. Because, using g-and-h distribution is a simple method to generate population data from a wide variety of distributions included extreme departures from normality in terms of skewness and kurtosis. g-and-h distribution function is given by following equation:

$$X = \frac{(\exp(gZ) - 1)\exp\left(\frac{hZ^2}{2}\right)}{g}. \tag{3.4}$$

where g is the skewness and h is the kurtosis. When $g=0$, g-and-h distribution reduces to

$$X = Z\exp\left(\frac{hZ^2}{2}\right). \tag{3.5}$$

We have taken the population size $N = 250$, since the real data includes 252 observations (in Chapter five). For RSS, set sizes are $k = \{3,5\}$. For PROS, set sizes are $k = \{6,10\}$. Also, the number of subsets are taken $H=3$ and $H=5$ for $k=6$ and $k=10$, respectively. Thus, the number of measured units for RSS is equal to the number of measured units for PROS in a cycle. In addition to set sizes, the number of cycles are taken as $l = \{1,2,3,4,5\}$. Also, ranking procedures in RSS and PROS is done by using the following ranking error model. This model was proposed by Dell and Clutter [3],

$$Y = \rho\left(\frac{X-\mu_x}{\sigma_x}\right) + \sqrt{1-\rho^2}\xi. \tag{3.6}$$

where Y is the auxiliary variable, ξ follows the standart normal distribution and independent from X and ρ is the magnitude of the correlation coefficient between X and Y. Here, the ranking quality is controlled by $\rho \in [-1,1]$. In the simulation study, it is assumed that $\rho = 1$ and $\rho = 0.25$ for perfect and imperfect ranking, respectively. Relative efficiencies (RE) of EDFs based on RSS and PROS w.r.t EDF based on SRS are obtained by using their integrated mean squared errors. As performance comparison criteria, Wang et al. [28] described IMSE of EDF using SRS as follow:

$$IMSE_{\hat{F}(x)} = \int_{-\infty}^{\infty} \left\{\hat{F}(x) - F(x)\right\}^2 dx = \int_0^1 \left\{\hat{F}(F^{-1}(p)) - p\right\}^2 dF^{-1}(p)$$

$$= \int_0^1 \left\{ \hat{F}(F^{-1}(p)) - p \right\}^2 \frac{1}{f(F^{-1}(p))} dp \tag{3.7}$$

$x = F^{-1}(p)$ where $p \in [0,1]$. This $IMSE_{\hat{F}(x)}$ can be calculated approximately by using composite trapezoidal rule,

$$IMSE_{\hat{F}(x)} = \frac{b-a}{2L} \left\{ \sum_{i=1}^{L} \left| F(q_i) - \hat{F}(q_i) \right| + \sum_{i=2}^{L-1} \left| F(q_i) - \hat{F}(q_i) \right| \right\} \tag{3.8}$$

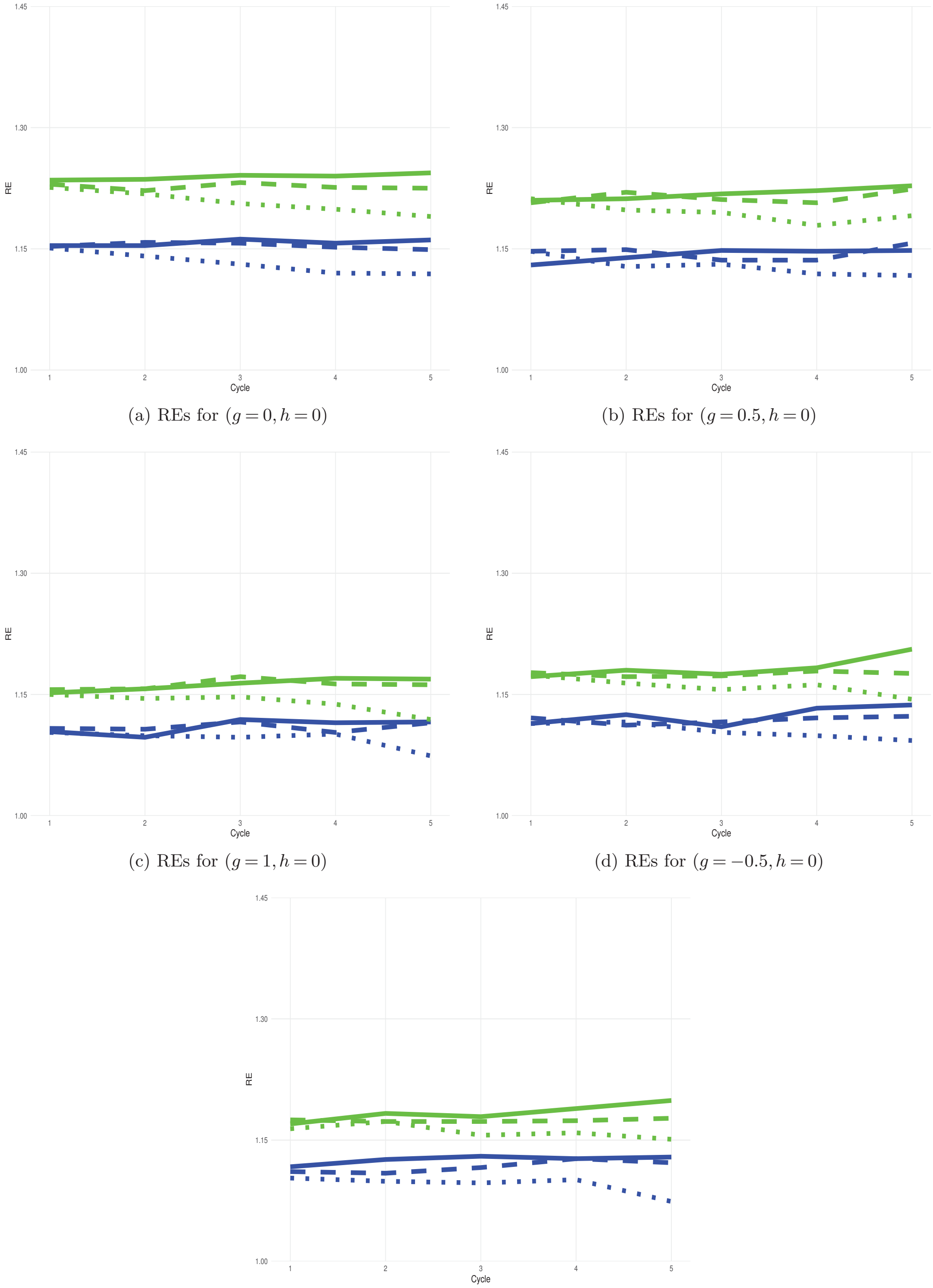
where b and a are upper and lower limits of integral, respectively. L is the number of cut points q_i, $L = \frac{b-a}{w}$ where w is the width of intervals. In the interval $[a,b]$, cut points q_i are obtained by $q_i = a + i(b-a)/L$, $i = 1, \cdots, L$. IMSEs based on sampling designs in RSS and PROS are given by

$$IMSE_{\hat{F}_{\Psi_{L-t}}(x)} = \frac{b-a}{2L} \left\{ \sum_{i=1}^{L} \left| F(q_i) - \hat{F}_{\Psi_{L-t}}(q_i) \right| + \sum_{i=2}^{L-1} \left| F(q_i) - \hat{F}_{\Psi_{L-t}}(q_i) \right| \right\} \tag{3.9}$$

where $\Psi =$ RSS, PROS and $t = 0,1,2$. $\hat{F}_{\Psi_{L-t}}(x)$ are calculated by using (3.2) and (3.3), where $x = \{q_1, q_2, \cdots, q_L\}$. In the simulation, the IMSEs of the EDFs based on SRS, RSS and PROS are calculated as taking $w = 0.01$. REs are computed as follow:

$$RE\left(\hat{F}(x), \hat{F}_{\Psi_{L-t}}(x)\right) = \frac{IMSE_{\hat{F}(x)}}{IMSE_{\hat{F}_{\Psi_{L-t}}(x)}} \tag{3.10}$$

If RE is larger than 1, it can be said that $\hat{F}_{L-t}(x)$ is more efficient than $\hat{F}(x)$. To obtain REs, $10,000$ samples are generated from SRS, RSS and PROS. The REs are illustrated by the Figures 1 and 2 for $\rho = 1$. According to all figures, we can say that the EDFs based on RSS and PROS are more efficient than the EDF based on SRS. Also, the REs of PROS are higher than the REs of RSS. Thus, it is clearly appeared that the proposed EDF estimators based on PROS sampling designs are more efficient than the EDFs based on sampling designs in RSS which are suggested by Yildiz and Sevil [19, 20]. Moreover, REs for symmetric distribution are higher than REs for skewed distributions. In addition to these, REs are almost the same for left-skewed and right-skewed distributions. While g gets closer to 1 (or -1), REs decrease. Among the sampling designs (Level-0, Level-1 and Level-2) in RSS and PROS, it is not obtained substantial difference when $k = 3$. The Figure 2 includes the REs for $k = 5$. First, we can say that the REs for $k = 5$ are larger than the REs for $k = 3$. Also, it can be appeared that EDFs based on Level-2 sampling design in RSS and PROS have higher efficient than EDFs based on the Level-0 and Level-1 in RSS and PROS. Obviously, it can also be seen that Level-2 in PROS is better than Level-2 in RSS, since PROS has higher performance than RSS according to all figures. On the other hand, the REs for $\rho = 0.25$ are not reported in this study, since REs varying around 1 are obtained for EDFs based on RSS and PROS under all distributions.



(a) REs for $(g = 0, h = 0)$

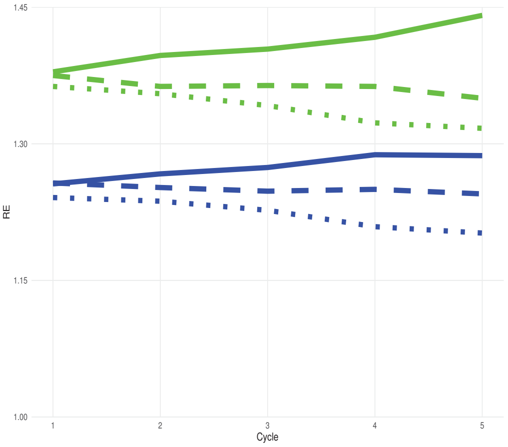
(b) REs for $(g = 0.5, h = 0)$

(c) REs for $(g = 1, h = 0)$

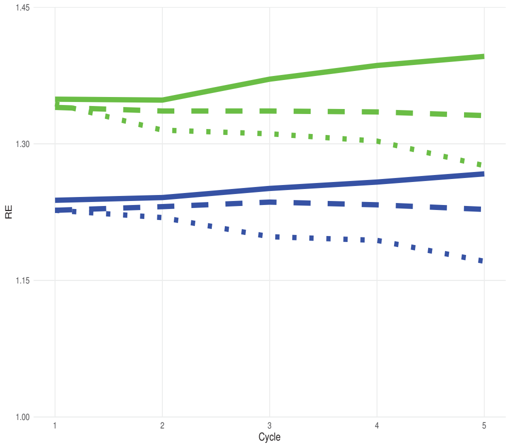
(d) REs for $(g = -0.5, h = 0)$

(e) REs for $(g = -1, h = 0)$

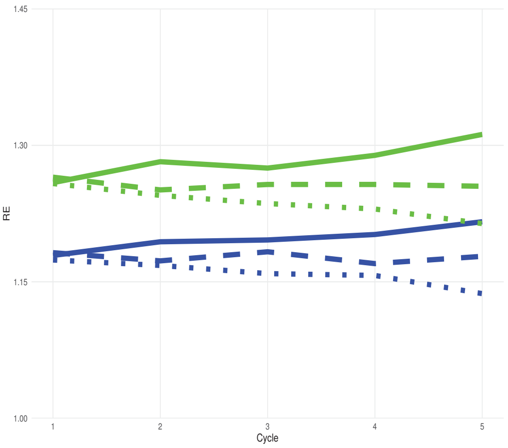
FIGURE 1. REs when $k = H = 3$ (blue: RSS, green: PROS, dotted: Level-0, dashed: Level-1 and solid: Level-2)



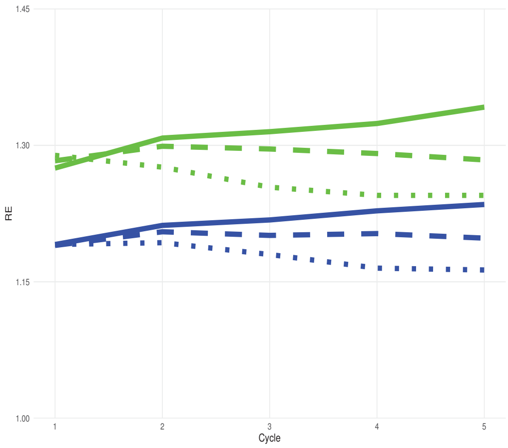
(a) REs for RSS $(g=0, h=0)$



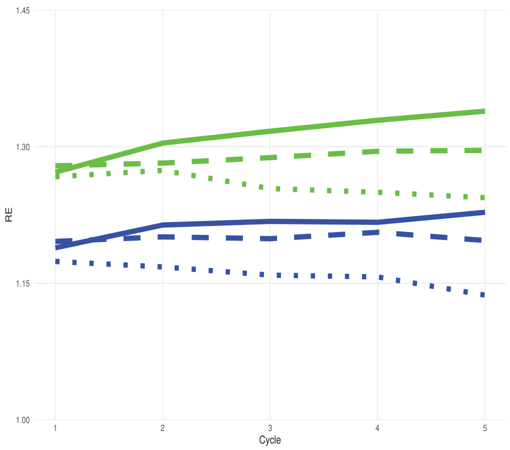
(b) REs for PROS $(g=0.5, h=0)$



(c) REs for PROS $(g=1, h=0)$



(d) REs for PROS $(g=-0.5, h=0)$



(e) REs for PROS $(g=-1, h=0)$

FIGURE 2. REs when $k=H=5$ (blue: RSS, green: PROS, dotted: Level-0, dashed: Level-1 and solid: Level-2)

4. Goodness-of Fit Tests

GOF test is a statistical hypothesis test to see how well sample data fit a distribution from a population with a specific distribution. In other words, these tests are used for making inference about the population distribution. Mostly, it is tested whether sampling observations are obtained from a population having normal distribution or not. In this situation, null hypothesis H_0 is simple hypothesis if we know parameters. On the other hand, H_0 is composite hypothesis when the parameters are not known. In this case, the parameters are estimated by using sampling observations. Also, alternative hypothesis H_1 is mostly composite hypothesis since we have little or no information about distribution of the data.

In this study, we investigate the powers of Kolmogorov-Smirnov (D), Kuiper (V), Cramér-von Mises (W^2), Watson (U^2), Anderson-Darling (A^2) tests under SRS, RSS and PROS. These are GOF tests based on EDF. These tests are divided into two different classes. Kolmogorov-Smirnov and Kuiper test statistics are in the supremum class. Cramér-von Mises, Watson and Anderson-Darling tests belong to the quadratic class.

These GOF tests based on SRS are intorduced in the Chapter 4 of the book "Goodness-of Fit Techiques", D'Agostino [29]. It is assumed that a random sample of size n, $X_1, \cdots, X_n$ is selected from a population and CDF of this population is $F(x)$. We test the null hypothesis $H_0: F(x) = F_0(x)$ against $H_1: F(x) \neq F_0(x)$. Then, the GOF tests are as follows:

- Kolmogorov-Smirnov test statistic:

$$\begin{aligned} D &= \sup_x \left| \hat{F}(x) - F(x) \right| = \max\left(D^+, D^-\right) \\ &= \max_j \left(\max\left\{ \frac{j}{n} - F_0(x_{(j)}) \right\}, \max\left\{ F_0(x_{(j)}) - \frac{j-1}{n} \right\} \right) \end{aligned} \tag{4.1}$$

- Kuiper test statistic:

$$\begin{aligned} V &= D^+ + D^- \\ &= \max\left\{ \frac{j}{n} - F_0(x_{(j)}) \right\} + \max\left\{ F_0(x_{(j)}) - \frac{j-1}{n} \right\} \end{aligned} \tag{4.2}$$

- Cramér-von Mises test statistic:

$$\begin{aligned} W^2 &= n \int_{-\infty}^{\infty} \left\{ \hat{F}(x) - F_0(x) \right\}^2 dF_0(x) \\ &= \sum_{j=1}^{n} \left\{ F_0(x_{(j)}) - \frac{2j-1}{2n} \right\}^2 + \frac{1}{12n} \end{aligned} \tag{4.3}$$

- Watson test statistic:

$$\begin{aligned} U^2 &= n \int_{-\infty}^{\infty} \left\{ \hat{F}(x) - F_0(x) - \int_{-\infty}^{\infty} \left[\hat{F}(x) - F_0(x) \right] dF_0(x) \right\}^2 dF_0(x) \\ &= W^2 - n \left(\frac{1}{n} \sum_{j=1}^{n} F_0(x_j) - 0.5 \right)^2 \end{aligned} \tag{4.4}$$

- Anderson-Darling test statistic:

$$\begin{aligned} A^2 &= n \int_{-\infty}^{\infty} \left\{ \hat{F}(x) - F_0(x) \right\}^2 \left[F_0(x) \left(1 - F_0(x)\right) \right]^{-1} dF_0(x) \\ &= -n - \frac{1}{n} \sum_{j=1}^{n} (2j-1) \left[\log F_0(x_{(j)}) + \log\left\{ 1 - F_0(x_{(n+1-j)}) \right\} \right] \end{aligned} \tag{4.5}$$

where $x_{(j)}$ is the jth order statistic of the random sample. Also, the GOF tests reject the null hypothesis of normality when the test statistics are larger than the their critical values which are the corresponding $100(1-\alpha)$ percentile of the null distribution of the test statistics. GOF tests based on Level-t in RSS were proposed by Yildiz and Sevil [19]. Now, we give the GOF tests based on Level-t in RSS and PROS. It is assumed that $\left\{X^{(t)}_{[r:k]i}, r=1,\cdots,k; i=1,\cdots,l\right\}$ and $\left\{X^{(t)}_{[d_v]i}, v=1,\cdots,H; i=1,\cdots,l\right\}$ are Level-t in RSS and PROS, respectively. Moreover, $\zeta^{RSS,t}_{(1)},\cdots,\zeta^{RSS,t}_{(n)}$ and $\zeta^{PROS,t}_{(1)},\cdots,\zeta^{PROS,t}_{(n)}$ are ordered Level-t sampling design in RSS and PROS, respectively. Then, the GOF tests based on Level-t in RSS and PROS are as follows:

- Kolmogorov-Smirnov test statistic:

$$\begin{aligned} D =& \sup_x \left|\hat{F}_{\Psi_{L-t}}(x) - F(x)\right| = \max\left(D^+, D^-\right) \\ =& \max_j \left(\max\left\{\frac{j}{n} - F_0(\zeta^{\Psi,t}_{(j)})\right\}, \max\left\{F_0(\zeta^{\Psi,t}_{(j)}) - \frac{j-1}{n}\right\}\right) \end{aligned} \tag{4.6}$$

- Kuiper test statistic:

$$\begin{aligned} V =& D^+ + D^- \\ =& \max\left\{\frac{j}{n} - F_0(\zeta^{\Psi,t}_{(j)})\right\} + \max\left\{F_0(\zeta^{\Psi,t}_{(j)}) - \frac{j-1}{n}\right\} \end{aligned} \tag{4.7}$$

- Cramér-von Mises test statistic:

$$\begin{aligned} W^2 =& n\int_{-\infty}^{\infty} \left\{\hat{F}_{\Psi_{L-t}}(x) - F_0(x)\right\}^2 dF_0(x) \\ =& \sum_{j=1}^{n}\left\{F_0(\zeta^{\Psi,t}_{(j)}) - \frac{2j-1}{2n}\right\}^2 + \frac{1}{12n} \end{aligned} \tag{4.8}$$

- Watson test statistic:

$$\begin{aligned} U^2 =& n\int_{-\infty}^{\infty} \left\{\hat{F}_{\Psi_{L-t}}(x) - F_0(x) - \int_{-\infty}^{\infty}\left[\hat{F}_{\Psi_{L-t}}(x) - F_0(x)\right] dF_0(x)\right\}^2 dF_0(x) \\ =& W^2 - n\left(\frac{1}{n}\sum_{j=1}^{n} F_0(\zeta^{\Psi,t}_{(j)}) - 0.5\right)^2 \end{aligned} \tag{4.9}$$

- Anderson-Darling test statistic:

$$\begin{aligned} A^2 =& n\int_{-\infty}^{\infty} \left\{\hat{F}_{\Psi_{L-t}}(x) - F_0(x)\right\}^2 \left[F_0(x)\left(1-F_0(x)\right)\right]^{-1} dF_0(x) \\ =& -n - \frac{1}{n}\sum_{j=1}^{n}(2j-1)\left[\log F_0(\zeta^{\Psi,t}_{(j)}) + \log\left\{1 - F_0(\zeta^{\Psi,t}_{(n+1-j)})\right\}\right] \end{aligned} \tag{4.10}$$

where $\Psi = RSS$ and $PROS$. The null hypothesis of normality is rejected when the test statistics based on Level-t in RSS and PROS are larger than $100(1-\alpha)$ percentile of the null distribution of the test statistics. Also, it is substantial to note that the distribution of the test statistics based on RSS and PROS still depend on the quality of ranking while they do not depend on the unknown parameters μ and σ^2. Therefore, it is not possible that the critical values for the GOF tests based on RSS and PROS are not obtained since the quality of ranking is not known in practice. Against the problem, we suggest that the critical values are always obtain under the assumption of perfect ranking. As a result of this suggestion, we will see that the type I errors under imperfect ranking are relatively larger than the type I errors under perfect ranking. For this reason, it is seen that powers of the GOF tests based on RSS and PROS under imperfect ranking are larger than powers of the GOF tests based on RSS and PROS under perfect ranking. Therefore, ranking error should still be minimized as much as possible for both RSS and PROS.

In simulation study, the null hypothesis $H_0 : F_0(x) = N(\mu, \sigma^2)$ is tested. For the GOF tests based on RSS and PROS, we first obtain critical values under H_0 by using the following algorithm.

(1) Select a sample using RSS and PROS from standard normal distribution.

(2) Calculate T^{RSS} and T^{PROS} by using the Equations (4.6)-(4.10).

(3) Repeat steps (1)-(2) to get $T_1^{RSS}, \cdots, T_{10,000}^{RSS}$ and $T_1^{PROS}, \cdots, T_{10,000}^{PROS}$.

(4) Approximate critical values, C_α^{RSS} and C_α^{PROS}, the $100(1-\alpha)$ percentage point of T^{RSS} and T^{PROS}, respectively.

In this algorithm, the all test statistics (4.6)-(4.9) are denoted by the notation T. Estimated critical values are given in Table 4. Then, the type I errors of GOF tests based on RSS and PROS sampling designs are examined for $\alpha = 0.05$. These type I errors are given by Table 2. In this table, it can be seen that the type I errors are almost equal to the nominal value ($\alpha = 0.05$) for perfect ranking ($\rho = 1$) while the type I errors are larger than the nominal value for imperfect ranking. This means that the GOF tests based on RSS and PROS hold the nominal alpha, $\alpha = 0.05$, for $\rho = 1$ while they do not hold the nominal alpha for $\rho = 0.25$.

The other algorithm is performed to calculate the power of GOF tests based on RSS and PROS. The steps of the algorithm are as following. Alternative distributions are obtained by g-and-h distribution, $g = 0.5, h = 0$ (right skewed), $g = 1, h = 0$ (right skewed), $g = -0.5, h = 0$ (left skewed) and $g = -1, h = 0$ (left skewed).

(1) Select a sample using RSS and PROS from an alternative distribution H_1.

(2) Calculate T^{RSS} and T^{PROS} by using the Equations (4.6)-(4.10).

(3) Repeat steps (1)-(2) to get $T_1^{RSS}, \cdots, T_{5,000}^{RSS}$ and $T_1^{PROS}, \cdots, T_{5,000}^{PROS}$.

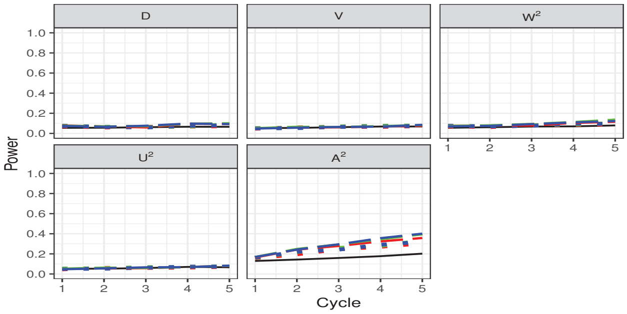
(4) $Power\, of\, T^{RSS} \approx \frac{1}{5,000} \sum_{t=1}^{5,000} I(T_t^{RSS} > C_{0.05}^{RSS})$ and

$Power\, of\, T^{PROS} \approx \frac{1}{5,000} \sum_{t=1}^{5,000} I(T_t^{PROS} > C_{0.05}^{PROS})$.

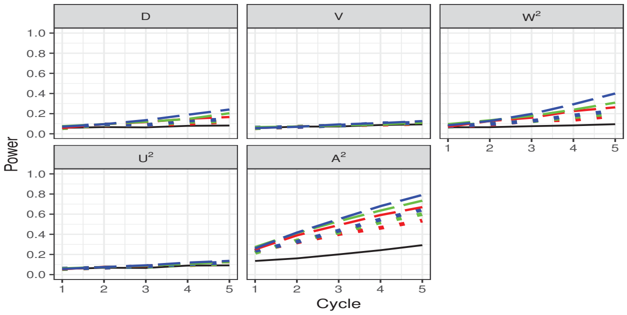
Figure 3-6 include estimated powers of GOF tests based on SRS, RSS and PROS. According to the Figures 3 and 4 (include the powers for $\rho = 1$), it is obviously seen that the best power performance among the GOF tests belong to Anderson-Darling GOF test (A^2) for SRS, RSS and PROS. Also, the highest powers are obtained for $k = H = 5$ and ($g = 1, h = 0$). On the other hand, the GOF tests based on PROS have higher power than the GOF tests based on SRS and RSS except for Kuiper (V) and Watson (U^2). For Kuiper (V) and Watson (U^2) tests, a difference is observed between among the GOF tests based on SRS, RSS and PROS only for $k = H = 5$ and ($g = 1, h = 0$). According to the Figures 5 and 6 (include the powers for $\rho = 0.25$), the powers of the GOF tests for RSS and PROS when $\rho = 0.25$ have higher than the powers when $\rho = 1$. This is not surprising result since this occurs as a result of increase in type I error. Among the right and left skewed distributions, the highest powers are obtained for $g = 1, h = 0$. The powers are almost equal among $g = 0.5, h = 0$, $g = -0.5, h = 0$ and $g = -1, h = 0$. The powers of GOF tests for $k = 5$ are higher than the powers of GOF tests for $k = 3$. Among the Level-0, Level-1 and Level-2 sampling designs in RSS and PROS, the GOF tests based on Level-2 have outperformance in most cases. Thus, it is shown that the GOF tests based on PROS Level-2 sampling design has the best power performance, especially, Kolmogorov-Smirnov (D), Cramér-von Mises (W^2) and Anderson-Darling(A^2). Kuiper (V) and Watson (U^2) test statistic have the lowest powers among all the GOF tests.

Table 2. Type I errors of GOF tests based on RSS and PROS sampling designs for $\alpha = 0.05$

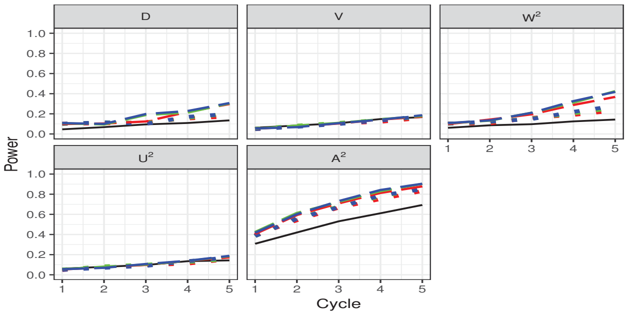
				D		V		W^2		U^2		A^2	
Methods	Designs	k	l	$\rho=0.25$	$\rho=1$	$\rho=0.25$	$\rho=1$	$\rho=0.25$	$\rho=1$	$\rho=0.25$	$\rho=1$	$\rho=0.25$	$\rho=1$
RSS	Level-0	3	1	0.164	0.052	0.095	0.048	0.174	0.051	0.096	0.048	0.151	0.053
			2	0.132	0.045	0.088	0.051	0.145	0.045	0.090	0.053	0.137	0.047
			3	0.146	0.051	0.090	0.052	0.163	0.051	0.091	0.049	0.153	0.050
			4	0.136	0.048	0.082	0.043	0.161	0.048	0.087	0.046	0.151	0.046
			5	0.133	0.051	0.084	0.047	0.154	0.050	0.090	0.050	0.155	0.051
		5	1	0.243	0.052	0.142	0.046	0.278	0.054	0.148	0.049	0.259	0.057
			2	0.225	0.050	0.127	0.048	0.256	0.049	0.144	0.050	0.250	0.049
			3	0.216	0.048	0.131	0.051	0.256	0.053	0.143	0.048	0.249	0.051
			4	0.200	0.046	0.125	0.051	0.243	0.047	0.136	0.051	0.234	0.045
			5	0.206	0.044	0.116	0.044	0.252	0.046	0.138	0.047	0.244	0.046
	Level-1	3	1	0.157	0.050	0.098	0.055	0.170	0.049	0.098	0.055	0.148	0.053
			2	0.145	0.050	0.087	0.055	0.158	0.047	0.096	0.057	0.155	0.049
			3	0.141	0.050	0.089	0.053	0.160	0.050	0.092	0.051	0.160	0.057
			4	0.138	0.047	0.088	0.051	0.160	0.050	0.092	0.052	0.151	0.050
			5	0.136	0.056	0.083	0.052	0.155	0.052	0.092	0.049	0.157	0.053
		5	1	0.228	0.047	0.140	0.049	0.256	0.047	0.149	0.047	0.239	0.041
			2	0.224	0.051	0.131	0.054	0.256	0.053	0.140	0.052	0.246	0.052
			3	0.211	0.049	0.125	0.051	0.248	0.050	0.140	0.052	0.244	0.051
			4	0.213	0.053	0.124	0.050	0.249	0.050	0.136	0.051	0.247	0.051
			5	0.199	0.048	0.129	0.051	0.245	0.050	0.141	0.051	0.244	0.049
	Level-2	3	1	0.167	0.047	0.095	0.046	0.183	0.049	0.095	0.046	0.154	0.047
			2	0.141	0.047	0.088	0.046	0.164	0.049	0.090	0.048	0.157	0.052
			3	0.133	0.046	0.081	0.050	0.155	0.046	0.079	0.050	0.144	0.046
			4	0.145	0.049	0.090	0.051	0.164	0.051	0.095	0.053	0.156	0.051
			5	0.139	0.048	0.089	0.052	0.166	0.050	0.099	0.054	0.159	0.047
		5	1	0.249	0.051	0.155	0.053	0.277	0.050	0.157	0.052	0.260	0.054
			2	0.235	0.051	0.131	0.047	0.268	0.049	0.144	0.049	0.256	0.045
			3	0.226	0.052	0.129	0.051	0.268	0.050	0.139	0.048	0.258	0.049
			4	0.236	0.052	0.134	0.054	0.285	0.052	0.145	0.050	0.277	0.048
			5	0.245	0.048	0.132	0.051	0.300	0.047	0.156	0.055	0.291	0.047
PROS	Level-0	3	1	0.256	0.054	0.160	0.052	0.267	0.051	0.161	0.052	0.204	0.049
			2	0.225	0.045	0.129	0.048	0.254	0.048	0.143	0.050	0.225	0.047
			3	0.203	0.047	0.114	0.048	0.241	0.047	0.125	0.047	0.219	0.050
			4	0.219	0.052	0.121	0.049	0.245	0.049	0.132	0.050	0.218	0.048
			5	0.229	0.057	0.119	0.052	0.252	0.054	0.134	0.054	0.230	0.052
		5	1	0.362	0.049	0.237	0.050	0.397	0.045	0.248	0.049	0.348	0.045
			2	0.343	0.050	0.215	0.052	0.402	0.048	0.247	0.053	0.367	0.049
			3	0.326	0.051	0.195	0.050	0.383	0.053	0.228	0.052	0.362	0.052
			4	0.343	0.054	0.209	0.055	0.401	0.052	0.246	0.056	0.383	0.055
			5	0.322	0.052	0.195	0.053	0.388	0.053	0.229	0.052	0.374	0.050
	Level-1	3	1	0.251	0.051	0.159	0.053	0.268	0.052	0.159	0.054	0.207	0.051
			2	0.220	0.053	0.130	0.050	0.243	0.050	0.137	0.052	0.225	0.052
			3	0.220	0.053	0.120	0.048	0.248	0.053	0.129	0.046	0.225	0.053
			4	0.211	0.051	0.122	0.052	0.247	0.053	0.125	0.047	0.232	0.051
			5	0.219	0.052	0.116	0.048	0.257	0.052	0.131	0.050	0.234	0.050
		5	1	0.386	0.054	0.250	0.054	0.432	0.058	0.261	0.054	0.375	0.053
			2	0.340	0.051	0.210	0.047	0.398	0.052	0.239	0.050	0.371	0.055
			3	0.336	0.057	0.201	0.052	0.390	0.056	0.230	0.052	0.368	0.055
			4	0.315	0.049	0.188	0.046	0.385	0.048	0.228	0.049	0.375	0.049
			5	0.317	0.054	0.185	0.047	0.374	0.053	0.223	0.050	0.363	0.051
	Level-2	3	1	0.257	0.052	0.147	0.052	0.267	0.047	0.146	0.052	0.211	0.049
			2	0.228	0.050	0.124	0.050	0.249	0.047	0.128	0.048	0.221	0.049
			3	0.232	0.050	0.129	0.047	0.256	0.046	0.139	0.047	0.226	0.043
			4	0.234	0.051	0.113	0.041	0.261	0.051	0.123	0.044	0.240	0.053
			5	0.239	0.047	0.127	0.051	0.262	0.046	0.137	0.048	0.242	0.047
		5	1	0.381	0.055	0.237	0.052	0.421	0.055	0.256	0.052	0.367	0.052
			2	0.360	0.046	0.213	0.049	0.424	0.045	0.239	0.050	0.397	0.048
			3	0.368	0.046	0.220	0.051	0.445	0.048	0.250	0.046	0.420	0.051
			4	0.388	0.048	0.220	0.047	0.461	0.045	0.260	0.049	0.431	0.046
			5	0.386	0.043	0.217	0.046	0.473	0.045	0.248	0.045	0.448	0.050



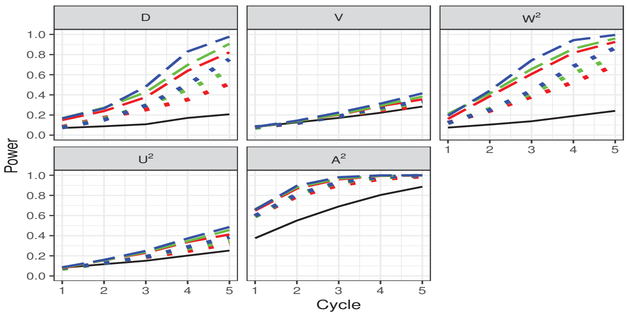
(a) For $k = H = 3$ and $(g = 0.5, h = 0)$



(b) For $k = H = 5$ and $(g = 0.5, h = 0)$

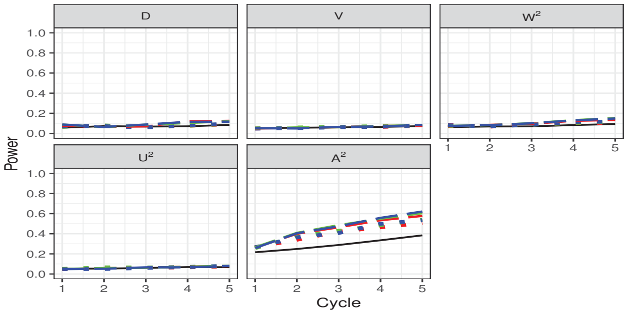


(c) For $k = H = 3$ and $(g = 1, h = 0)$

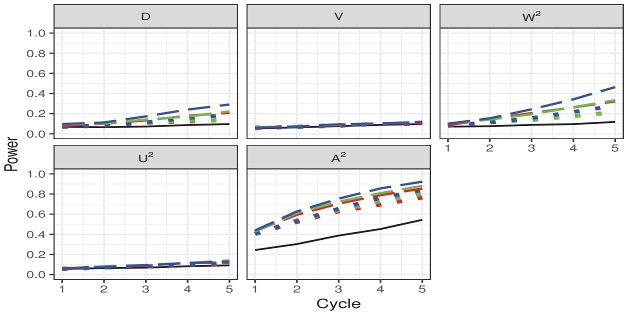


(d) For $k = H = 5$ and $(g = 1, h = 0)$

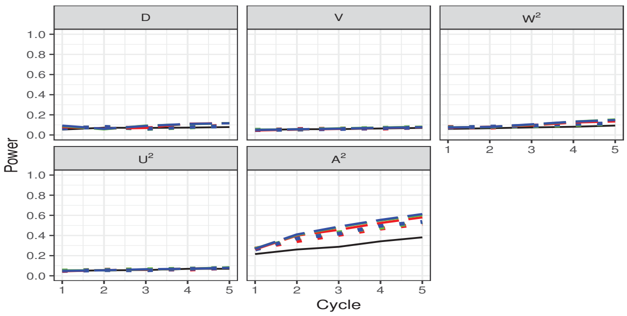
FIGURE 3. The power of GOF tests at $\alpha = 0.05$ for $\rho = 1$, solid: SRS, dotted: RSS and longdash: PROS (For RSS and PROS, red: Level-0, green: Level-1 and blue: Level-2)



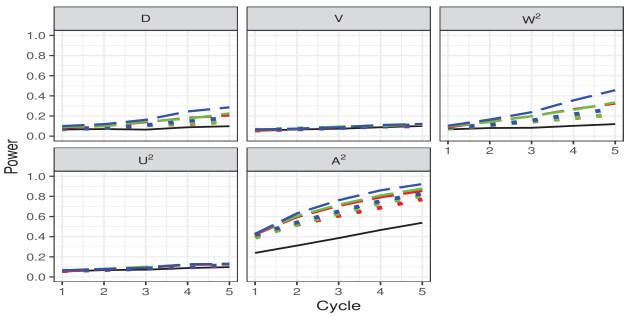
(a) For $k = H = 3$ and $(g = -0.5, h = 0)$



(b) For $k = H = 5$ and $(g = -0.5, h = 0)$

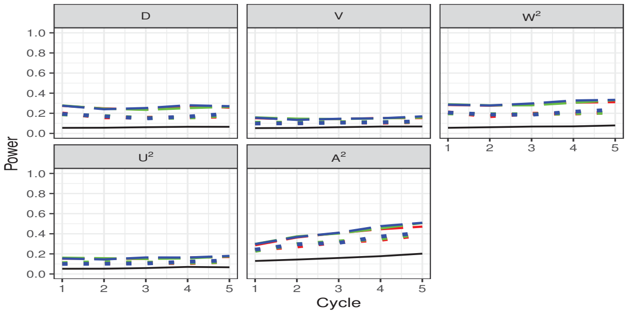


(c) For $k = H = 3$ and $(g = -1, h = 0)$

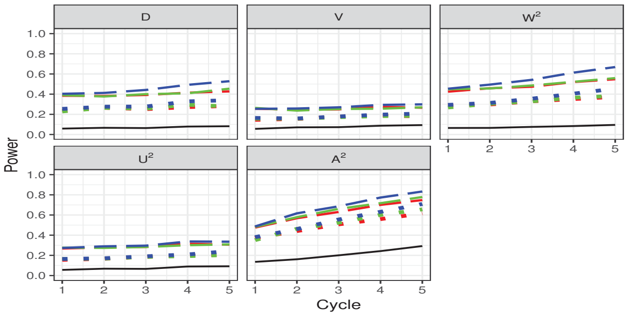


(d) For $k = H = 5$ and $(g = -1, h = 0)$

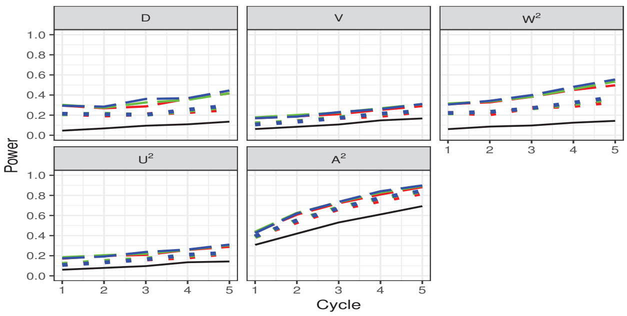
FIGURE 4. The power of GOF tests at $\alpha = 0.05$ for $\rho = 1$, solid: SRS, dotted: RSS and longdash: PROS (For RSS and PROS, red: Level-0, green: Level-1 and blue: Level-2)



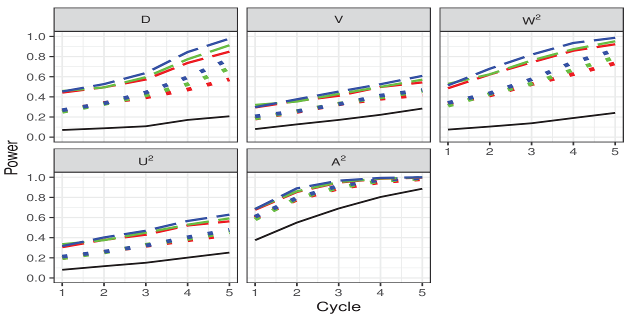
(a) For $k = H = 3$ and $(g = 0.5, h = 0)$



(b) For $k = H = 5$ and $(g = 0.5, h = 0)$

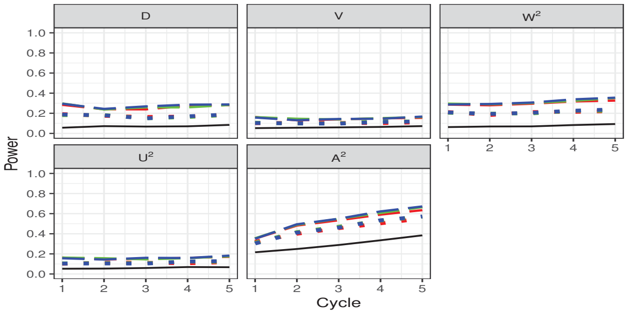


(c) For $k = H = 3$ and $(g = 1, h = 0)$

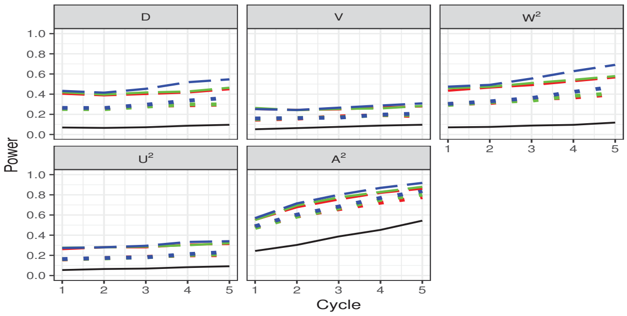


(d) For $k = H = 5$ and $(g = 1, h = 0)$

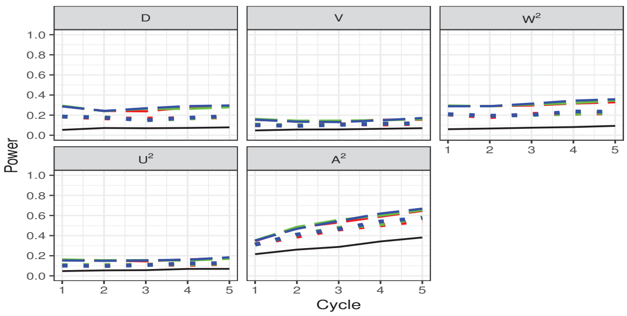
FIGURE 5. The power of GOF tests at $\alpha = 0.05$ for $\rho = 0.25$, solid: SRS, dotted: RSS and longdash: PROS (For RSS and PROS, red: Level-0, green: Level-1 and blue: Level-2)



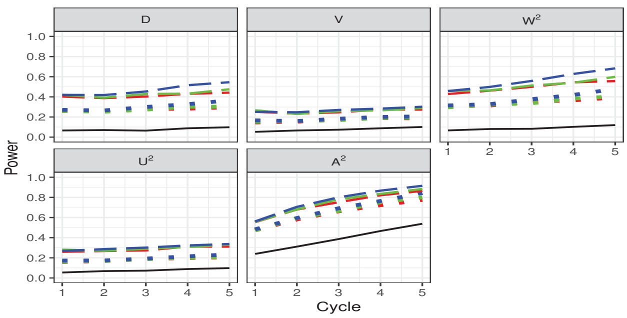
(a) For $k = H = 3$ and $(g = -0.5, h = 0)$



(b) For $k = H = 5$ and $(g = -0.5, h = 0)$



(c) For $k = H = 3$ and $(g = -1, h = 0)$



(d) For $k = H = 5$ and $(g = -1, h = 0)$

FIGURE 6. The power of GOF tests at $\alpha = 0.05$ for $\rho = 0.25$, solid: SRS, dotted: RSS and longdash: PROS (For RSS and PROS, red: Level-0, green: Level-1 and blue: Level-2)

5. Real Data Example

In this section, an illustrative example was considered using body fat data. This data set "http://lib.stat.cmu.edu/datasets/bodyfat" for 252 men collected by Penrose et al. [30] Suppose the set of 252 men constitutes a hypothetical population. It is made up of 15 measured variables on 252 men. Variables in the data are density, percentage of body fat (PBF), age, weight, height and 10 circumferences: neck, chest, abdominal, hip, thigh, knee, ankle, biceps, forearm and wrist. The body fat percentage of a human or other living being is the total mass of fat divided by total body mass. It is determined by underwater weighing and can be estimated using Equation 5.1.

$$\begin{aligned} &D = 1/[(A/a)+(B/b)] \\ &B = (1/D)[ab/(a-b)] - [b/(a-b)] \\ &PBF = 100 \times B, \end{aligned} \tag{5.1}$$

where $D =$ body density, $W =$ body weight, $A =$ proportion of lean tissue, $B =$ proportion of fat tissue $(A+B=1)$, $a =$ density of lean tissue and $b =$ density of fat tissue.

Our target parameter is the distribution function of percentage of body fat. The interested variable, X, has normal distribution with parameters $\mu = 19.15$ and $\sigma^2 = 70.03$. We used ages (Y)

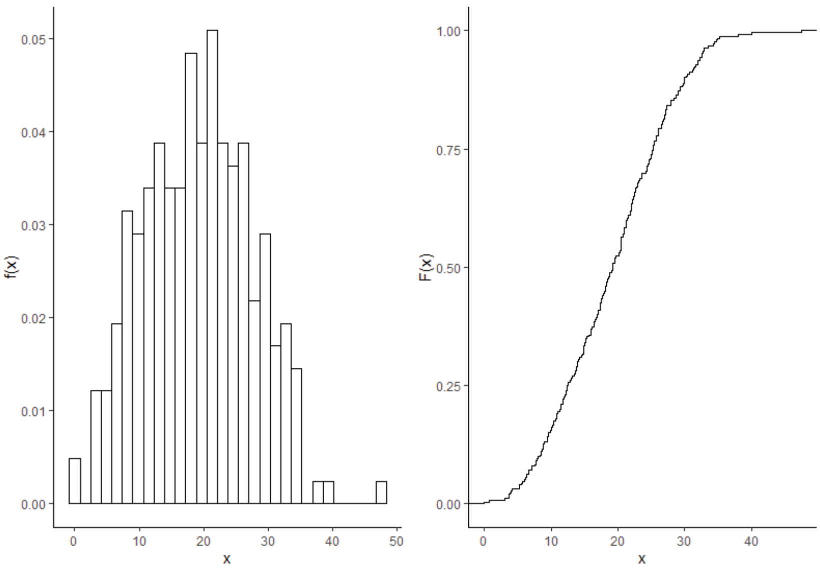


FIGURE 7. The PDF (left) and CDF (right) of the percentage of body fat

of the 252 men as auxiliary variable in ranking process since $cor(X,Y) = 0.813$. To obtain PROS Level-2 sampling design of size $n = 25$, we take set size $k = 10$, the number of subsets $H = 5$ and the number of cycles $l = 5$. The PROS Level-2 sampling procedure is illustrated in the Table 3. In this table, the set of size $k = 10$ is selected and divided into the $H = 5$ mutually exclusive subsets of size $s = 2$ in each row. Then, an observation is selected at random from the bold subsets. Then, this observation is measured. Although 250 of 252 men are used in ranking processs, only 25 men's percentage of body fat are measured. Based on the PROS Level-2 sampling design, the null hypothesis $H_0 : F_0(x) = N(\mu = 19.15, \sigma^2 = 70.03)$ is tested by using the all test statistics. Obtained test statistics are $D = 0.167$, $V = 0.245$, $W^2 = 0.099$, $U^2 = 0.078$ and $A^2 = 0.57$. According to the test statistics, the null hypothesis is not rejected at $\alpha = 0.05$. Thus, we can say that the percentage of body fats of 25 men come from normal distribution with parameters $\mu = 19.15$ and $\sigma^2 = 70.03$.

Table 3. PROS Level-2 sampling design procedure

l	Set	Subsets	Measured Observations
1	S_1	$\{\boldsymbol{d_1}, d_2, d_3, d_4, d_5\} = \{\{\mathbf{1,2}\}, \{3,4\}, \{5,6\}, \{7,8\}, \{9,10\}\}$	$X_{[d_1]1} = 8.5$
	S_2	$\{d_1, \boldsymbol{d_2}, d_3, d_4, d_5\} = \{\{1,2\}, \{\mathbf{3,4}\}, \{5,6\}, \{7,8\}, \{9,10\}\}$	$X_{[d_2]1} = 13.8$
	S_3	$\{d_1, d_2, \boldsymbol{d_3}, d_4, d_5\} = \{\{1,2\}, \{3,4\}, \{\mathbf{5,6}\}, \{7,8\}, \{9,10\}\}$	$X_{[d_3]1} = 24.7$
	S_4	$\{d_1, d_2, d_3, \boldsymbol{d_4}, d_5\} = \{\{1,2\}, \{3,4\}, \{5,6\}, \{\mathbf{7,8}\}, \{9,10\}\}$	$X_{[d_4]1} = 20.5$
	S_5	$\{d_1, d_2, d_3, d_4, \boldsymbol{d_5}\} = \{\{1,2\}, \{3,4\}, \{5,6\}, \{7,8\}, \{\mathbf{9,10}\}\}$	$X_{[d_5]1} = 21.0$
2	S_1	$\{\boldsymbol{d_1}, d_2, d_3, d_4, d_5\} = \{\{\mathbf{1,2}\}, \{3,4\}, \{5,6\}, \{7,8\}, \{9,10\}\}$	$X_{[d_2]2} = 12.1$
	S_2	$\{d_1, \boldsymbol{d_2}, d_3, d_4, d_5\} = \{\{1,2\}, \{\mathbf{3,4}\}, \{5,6\}, \{7,8\}, \{9,10\}\}$	$X_{[d_2]2} = 17.4$
	S_3	$\{d_1, d_2, \boldsymbol{d_3}, d_4, d_5\} = \{\{1,2\}, \{3,4\}, \{\mathbf{5,6}\}, \{7,8\}, \{9,10\}\}$	$X_{[d_3]2} = 29.9$
	S_4	$\{d_1, d_2, d_3, \boldsymbol{d_4}, d_5\} = \{\{1,2\}, \{3,4\}, \{5,6\}, \{\mathbf{7,8}\}, \{9,10\}\}$	$X_{[d_4]2} = 22.0$
	S_5	$\{d_1, d_2, d_3, d_4, \boldsymbol{d_5}\} = \{\{1,2\}, \{3,4\}, \{5,6\}, \{7,8\}, \{\mathbf{9,10}\}\}$	$X_{[d_5]2} = 29.8$
3	S_1	$\{\boldsymbol{d_1}, d_2, d_3, d_4, d_5\} = \{\{\mathbf{1,2}\}, \{3,4\}, \{5,6\}, \{7,8\}, \{9,10\}\}$	$X_{[d_1]3} = 9.4$
	S_2	$\{d_1, \boldsymbol{d_2}, d_3, d_4, d_5\} = \{\{1,2\}, \{\mathbf{3,4}\}, \{5,6\}, \{7,8\}, \{9,10\}\}$	$X_{[d_2]3} = 7.1$
	S_3	$\{d_1, d_2, \boldsymbol{d_3}, d_4, d_5\} = \{\{1,2\}, \{3,4\}, \{\mathbf{5,6}\}, \{7,8\}, \{9,10\}\}$	$X_{[d_3]3} = 21.5$
	S_4	$\{d_1, d_2, d_3, \boldsymbol{d_4}, d_5\} = \{\{1,2\}, \{3,4\}, \{5,6\}, \{\mathbf{7,8}\}, \{9,10\}\}$	$X_{[d_4]3} = 19.2$
	S_5	$\{d_1, d_2, d_3, d_4, \boldsymbol{d_5}\} = \{\{1,2\}, \{3,4\}, \{5,6\}, \{7,8\}, \{\mathbf{9,10}\}\}$	$X_{[d_5]3} = 24.4$
4	S_1	$\{\boldsymbol{d_1}, d_2, d_3, d_4, d_5\} = \{\{\mathbf{1,2}\}, \{3,4\}, \{5,6\}, \{7,8\}, \{9,10\}\}$	$X_{[d_1]4} = 19.3$
	S_2	$\{d_1, \boldsymbol{d_2}, d_3, d_4, d_5\} = \{\{1,2\}, \{\mathbf{3,4}\}, \{5,6\}, \{7,8\}, \{9,10\}\}$	$X_{[d_2]4} = 9.9$
	S_3	$\{d_1, d_2, \boldsymbol{d_3}, d_4, d_5\} = \{\{1,2\}, \{3,4\}, \{\mathbf{5,6}\}, \{7,8\}, \{9,10\}\}$	$X_{[d_3]4} = 17.0$
	S_4	$\{d_1, d_2, d_3, \boldsymbol{d_4}, d_5\} = \{\{1,2\}, \{3,4\}, \{5,6\}, \{\mathbf{7,8}\}, \{9,10\}\}$	$X_{[d_4]4} = 18.1$
	S_5	$\{d_1, d_2, d_3, d_4, \boldsymbol{d_5}\} = \{\{1,2\}, \{3,4\}, \{5,6\}, \{7,8\}, \{\mathbf{9,10}\}\}$	$X_{[d_5]4} = 21.2$
5	S_1	$\{\boldsymbol{d_1}, d_2, d_3, d_4, d_5\} = \{\{\mathbf{1,2}\}, \{3,4\}, \{5,6\}, \{7,8\}, \{9,10\}\}$	$X_{[d_1]5} = 9.4$
	S_2	$\{d_1, \boldsymbol{d_2}, d_3, d_4, d_5\} = \{\{1,2\}, \{\mathbf{3,4}\}, \{5,6\}, \{7,8\}, \{9,10\}\}$	$X_{[d_2]5} = 20.5$
	S_3	$\{d_1, d_2, \boldsymbol{d_3}, d_4, d_5\} = \{\{1,2\}, \{3,4\}, \{\mathbf{5,6}\}, \{7,8\}, \{9,10\}\}$	$X_{[d_3]5} = 21.8$
	S_4	$\{d_1, d_2, d_3, \boldsymbol{d_4}, d_5\} = \{\{1,2\}, \{3,4\}, \{5,6\}, \{\mathbf{7,8}\}, \{9,10\}\}$	$X_{[d_4]5} = 5.2$
	S_5	$\{d_1, d_2, d_3, d_4, \boldsymbol{d_5}\} = \{\{1,2\}, \{3,4\}, \{5,6\}, \{7,8\}, \{\mathbf{9,10}\}\}$	$X_{[d_5]5} = 38.1$

6. Conclusions

In scientific researches, time and cost of the study determine how many observations can be used. Therefore, researchers prefer to study with fewer observations. For example, we showed that only 25 observations can be used instead of 252 observations by using PROS Level-2 sampling design. On the other hand, normality assumption is vital for parametric tests. For this purpose, we studied GOF tests for normality in this study.

According to the simulation results, it is proved that the EDF based on PROS Level-2 sampling design is the most efficient estimator among the other EDF estimators for symmetric, skewed distributions with light tail or heavy tail. Also, in general, the quadratic class GOF tests (W^2 and A^2) have better performance than supremum class GOF tests (D and V) for SRS, RSS and PROS. Espicially, the Anderson-Darling GOF test (A^2) has the highest powers among the GOF tests. On the other hand, it is seen that the powers of the GOF tests based on PROS Level-2 sampling design, especially, Kolmogorov-Smirnov, Cramér-von Mises and Anderson-Darling GOF tests, are the highest when $k = 5$. It is very important that the quality of ranking should be almost perfect since the proposed test statistics have larger type I errors than nominal value when ranking is poor ($\rho = 0.25$). According to the Figures 3-6, it is observed that the powers of GOF tests based on SRS, RSS and PROS get close to 1 when the set size is 5 and the distribution is ($g = 1, h = 0$). In the other distributions, the sample size must be larger than 25 (when the set size and the number of cycles are 5) for the powers close to 1. The largest sample size is taken as 25 in the simulation study. For the sample size to be greater than 25, the population size that is larger than 250 must be considered. In real data application, the PROS Level-2 sampling design is applied to

252 men's percentage of body fats (X). Ranking procedure is done using ages (Y) of the 252 men, $cor(X,Y) = 0.813$.

Another important note is that the critical values of GOF tests based on the PROS sampling design can be obtained for any set size k, the number subsets H and the number of cycles l using the algorithm which is given in Section four. Therefore, the proposed GOF tests can be used for any case studies such as in Section 5.

Acknowledgments

This study was supported by Department of Scientific Research Projects in Dokuz Eylül Univesity (Project No: 2018.KB.FEN.003). Also, Yusuf Can SEVIL would like to thank to the Scientific and Technological Research Council of Turkey (TUBITAK) for funding the study through 2211/A Fellowship programme. Moreover, the authors thank the reviewers and the editor for helpful comments that have improved the paper.

References

[1] McIntyre, G.A. (1952). A method for unbiased selective sampling, using ranked sets. *Australian Journal of Agricultural Research*, 3(4), 385-390.

[2] Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Annals of the Institute of Statistical Mathematics*, 20(1), 1-31.

[3] Dell, T. and Clutter, J. (1972). Ranked set sampling theory with order statistics background. *Biometrics*, 545-555.

[4] Kaur, A., Patil, G., Sinha, A. and Taillie, C. (1995). Ranked set sampling: an annotated bibliography. *Environmental and Ecological Statistics*, 2(1), 25-54.

[5] Chen, Z., Bai, Z. and Sinha, B. (2003). Ranked set sampling: theory and applications. Springer Science & Business Media.

[6] Al-Omari, A. I. and Bouza, C. N. (2014). Review of ranked set sampling: modifications and applications. *Revista Investigación Operacional*, 3, 215-240.

[7] Deshpande, J. V. and Frey, J. and Ozturk, O. (2006). Nonparametric ranked-set sampling confidence intervals for quantiles of a finite population. *Environmental and Ecological Statistics*, 13(1), 25-40.

[8] Al-Saleh, M. F. and Samawi, H. M. (2007). A note on inclusion probability in ranked set sampling and some of its variations. *Test*, 16(1), 198-209.

[9] Ozdemir, Y. A. and Gokpinar, F. (2007). A generalized formula for inclusion probabilities in ranked set sampling. *Hacettepe Journal of Mathematics and Statistics*, 36(1), 89-99.

[10] Ozdemir, Y. A. and Gokpinar, F. (2008). A new formula for inclusion probabilities in median ranked set sampling. *Communication in Statistics- Theory and Methods*, 37(13), 2022-2033.

[11] Gokpinar, F, Ozdemir, Y. A. (2010). Generalization of inclusion probabilities in ranked set sampling. *Hacettepe Journal of Mathematics and Statistics*, 39(1), 89-95.

[12] Frey, J. (2011). Recursive computation of inclusion probabilities in ranked set sampling. *Journal of Statistical Planning and Inference*, 141(11), 3632-3639.

[13] Jafari Jozani, M., Johnson, B. C. (2011). Design based estimation for ranked set sampling in finite population. *Environmental and Ecological Statistics*, 18(4), 663-685.

[14] Jafari Jozani, M., Johnson, B. C. (2011). Randomized nomination sampling in finite populations. *Journal of Statistical Planning and inference*, 142(7), 2103-2115.

[15] Stokes, S L. and Sager, T. W. (1988). Characterization of a ranked-set sample with application to estimating distribution functions. *Journal of the American Statistical Association*, 83(402), 374-381.

[16] Frey, J. and Wang, L. (2014). EDF-based goodness-of-fit tests for ranked-set sampling. *Canadian Journal of Statistics*, 42(3), 451-469.

[17] Nazari, S., Jafari Jozani, M. and Kharrati-Kopaei, M. (2014). Nonparametric density estimation using partially rank-ordered set samples with application in estimating the distribution of wheat yield. *Electronic Journal of Statistics*, 8(1), 738-761.

[18] Sevil, Y. C. and Yildiz, T. O. (2017). Power comparison of the Kolmogorov–Smirnov test under ranked set sampling and simple random sampling. *Journal of Statistical Computation and Simulation*, 87(11), 2175-2185.

[19] Yildiz, T. O., and Sevil, Y. C. (2018). Performances of some goodness-of-fit tests for sampling designs in ranked set sampling. *Journal of Statistical Computation and Simulation*, 88(9), 1702-1716.

[20] Yildiz, T. O., and Sevil, Y. C. (2019). Empirical distribution function estimators based on sampling designs in a finite population using single auxiliary variable. *Journal of Applied Statistics*, 46(16), 2962-2974.

[21] Sevil, Y. C. and Yildiz, T. O. (2020). Performances of the distribution function estimators based on ranked set sampling using body fat data. *Turkiye Klinikleri Journal of Biostatistics*, 12(2), 218-228.

[22] Ozturk, O. (2011). Sampling from partially rank-ordered sets. *Environmental and Ecological statistics*, 18(4), 757-779.

[23] Ozturk, O. (2012a). Combining ranking information in judgment post stratified and ranked set sampling designs. *Environmental and Ecological Statistics*, 19(1), 73-93.

[24] Ozturk, O. (2012b). Quantile inference based on partially rank-ordered set samples. *Journal of Statistical Planning and Inference*, 142(7), 2116-2127.

[25] Ozturk, O. (2014). Estimation of population mean and total in a finite population setting using multiple auxiliary variables. *Journal of Agricultural, Biological, and Environmental Statistics*, 19(2), 161-184.

[26] Ozturk, O. and Jafari Jozani, M. (2014). Inclusion probabilities in partially rank ordered set sampling. *Computational Statistics & Data Analysis*, 69, 122-132.

[27] Hatefi, A., Jafari Jozani, M. and Ozturk, O. (2015). Mixture model analysis of partially rank-ordered set samples: age groups of fish from length-frequency data. *Scandinavian Journal of Statistics*, 42(3), 848-871.

[28] Wang, X., Wang, K., and Lim, J. (2012). Isotonized CDF estimation from judgment poststratification data with empty strata. *Biometrics*, 68(1), 194-202.

[29] D'Agostino, R. B. (1986). Goodness-of-fit techniques. CRC press.

[30] Penrose, K. W., Nelson, A. G., and Fisher, A. G. (1985). Generalized body composition prediction equation for men using simple measurement techniques. *Medicine & Science in Sports & Exercise*, 17(2), 189.

Appendices

TABLE 4. Critical values of GOF tests based on RSS and PROS sampling designs when $\alpha = 0.05$

Methods	Designs	k	l	D	V	W^2	U^2	A^2
RSS	Level-0	3	1	0.581	0.820	0.248	0.148	1.557
			2	0.439	0.608	0.277	0.151	1.610
			3	0.366	0.509	0.265	0.155	1.548
			4	0.320	0.451	0.272	0.159	1.576
			5	0.290	0.406	0.277	0.158	1.566
		5	1	0.429	0.626	0.197	0.130	1.192
			2	0.310	0.463	0.204	0.133	1.217
			3	0.263	0.383	0.204	0.134	1.216
			4	0.232	0.337	0.209	0.136	1.243
			5	0.210	0.309	0.209	0.138	1.251
	Level-1	3	1	0.583	0.813	0.251	0.145	1.573
			2	0.430	0.604	0.261	0.146	1.508
			3	0.366	0.503	0.262	0.150	1.485
			4	0.314	0.438	0.258	0.149	1.492
			5	0.279	0.397	0.260	0.150	1.470
		5	1	0.429	0.619	0.202	0.127	1.221
			2	0.307	0.454	0.194	0.129	1.162
			3	0.258	0.376	0.197	0.127	1.158
			4	0.221	0.328	0.190	0.127	1.123
			5	0.202	0.294	0.191	0.126	1.128
	Level-2	3	1	0.581	0.818	0.249	0.147	1.561
			2	0.431	0.607	0.261	0.150	1.513
			3	0.366	0.506	0.263	0.153	1.528
			4	0.311	0.439	0.249	0.147	1.447
			5	0.277	0.396	0.250	0.147	1.454
		5	1	0.419	0.612	0.192	0.125	1.157
			2	0.304	0.455	0.191	0.128	1.150
			3	0.254	0.375	0.185	0.127	1.107
			4	0.214	0.324	0.174	0.123	1.060
			5	0.192	0.291	0.168	0.120	1.023
PROS	Level-0	3	1	0.531	0.773	0.201	0.128	1.339
			2	0.399	0.586	0.207	0.134	1.290
			3	0.335	0.492	0.211	0.140	1.299
			4	0.290	0.430	0.208	0.138	1.286
			5	0.261	0.390	0.212	0.139	1.299
		5	1	0.384	0.577	0.148	0.107	0.967
			2	0.283	0.430	0.146	0.107	0.948
			3	0.235	0.361	0.150	0.111	0.958
			4	0.206	0.316	0.149	0.109	0.942
			5	0.187	0.289	0.152	0.114	0.967
	Level-1	3	1	0.528	0.768	0.197	0.125	1.318
			2	0.397	0.577	0.204	0.131	1.244
			3	0.326	0.484	0.201	0.134	1.240
			4	0.289	0.422	0.199	0.134	1.223
			5	0.257	0.381	0.194	0.131	1.203
		5	1	0.373	0.569	0.138	0.102	0.932
			2	0.279	0.425	0.142	0.104	0.913
			3	0.229	0.352	0.142	0.104	0.896
			4	0.201	0.310	0.141	0.104	0.889
			5	0.178	0.277	0.139	0.103	0.876
	Level-2	3	1	0.528	0.770	0.198	0.127	1.319
			2	0.395	0.583	0.201	0.134	1.252
			3	0.323	0.484	0.197	0.132	1.230
			4	0.282	0.423	0.190	0.133	1.169
			5	0.254	0.377	0.191	0.128	1.174
		5	1	0.372	0.572	0.137	0.102	0.924
			2	0.273	0.423	0.134	0.103	0.865
			3	0.222	0.348	0.128	0.102	0.833
			4	0.188	0.303	0.120	0.098	0.790
			5	0.169	0.272	0.115	0.099	0.762

İSTATİSTİK: JOURNAL OF THE TURKISH STATISTICAL ASSOCIATION

İSTATİSTİK

Vol. 13, No. 2, July 2021, pp. 74–87

ISSN 1300-4077 | 21 | 2 | 74 | 87

ESTIMATION OF STRESS-STRENGTH RELIABILITY OF A PARALLEL SYSTEM WITH COLD STANDBY REDUNDANCY AT COMPONENT LEVEL

Gülce Cüran

Department of Mathematics,
Yeditepe University,
34755, Istanbul, Turkey,

Fatih Kızılaslan *

Department of Statistics,
Marmara University,
34722, Istanbul, Turkey

Abstract: In this paper, we consider the estimation problem of stress-strength reliability of a parallel system with cold standby redundancy. The reliability of the system is estimated when both strength and stress variables follow the exponential distribution and associated approximate confidence interval is constructed. Two different maximum likelihood and Bayes estimates are obtained. Lindley's approximation method has been utilized for Bayesian calculations. A real-life data set is analysed for illustrative purposes of the findings.

Key words: Stress-strength reliability; Parallel system; Cold standby redundancy

1. Introduction

In reliability analysis, stress-strength probability is a major concern for some experiments in the fields including industrial engineering, hydrology, economics and survival analysis. Stress-strength models are introduced originally by Birnbaum [2] then developed by Birnbaum and McCarty [3] and Church and Harris [6]. When X is the random strength of a system under the random stress Y, the probability $R = P(X > Y)$ indicates the measure of the performance of the system. If the stress exceeds the strength, the system stops operating, otherwise it continues to work. In 2003, Kotz et al. [13] interpreted the concept of stress-strength combined with the theory and applications.

A k-out-of-n: G system consists of n independent and identically distributed strength components and a common stress, and functions at least k out of the n components operate. When $k = 1$ and $k = n$, the k-out-of-n: G system becomes a parallel and series systems, respectively. A parallel system fails if and only if its each component fails so that this system works whenever at least one component works. Multicomponent stress-strength reliability has been of great interest among researchers in recent years. In this context, we can refer to Eryilmaz [8], Pakdaman and Ahmadi [18], Kızılaslan [12], Akgül [1] and Dey et al. [7].

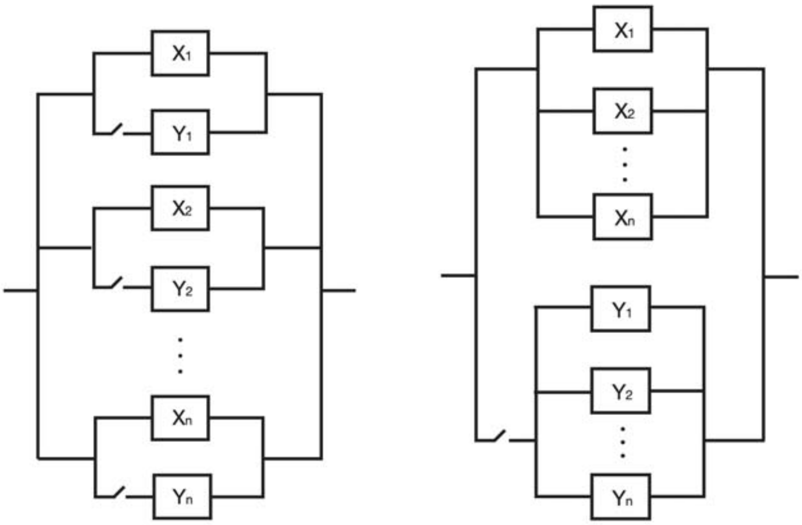
Standby redundancy allocation to a system or components makes a great impact on system lifetime. Hence, it is widely used to improve system reliability. Different types of standby redundancy have been introduced and studied in the reliability literature. The component is said to be in the case of cold standby if it does not fail while in standby. When the components of a system fail, cold standby redundancy puts into operation and system operates until the standby component fails.

The cold standby redundancy can be implemented to a system at system and component levels. At component level, the standby components are connected to the original components one by one. At system level, the standby components construct an alternative system for the system of

* Corresponding author. E-mail address:fatih.kizilaslan@marmara.edu.tr

original components. When standby components are added to a n-component parallel system at the component or system level, new systems are obtained as in Figure 1.

We can consider the following example for the parallel system with standby components at component level. Computer systems are used in many areas. Saving customer transactions data automatically in banking system is one of them. For example, six computers in parallel design be used for data saving. The random variables $X_1, \ldots, X_6$ represent the lifetime of these computers. We add additional six computers to the parallel system as standby components at component level. The random variables $Y_1, \ldots, Y_6$ represent the lifetime of these new standby computers. Some factors such as lifetime of the subcomponents, cyber attacks, density of the data, etc. can be considered as stress variables of this system. The random variable T represents the stress variable. In this scenario, a probability for the total lifetime of the parallel system under the stress, that is $P(\max(X_1+Y_1, X_2+Y_2, \ldots, X_6+Y_6) > T)$, is the probability of saving data successfully.



(a) Component level (b) System level

FIGURE 1. Standby redundancy of parallel systems

In the literature, the stochastic comparisons of coherent systems when the cold standby redundancy applied at component level and system level have been investigated by many researchers. Chen and Xie [23] studied the effect of adding standby redundancy at system and component levels in series and parallel systems, and compared the superiority of the levels. Boland and El-Neweihi [4] extended the stochastic comparison results of the cold standby levels for the series and parallel systems from the usual stochastic ordering to the hazard rate ordering. They also obtained some results about the hazard rate ordering of the k-out-of-n: G system with cold standby redundancy. Zhao et al. [27] presented the likelihood ratio ordering result for the series system with n exponential components when the active components and standby components are identical. Similar results for the parallel system was proved for $n = 2$ case. Eryilmaz and Tank [9] considered a series system with two dependent components and a single cold standby component. Eryilmaz [10] investigated the effect of adding cold standby redundancy to a general coherent structure at system and component levels. Tuncel [25] studied the residual lifetime of a single component system with a cold standby component when the lifetimes of these components were dependent. Chen et al. [5] discussed the problem of optimal allocation methods for two standby spares in a two-component series/parallel structure. Yan et al. [26] studied series and parallel systems of two components with one standby redundancy component when all components having exponential distribution. Roy and Gupta [21, 22] considered the reliability of a k-out-of-n and coherent systems equipped with two cold standby components, respectively.

The stress-strength reliability estimation of the k-out-of-n: G system has been paying great attention by researchers in decades. Many studies are available in the literature about this topic. Since adding standby component(s) to the system increases the system reliability, to investigate the reliability of this type of a system is useful for practitioners. Some recent studies are mentioned in the next references. Siju and Kumar [24] considered the estimation of reliability of a parallel system under the hybrid of active, warm and cold standby components by applying maximum likelihood method. The reliability estimation for the standby redundancy system consists of a certain number of same subsystems with series structure was considered by Liu et al. [15] for the generalized half-logistic distribution based on progressive Type-II censoring sample.

In this study, we consider a parallel system of n components with n standby components whenever the standby components connected to working components at component level. The estimation problem for the stress-strength reliability of this parallel system has been studied.

The rest of this article is organized as follows: Section 2 presents the formulation of the model that includes standby components. Section 3 contains the maximum likelihood estimate (MLE) and Bayes estimate of the reliability of the parallel system under the common stress. Section 4 contains the simulation study in order to compare the performance of the obtained estimators. A real life data analysis is considered for illustrative purposes of the proposed estimates.

2. Model definition

In this section, we have considered the problem of stress-strength reliability estimation of a parallel system with standby redundancy at component level. Suppose an n component parallel system with standby components that are independent but not identically distributed with original components. We assume that $X_1,\ldots,X_n$ represent the lifetimes of strength components and $Y_1,\ldots,Y_n$ are the lifetimes of independent standby strength components having exponential distributions with parameters α and β, respectively. T represents the common stress variable and follows the exponential distribution with parameter θ. This parallel system structure is shown in Figure 1a.

In the case of component level, standby redundancy is applied to the components one by one. The total lifetime of each strength component is $Z_i = X_i + Y_i$, $i = 1,\ldots,n$. Then, the cumulative distribution function (cdf) and probability density function (pdf) of Z_i, $i = 1,\ldots,n$ are given by

$$F_{Z_i}(z) = \int_0^z F_x(z-y)f_y(y)dy = \begin{cases} 1+\frac{\alpha e^{-\beta z}-\beta e^{-\alpha z}}{\beta-\alpha} & , \alpha \neq \beta \\ 1-e^{-\alpha z}(1+\alpha z) & , \alpha = \beta \end{cases} \tag{2.1}$$

and

$$f_{Z_i}(z) = \begin{cases} \frac{\alpha\beta}{\beta-\alpha}(e^{-\alpha z}-e^{-\beta z}) & , \alpha \neq \beta \\ \alpha^2 z e^{-\alpha z} & , \alpha = \beta \end{cases}, \quad i = 1,\ldots,n. \tag{2.2}$$

When the active strength and standby redundancy components are identical, i.e. $\alpha = \beta$, the total lifetime distributions Z_i, $i = 1,\ldots,n$ follow the gamma distribution with shape parameter 2 and rate parameter α. In this case, the stress-strength reliability problem is reduced to the simple stress-strength reliability of the gamma components in Nojosa and Rathie [17]. Hence, we consider the case where parameters α and β are not the same in this study. Since $Z_1,\ldots,Z_n$ denote n

independent total strength random variables given in (2.1), the cdf for the lifetime of the parallel system is

$$F_{Z_{(n)}}(z) = \sum_{k=0}^{n} \binom{n}{k} \frac{(\alpha e^{-\beta z} - \beta e^{-\alpha z})^k}{(\beta-\alpha)^k}$$
$$= \sum_{k=0}^{n} \sum_{j=0}^{k} \binom{n}{k} \binom{k}{j} (-1)^j \frac{\alpha^{k-j}\beta^j}{(\beta-\alpha)^k} e^{-z[j(\alpha-\beta)+\beta k]}$$

where $Z_{(n)} = \max(Z_1,\dots,Z_n)$.

When the maximum strength component $Z_{(n)}$ is subjected to the common stress component T, the stress-strength reliability of the parallel system is obtained as

$$R = \int_0^\infty P(Z_{(n)} > T | T = t) f_T(t) dt$$
$$= 1 - \sum_{k=0}^{n} \sum_{j=0}^{k} \binom{n}{k} \binom{k}{j} \frac{(-1)^j}{(\beta-\alpha)^k} \frac{\alpha^{k-j}\beta^j\theta}{\alpha j + \beta(k-j) + \theta}. \tag{2.3}$$

Moreover, if we use n standby strength components as original components in the parallel system, we construct a new parallel system with $2n$ strength components. The stress-strength reliability of the new parallel system is derived under the common stress for comparison. Let V_i be the lifetime of i^{th} strength components $i = 1,\dots,2n$, and follow the exponential distribution with parameter α for $1 \le i \le n$ and β for $n+1 \le i \le 2n$. In this case, the cdf for the lifetime of the $2n$ components parallel system is

$$F_{V_{(2n)}}(t) = P(\max(V_1,\dots,V_{2n}) \le t) = (1-e^{-\alpha t})^n (1-e^{-\beta t})^n.$$

Then, the stress-strength reliability of this parallel system is obtained as

$$R_V = \int_0^\infty P(V_{(2n)} > T | T = t) f_T(t) dt$$
$$= 1 - \sum_{k=0}^{n} \sum_{j=0}^{n} \binom{n}{k} \binom{n}{j} (-1)^{k+j} \frac{\theta}{\alpha k + \beta j + \theta} \tag{2.4}$$

under the common stress T.

To show the effect of the standby components at component level in the stress-strength reliability of the parallel system, we present the following graphics. In Figure 2, the comparison of two stress-strength reliabilities R in (2.3) and R_V in (2.4) are plotted with respect to the different parameters. It is seen that adding standby components increases system reliability, as expected.

3. Estimation of R

In this section, the ML and Bayes estimates of the stress-strength reliability of the aforementioned parallel system are investigated.

3.1. MLE of R

Let m systems be put on a test each with n original components and n cold standby components in the parallel system. The strength data is represented as $Z_{i1},\dots,Z_{in}$, $i = 1,\dots,m$ and stress is $T_i, i = 1,\dots,m$. Then, the likelihood function of the observed sample is

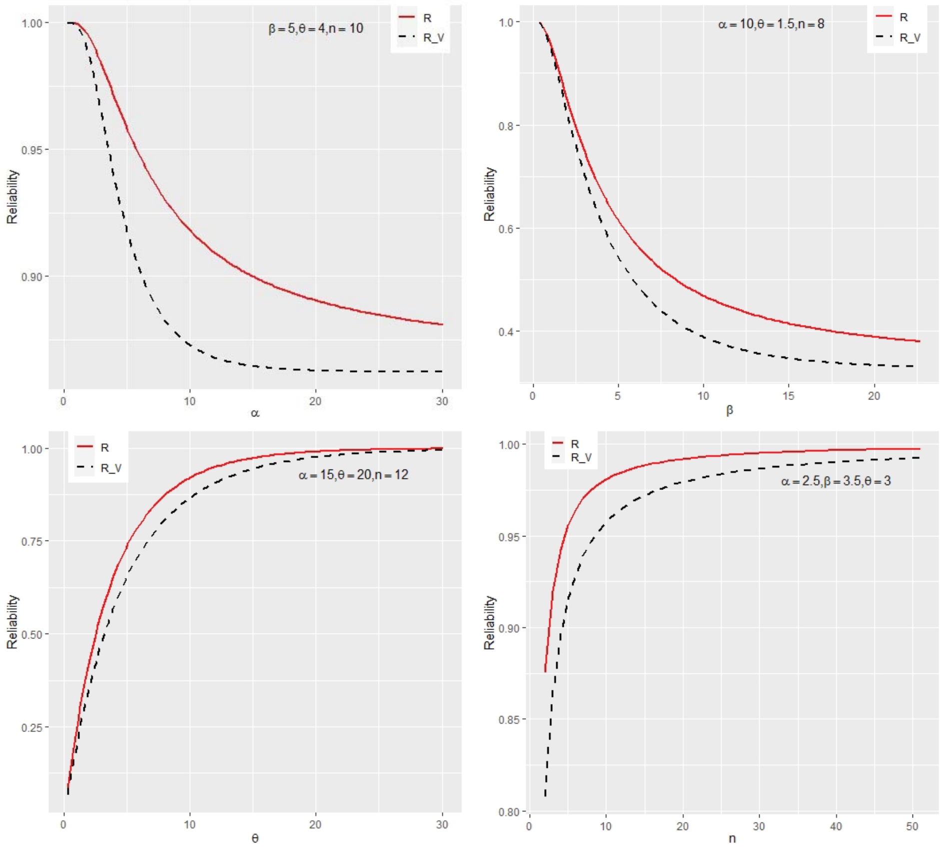


FIGURE 2. Plots for the reliabilities of R and R_V

$$
\begin{aligned}
L(\alpha,\beta,\theta|\mathbf{z},\mathbf{t}) &= \prod_{i=1}^{m}\left(\prod_{j=1}^{n} f_Z(z_{ij})\right) f_T(t_i) \\
&= \left(\frac{\alpha\beta}{\beta-\alpha}\right)^{nm} \exp\left(\sum_{i=1}^{m}\sum_{j=1}^{n}\ln\left(e^{-\alpha z_{ij}}-e^{-\beta z_{ij}}\right)\right)\theta^m e^{-\theta\sum_{i=1}^{m} t_i},
\end{aligned}
$$

and the log-likelihood function is given by

$$
\ell(\alpha,\beta,\theta;\mathbf{z},\mathbf{t}) = nm\left(\ln(\alpha\beta)-\ln(\beta-\alpha)\right)+m\ln\theta+\sum_{i=1}^{m}\sum_{j=1}^{n}\ln\left(e^{-\alpha z_{ij}}-e^{-\beta z_{ij}}\right)-\theta\sum_{i=1}^{m} t_i.
$$

The partial derivatives with respect to three parameters are obtained in the following forms:

$$
\frac{\partial\ell}{\partial\alpha} = nm\left(\frac{1}{\alpha}+\frac{1}{\beta-\alpha}\right)-\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{z_{ij}e^{-\alpha z_{ij}}}{e^{-\alpha z_{ij}}-e^{-\beta z_{ij}}}, \tag{3.1}
$$

$$
\frac{\partial\ell}{\partial\beta} = nm\left(\frac{1}{\beta}-\frac{1}{\beta-\alpha}\right)+\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{z_{ij}e^{-\beta z_{ij}}}{e^{-\alpha z_{ij}}-e^{-\beta z_{ij}}} \tag{3.2}
$$

and

$$\frac{\partial \ell}{\partial \theta} = \frac{m}{\theta} - \sum_{i=1}^{m} t_i.$$

Hence, the ML estimate of θ is $\widehat{\theta} = 1/\overline{T}$, and the ML estimates of α and β, that is $\widehat{\alpha}$ and $\widehat{\beta}$, are the solutions of non-linear equation system given in (3.1) and (3.2). $\widehat{\alpha}$ and $\widehat{\beta}$ can be derived with the help of numerical methods, like Newton-Raphson or Broyden's method. Therefore, $\widehat{R}$ can be obtained as

$$\widehat{R} = 1 - \sum_{k=0}^{n}\sum_{j=0}^{k} \binom{n}{k}\binom{k}{j} \frac{(-1)^j}{(\widehat{\beta}-\widehat{\alpha})^k} \frac{\widehat{\alpha}^{k-j}\widehat{\beta}^j\widehat{\theta}}{\widehat{\alpha}j + \widehat{\beta}(k-j) + \widehat{\theta}},$$

from (2.3) by using the invariance property of MLE.

3.2. Asymptotic distribution and confidence interval of R

The observed information matrix of $\tau = (\alpha, \beta, \theta)$ is given as

$$J(\tau) = -\begin{pmatrix} \frac{\partial^2 \ell}{\partial \alpha^2} & \frac{\partial^2 \ell}{\partial \alpha \partial \beta} & \frac{\partial^2 \ell}{\partial \alpha \partial \theta} \\ \frac{\partial^2 \ell}{\partial \beta \partial \alpha} & \frac{\partial^2 \ell}{\partial \beta^2} & \frac{\partial^2 \ell}{\partial \beta \partial \theta} \\ \frac{\partial^2 \ell}{\partial \theta \partial \alpha} & \frac{\partial^2 \ell}{\partial \theta \partial \beta} & \frac{\partial^2 \ell}{\partial \theta^2} \end{pmatrix} = \begin{pmatrix} J_{11} & J_{12} & J_{13} \\ J_{21} & J_{22} & J_{23} \\ J_{31} & J_{32} & J_{33} \end{pmatrix}.$$

Since $J_{13} = J_{31} = J_{23} = J_{32} = 0$, other entries of the matrix are

$$J_{11} = nm\left(\frac{1}{\alpha^2} - \frac{1}{(\beta-\alpha)^2}\right) + \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{z_{ij}^2 e^{-z_{ij}(\beta-\alpha)}}{(1-e^{-z_{ij}(\beta-\alpha)})^2},$$

$$J_{12} = J_{21} = \frac{nm}{(\beta-\alpha)^2} - \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{z_{ij}^2 e^{-z_{ij}(\beta-\alpha)}}{(1-e^{-z_{ij}(\beta-\alpha)})^2}$$

and

$$J_{22} = nm\left(\frac{1}{\beta^2} - \frac{1}{(\beta-\alpha)^2}\right) + \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{z_{ij}^2 e^{-z_{ij}(\beta-\alpha)}}{(1-e^{-z_{ij}(\beta-\alpha)})^2}, \qquad J_{33} = \frac{m}{\theta^2}.$$

The expectations of the entries of the observed information matrix cannot be obtained analytically. Therefore, the Fisher Information matrix $I(\tau) = E(J(\tau))$ can be obtained by using numerical methods. The MLE of R is asymptotically normal with mean R and asymptotic variance

$$\sigma_R^2 = \sum_{j=1}^{3}\sum_{i=1}^{3} \frac{\partial R}{\partial \tau_i}\frac{\partial R}{\partial \tau_j} I_{ij}^{-1},$$

where I_{ij}^{-1} is the $(i,j)^{th}$ element of the inverse of $I(\tau)$, see Rao [20]. Afterwards,

$$\sigma_R^2 = \left(\frac{\partial R}{\partial \alpha}\right)^2 I_{11}^{-1} + 2\frac{\partial R}{\partial \alpha}\frac{\partial R}{\partial \beta} I_{12}^{-1} + \left(\frac{\partial R}{\partial \beta}\right)^2 I_{22}^{-1} + \left(\frac{\partial R}{\partial \theta}\right)^2 I_{33}^{-1}. \tag{3.3}$$

Note that $I(\tau)$ can be replaced by $J(\tau)$ when $I(\tau)$ is not obtained. Therefore, an asymptotic $100(1-\gamma)\%$ confidence interval of R is given by $\left(\widehat{R}^{MLE} \pm z_{\gamma/2}\widehat{\sigma}_R\right)$ where $z_{\gamma/2}$ is the upper $\gamma/2$th quantile of the standard normal distribution and $\widehat{\sigma}_R$ is the value of σ_R at the MLE of the parameters.

3.3. Bayes estimation of R

In this section, we assume that α, β and θ are random variables that follow independent gamma prior distributions with parameters (a_i, b_i), $i = 1, 2, 3$, respectively. The pdf of a gamma random variable X with parameters (a_i, b_i) is given as

$$f(x) = \frac{{b_i}^{a_i}}{\Gamma(a_i)} x^{a_i - 1} e^{-xb_i}, \qquad x > 0, \quad a_i, b_i > 0 \quad \text{and} \quad i = 1, 2, 3.$$

The joint posterior density function of α, β and θ is

$$\pi(\alpha, \beta, \theta | \mathbf{z}, \mathbf{t}) = I(\mathbf{z}, \mathbf{t}) \alpha^{nm + a_1 - 1} \beta^{nm + a_2 - 1} (\beta - \alpha)^{-nm} \theta^{a_3 + m - 1} e^{-\alpha b_1 - \beta b_2 - \theta \left(b_3 + \sum_{i=1}^m t_i\right)} z_{\alpha,\beta} \tag{3.4}$$

where $I(\mathbf{z}, \mathbf{t})$ is the normalizing constant and written by

$$\frac{I(\mathbf{z}, \mathbf{t})^{-1}}{\Gamma(a_3 + m)} \left(b_3 + \sum_{i=1}^{m} t_i \right)^{a_3 + m} = \int_0^\infty \int_0^\infty \left(\frac{\alpha\beta}{\beta - \alpha} \right)^{nm} \alpha^{a_1 - 1} \beta^{a_2 - 1} e^{-\alpha b_1 - \beta b_2} z_{\alpha,\beta} d\alpha d\beta$$

where

$$z_{\alpha,\beta} = \exp\left(\sum_{i=1}^{m} \sum_{j=1}^{n} \ln\left(e^{-\alpha z_{ij}} - e^{-\beta z_{ij}} \right) \right).$$

The Bayes estimator of R under the SE loss function is

$$\widehat{R}_{Bayes} = \int_0^\infty \int_0^\infty \int_0^\infty R \ \pi(\alpha, \beta, \theta | \mathbf{z}, \mathbf{t}) d\alpha d\beta d\theta.$$

This integral cannot be easily computed analytically; thus, some approximation methods are needed. In order to obtain the Bayes estimate of R, we use Lindley's approximation method.

3.3.1. Lindley's approximation

Lindley [14] proposed an approximate method in order to obtain a numerical result for the computation of the ratio of two integrals. This procedure, applied to the posterior expectation of the function $u(\theta)$ for a given $\mathbf{x}$, is

$$E(u(\theta) | \mathbf{x}) = \frac{\int u(\theta) e^{Q(\theta)} d\lambda}{\int e^{Q(\theta)} d\lambda},$$

where $Q(\theta) = l(\theta) + \rho(\theta)$, $l(\theta)$ is the logarithm of the likelihood function and $\rho(\theta)$ is the logarithm of the prior density of θ. Using Lindley's approximation, $E(u(\theta) | \mathbf{x})$ is approximately estimated by

$$E(u(\theta) | \mathbf{x}) = \left[u + \frac{1}{2} \sum_i \sum_j (u_{ij} + 2u_i \rho_j) \sigma_{ij} + \frac{1}{2} \sum_i \sum_j \sum_k \sum_l L_{ijk} \sigma_{ij} \sigma_{kl} u_l \right]_{\widehat{\lambda}}$$
$$+ \text{terms of order } n^{-2} \text{ or smaller},$$

where $\theta = (\theta_1, \theta_2, \ldots, \theta_m)$, $i, j, k, l = 1, \ldots, m$, $\widehat{\theta}$ is the MLE of θ, $u = u(\theta)$, $u_i = \partial u / \partial \theta_i$, $u_{ij} = \partial^2 u / \partial \theta_i \partial \theta_j$, $L_{ijk} = \partial^3 l / \partial \theta_i \partial \theta_j \partial \theta_k$, $\rho_j = \partial \rho / \partial \theta_j$, and $\sigma_{ij} = (i, j)$th element in the inverse of the matrix $\{-L_{ij}\}$ all evaluated at the MLE of the parameters.

For three parameter case $\theta = (\theta_1, \theta_2, \theta_3)$, Lindley's approximation gives the approximate Bayes estimate as

$$\widehat{u}_B = E(u(\theta) | \mathbf{x}) = u + (u_1 a_1 + u_2 a_2 + u_3 a_3 + a_4 + a_5) + \frac{1}{2} [A (u_1 \sigma_{11} + u_2 \sigma_{12}$$
$$+ u_3 \sigma_{13}) + B (u_1 \sigma_{21} + u_2 \sigma_{22} + u_3 \sigma_{23}) + C (u_1 \sigma_{31} + u_2 \sigma_{32} + u_3 \sigma_{33})],$$

evaluated at $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\theta}_3)$, where

$$a_i = \rho_1\sigma_{i1} + \rho_2\sigma_{i2} + \rho_3\sigma_{i3},\ i = 1,2,3,$$

$$a_4 = u_{12}\sigma_{12} + u_{13}\sigma_{13} + u_{23}\sigma_{23},\ a_5 = \frac{1}{2}(u_{11}\sigma_{11} + u_{22}\sigma_{22} + u_{33}\sigma_{33}),$$
$$A = \sigma_{11}L_{111} + 2\sigma_{12}L_{121} + 2\sigma_{13}L_{131} + 2\sigma_{23}L_{231} + \sigma_{22}L_{221} + \sigma_{33}L_{331},$$
$$B = \sigma_{11}L_{112} + 2\sigma_{12}L_{122} + 2\sigma_{13}L_{132} + 2\sigma_{23}L_{232} + \sigma_{22}L_{222} + \sigma_{33}L_{332},$$
$$C = \sigma_{11}L_{113} + 2\sigma_{12}L_{123} + 2\sigma_{13}L_{133} + 2\sigma_{23}L_{233} + \sigma_{22}L_{223} + \sigma_{33}L_{333}.$$

In our system, for $(\theta_1, \theta_2, \theta_3) \equiv (\alpha, \beta, \theta)$ and $u \equiv u(\alpha, \beta, \theta) = R$ in (2.3). We compute σ_{ij}, $i, j = 1,2,3$ by using the following partial derivatives $L_{11} = -J_{11}, L_{12} = L_{21} = -J_{12}$ and $L_{22} = -J_{22}$. Using logarithm of the prior density, we have

$$\rho_1 = \frac{(a_1 - 1)}{\alpha} - b_1, \quad \rho_2 = \frac{(a_2 - 1)}{\beta} - b_2, \quad \rho_3 = \frac{(a_3 - 1)}{\theta} - b_3.$$

Additionally, we obtain $L_{333} = \frac{2m}{\theta^3}$ and

$$L_{111} = 2nm\left(\frac{1}{\alpha^3} + \frac{1}{(\beta-\alpha)^3}\right) - \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{z_{ij}^3 e^{-z_{ij}(\beta-\alpha)}\left(1 + e^{-z_{ij}(\beta-\alpha)}\right)}{\left(1 - e^{-z_{ij}(\beta-\alpha)}\right)^3},$$

$$L_{121} = L_{112} = -2nm\frac{1}{(\beta-\alpha)^3} + \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{z_{ij}^3 e^{-z_{ij}(\beta-\alpha)}\left(1 + e^{-z_{ij}(\beta-\alpha)}\right)}{\left(1 - e^{-z_{ij}(\beta-\alpha)}\right)^3},$$

$$L_{122} = L_{221} = 2nm\frac{1}{(\beta-\alpha)^3} - \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{z_{ij}^3 e^{-z_{ij}(\beta-\alpha)}\left(1 + e^{-z_{ij}(\beta-\alpha)}\right)}{\left(1 - e^{-z_{ij}(\beta-\alpha)}\right)^3},$$

$$L_{222} = 2nm\left(\frac{1}{\beta^3} - \frac{1}{(\beta-\alpha)^3}\right) + \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{z_{ij}^3 e^{-z_{ij}(\beta-\alpha)}\left(1 + e^{-z_{ij}(\beta-\alpha)}\right)}{\left(1 - e^{-z_{ij}(\beta-\alpha)}\right)^3}.$$

Let

$$S = \binom{n}{k}(-1)^k \frac{\alpha^k \beta^{n-k}\theta}{(\beta-\alpha)^n\left[\alpha\left(n-k\right) + \beta k + \theta\right]} \tag{3.5}$$

denote the common term in the first and second order partial derivatives of the parallel system reliability. We have the derivatives of reliability as given below

$$u_1 = \sum_{k=0}^{n} S \frac{\{[n\alpha + k(\beta-\alpha)][\alpha(n-k) + \beta k + \theta] - \alpha(\beta-\alpha)(n-k)\}}{\alpha(\beta-\alpha)\left[\alpha\left(n-k\right) + \beta k + \theta\right]},$$

$$u_2 = \sum_{k=0}^{n} S \frac{(-k)[\alpha(n-k) + \beta(k+1) + \theta]}{\beta\left[\alpha\left(n-k\right) + \beta k + \theta\right]},$$

$$u_3 = \sum_{k=0}^{n} S \frac{[\alpha(n-k) + \beta k]}{\theta\left[\alpha\left(n-k\right) + \beta k + \theta\right]},$$

$$u_{11} = \sum_{k=0}^{n} S \Bigg[\frac{n(n+1)}{(\beta-\alpha)^2} + \frac{2(n-k)}{\alpha(n-k) + \beta k + \theta}\left(-\frac{n}{\beta-\alpha} - \frac{k}{\alpha} + \frac{n-k}{\alpha(n-k) + \beta k + \theta}\right)$$
$$+ \frac{2kn}{\alpha(\beta-\alpha)} + \frac{k(k-1)}{\alpha^2}(k-1)\Bigg],$$

$$u_{12}=\sum_{k=0}^{n}S\left[\frac{n(n+1)}{(\beta-\alpha)^2}-\frac{2(n-k)}{\alpha(n-k)+\beta k+\theta}\left(\frac{n}{\beta-\alpha}+\frac{k}{\alpha}-\frac{n-k}{\alpha(n-k)+\beta k+\theta}\right)+\frac{2kn}{\alpha(\beta-\alpha)}+\frac{k(k-1)}{\alpha^2}\right],$$

$$u_{22}=\sum_{k=0}^{n}S\left[\frac{n(n+1)}{(\beta-\alpha)^2}+\frac{2k}{\alpha(n-k)+\beta k+\theta}\left(\frac{n}{\beta-\alpha}-\frac{n-k}{\beta}+\frac{k}{\alpha(n-k)+\beta k+\theta}\right)\right.$$
$$\left.-\frac{2n(n-k)}{\beta(\beta-\alpha)}+\frac{(k-n)(k-n+1)}{\beta^2}\right],$$

$$u_{13}=\sum_{k=0}^{n}\frac{S}{\theta}\left[\frac{1}{\alpha(n-k)+\beta k+\theta}\left(-\theta\left(\frac{n}{\beta-\alpha}+\frac{k}{\alpha}\right)+(k-n)\left(1-\frac{2\theta}{\alpha(n-k)+\beta k+\theta}\right)\right)+\frac{n}{(\beta-\alpha)}+\frac{k}{\alpha}\right],$$

$$u_{23}=\sum_{k=0}^{n}\frac{S}{\theta}\left[\frac{1}{\alpha(n-k)+\beta k+\theta}\left(\theta\left(\frac{n}{\beta-\alpha}-\frac{n-k}{\beta}\right)+k\left(\frac{2\theta}{\alpha(n-k)+\beta k+\theta}-1\right)\right)+\frac{n-k}{\beta}\right],$$

$$u_{33}=\sum_{k=0}^{n}\frac{S}{\theta}\left(\frac{\theta}{\alpha(n-k)+\beta k+\theta}-1\right).$$

Hence, we obtain $A=\sigma_{11}L_{111}+2\sigma_{12}L_{121}+\sigma_{22}L_{221}$, $B=\sigma_{11}L_{112}+2\sigma_{12}L_{122}+\sigma_{22}L_{222}$ and $C=\sigma_{33}L_{333}$. Then, Bayes estimator of R, i.e. $\widehat{R}_{Lindley}$, is given as

$$\begin{aligned}\widehat{R}_{Lindley}=R+[u_1a_1+u_2a_2+u_3a_3+a_4+a_5]+\frac{1}{2}[A(u_1\sigma_{11}+u_2\sigma_{12}+u_3\sigma_{13})\\+B(u_1\sigma_{21}+u_2\sigma_{22}+u_3\sigma_{23})+C(u_1\sigma_{31}+u_2\sigma_{32}+u_3\sigma_{33})]\end{aligned}\tag{3.6}$$

where all the parameters are evaluated at MLEs $(\widehat{\alpha},\widehat{\beta},\widehat{\theta})$.

4. Numerical results

In this section, a simulation study and a real-life example are presented to illustrate the obtained estimates.

4.1. Simulation study

In this subsection, we perform a simulation study to compare the performance of two different ML estimates and Bayesian estimates under informative and non-informative priors with respect to mean squared error (MSE) and estimated risk (ER). The ER of θ for the estimate $\widehat{\theta}$, is computed as

$$ER(\theta)=\frac{1}{N}\sum_{i=1}^{N}(\widehat{\theta}_i-\theta_i)^2,$$

under the SE loss function. We have used statistical software R [19] for all computations. The nleqslv package [11] in software R is used to solve the non-linear equations. The point and interval estimates are compared with respect to average lengths (AL) and coverage probabilities (CP) of 95% confidence intervals. All results are obtained based on 2500 replications.

In simulation, we have considered different sample sizes $n=5(5)15$, $m=25(25)100$ for $(\alpha,\beta,\theta)=(8,2,1.5)$ and $n=12(4)20$, $m=10,20,40,60$ for $(\alpha,\beta,\theta)=(12,7,3)$. The biases and MSEs for two different ML estimates, biases and ERs of Bayes estimates under the informative and non-informative priors are presented in Table 1. The first MLE of R is computed by using (2.3) based on the solution of the non-linear equation system given in (3.1) and (3.2). ML estimates of unknown parameters α, β and θ are also obtained from the random sample of exponential distributions as $\widetilde{\alpha}=1/\overline{X}$, $\widetilde{\beta}=1/\overline{Y}$ and $\widetilde{\theta}=1/\overline{T}$. The second MLE of R, called MLE2, is computed by using $\widetilde{\alpha},\widetilde{\beta}$ and $\widetilde{\theta}$ in (2.3). In Bayesian case, the hyperparameters are selected as $a_1=2\alpha$, $b_1=2$, $a_2=2\beta$, $b_2=2$, $a_3=2\theta$, $b_3=2$ for the informative prior and zero for the non-informative case. The AL for the %95 asymptotic confidence interval of R and corresponding CP value are reported in Table 2.

From Table 1, MSE and ERs of the estimates decrease as the sample size increases as expected. We observe that Bayes estimates of R based on the informative prior perform better than that of other estimates except for $n=12$, $m=10$ case. The two ML estimates give nearly same error values but the MLE2 mostly gives better results with a small difference. For this reason, MLE2 can be preferable as an alternative method. It is observed that the AL for each interval decreases as the sample size increases and all the CPs are satisfactory.

4.2. Real data analysis

In this subsection, a real-life data set analysis is presented to illustrate the proposed methods. Nelson [16] considered the graphical methods for analyzing accelerated life test data about times to breakdown of an insulating fluid subjected to various constant elevated test voltages. The number of times to breakdown were observed and saved for each test voltage, and all the data was given in Nelson [16] (Table 1).

An engineer has a parallel system with three components and each component is isolated by using this insulating fluid. This engineer wants to decide what voltage values should be used for these components to minimize the breakdown times. For this purpose, he/she wants to compare the low and mid voltage levels (32 Kv and 34 Kv) against the high voltage level (36 Kv). Then, three components parallel system with standby components is constructed by using 32 Kv and 34 Kv data sets. It is assumed that 32 Kv, 34 Kv and 36 Kv data sets represent the strength component (**X**), standby strength component (**Y**) and stress component (**T**) observations, respectively. For the strength data sets, **X** and **Y** are used for three components of the parallel system. Based on the previous information, if the reliability value of this system exceeds 0.90, he/she prefers the parallel system with standby components.

We use 15 observations for strength components and 5 observations for stress component. Hence, a random sample of the size 15 is taken from 34 Kv data for **Y** and a random sample of the size 5 is taken from 36 Kv data for **T**. Every column of **X** and **Y** data sets represents the observations of the three components for the parallel system. Then, the observed data **X**, **Y** and (**Z**, **T**) for $n=3$, $m=5$ are given as

$$\mathbf{X}=\begin{bmatrix}0.27 & 3.91 & 53.24\\ 0.40 & 9.88 & 82.85\\ 0.69 & 13.95 & 89.29\\ 0.79 & 15.93 & 100.58\\ 2.75 & 27.80 & 215.10\end{bmatrix},\ \mathbf{Y}=\begin{bmatrix}4.15 & 33.91 & 8.27\\ 4.67 & 7.35 & 36.71\\ 72.89 & 0.78 & 2.78\\ 31.75 & 4.85 & 0.19\\ 0.96 & 3.16 & 8.01\end{bmatrix},\ \mathbf{Z}=\begin{bmatrix}4.42 & 37.82 & 61.51\\ 5.07 & 17.23 & 119.56\\ 73.58 & 14.73 & 92.07\\ 32.54 & 20.78 & 100.77\\ 3.71 & 30.96 & 223.11\end{bmatrix},\ \mathbf{T}=\begin{bmatrix}1.97\\ 3.99\\ 0.99\\ 2.58\\ 25.50\end{bmatrix}.$$

We check whether data sets **X**, **Y** and **T** come from the exponential distribution or not. Kolmogorov-Smirnov (K-S), Anderson-Darling (A-D) and Cramer-von Mises (C-VM) tests are carried out for the goodness-of-fit test. Test results are listed in Table 3. It is observed that the exponential distribution provides a good fit to these data sets.

TABLE 1. Estimates of R for $(\alpha,\beta,\theta)=(8,2,1.5)$ and $(12,7,3)$

			MLE			$MLE2$			$Bayes$ (Inf. prior)			$Bayes$ (Non-inf. prior)		
n	m	R	$\widehat{R}$	Bias	MSE	$\widehat{R}$	Bias	MSE	$\widehat{R}$	Bias	ER	$\widehat{R}$	Bias	ER
							$(\alpha,\beta,\theta)=(8,2,1.5)$							
5	25	0.80313	0.80398	0.00084	0.00344	0.80391	0.00078	0.00342	0.79163	-0.011507	0.00318	0.79527	-0.00787	0.00364
	50		0.80270	-0.00043	0.00170	0.80287	-0.00026	0.00169	0.79804	-0.00510	0.00155	0.79866	-0.00448	0.00173
	75		0.80264	-0.00049	0.00114	0.80274	-0.00040	0.00114	0.79972	-0.00342	0.00107	0.80000	-0.00313	0.00115
	100		0.80410	0.00097	0.00088	0.80423	0.00109	0.00088	0.80197	-0.00117	0.00083	0.80213	-0.00100	0.00088
10	25	0.87624	0.87350	-0.00273	0.00227	0.87382	-0.00242	0.00221	0.86402	-0.01222	0.00201	0.86437	-0.01187	0.00252
	50		0.87483	-0.00141	0.00114	0.87523	-0.00101	0.00111	0.87036	-0.00588	0.00105	0.87032	-0.00591	0.00118
	75		0.87564	-0.00060	0.00074	0.87582	-0.00041	0.00073	0.87261	-0.00363	0.00070	0.87263	-0.00361	0.00076
	100		0.87605	-0.00019	0.00056	0.87617	-0.00007	0.00056	0.87376	-0.00248	0.00054	0.87379	-0.00245	0.00057
15	25	0.90685	0.90550	-0.00134	0.00159	0.90606	-0.00079	0.00154	0.89698	-0.00987	0.00141	0.89670	-0.01015	0.00177
	50		0.90531	-0.00153	0.00080	0.90555	-0.00130	0.00079	0.90093	-0.00591	0.00076	0.90084	-0.00600	0.00085
	75		0.90531	-0.00154	0.00055	0.90549	-0.00135	0.00055	0.90237	-0.00448	0.00053	0.90231	-0.00454	0.00058
	100		0.90677	-0.00008	0.00041	0.90677	-0.00007	0.00040	0.90450	-0.00235	0.00039	0.90452	-0.00233	0.00042
							$(\alpha,\beta,\theta)=(12,7,3)$							
12	10	0.78932	0.80325	0.01392	0.00841	0.79725	0.00792	0.00856	0.73402	-0.05530	0.00897	0.76772	-0.02161	0.01178
	20		0.79812	0.00880	0.00427	0.79439	0.00506	0.00429	0.76829	-0.02104	0.00409	0.77702	-0.01231	0.00586
	40		0.79243	0.00310	0.00239	0.79050	0.00117	0.00238	0.77949	-0.00984	0.00238	0.78039	-0.00894	0.00315
	60		0.79170	0.00237	0.00158	0.79061	0.00128	0.00157	0.78344	-0.00589	0.00150	0.78288	-0.00645	0.00201
16	10	0.81363	0.82180	0.00817	0.00801	0.81633	0.00270	0.00821	0.76609	-0.04753	0.00735	0.78599	-0.02763	0.01130
	20		0.81862	0.00499	0.00420	0.81552	0.00189	0.00425	0.79410	-0.01954	0.00349	0.79753	-0.01610	0.00577
	40		0.81374	0.00011	0.00223	0.81211	-0.00152	0.00225	0.80219	-0.01144	0.00215	0.80139	-0.01224	0.00314
	60		0.81514	0.00150	0.00150	0.81416	0.00053	0.00147	0.80773	-0.00590	0.00139	0.80720	-0.00643	0.00182
20	10	0.83059	0.83843	0.00784	0.00666	0.83377	0.00318	0.00677	0.79058	-0.04001	0.00551	0.80129	-0.02930	0.01063
	20		0.83485	0.00426	0.00375	0.83201	0.00142	0.00377	0.81131	-0.01927	0.00340	0.81240	-0.01818	0.00616
	40		0.83167	0.00109	0.00197	0.83044	-0.00014	0.00196	0.82120	-0.00939	0.00173	0.82040	-0.01019	0.00248
	60	0.83059	0.83121	0.00618	0.00138	0.83055	-0.00003	0.00137	0.82407	-0.00651	0.00135	0.82320	-0.00739	0.00177

TABLE 2. Average lengths and coverage probabilities of R for $(\alpha, \beta, \theta) = (8, 2, 1.5)$ and $(12, 7, 3)$

n	m	R	AL	CP	n	m	R	AL	CP
5	25	0.80314	0.22488	0.92200	12	10	0.78933	0.35302	0.87320
	50		0.16184	0.93560		20		0.26235	0.91960
	75		0.13273	0.94360		40		0.19079	0.93040
	100		0.11477	0.93840		60		0.15704	0.93680
10	25	0.87624	0.18283	0.91240	16	10	0.81363	0.33748	0.86760
	50		0.13070	0.93320		20		0.25054	0.90360
	75		0.10691	0.93600		40		0.18285	0.92520
	100		0.09266	0.93520		60		0.14982	0.93240
15	25	0.90685	0.15545	0.90200	20	10	0.83059	0.32468	0.87360
	50		0.11211	0.93240		20		0.24059	0.90520
	75		0.09208	0.94520		40		0.17522	0.93720
	100		0.07921	0.94120		60		0.14416	0.93400

In this case, the MLE of the parameters are obtained as $(\widehat{\alpha}, \widehat{\beta}, \widehat{\theta}) = (0.0183, 0.8230, 0.1427)$ for our model. The estimates of R are listed in Table 4 based on these ML estimates. Bayes estimates of R are obtained under three different priors like as Prior 1: $a_i = b_i = 2$, $i = 1, 2, 3$, Prior 2: $a_1 = 0.5084$, $b_1 = 0.0123$, $a_2 = 0.5539$, $b_2 = 0.0377$, $a_3 = 0.5677$, $b_3 = 0.0810$ and Prior 3: $a_i = b_i = 0$, $i = 1, 2, 3$. The hyperparameters in Prior 2 are obtained by using moment estimation of Gamma distribution based on the data **X**, **Y** and **T**. This method can be preferable when the prior information is not available.

TABLE 3. Goodness-of-fit test for the real data set

Data	*MLE*	$K-S$	$p-value$	$A-D$	$p-value$	$C-VM$	$p-value$
X	$\widehat{\alpha}$ =0.0243	0.3094	0.0900	3.7203	0.0123	0.4030	0.0697
Y	$\widehat{\beta}$ =0.0680	0.3030	0.1020	1.5471	0.1659	0.2900	0.1439
T	$\widehat{\theta}$ =0.1427	0.3658	0.4134	0.7282	0.5276	0.1336	0.4562

TABLE 4. Estimates of R for the real data set

(n, m)	MLE	MLE2	Lindley (Prior 1)	Lindley (Prior 2)	Lindley (Prior 3)
(3,5)	0.99459	0.99801	0.93044	0.91603	0.94869
	(0.98770,1.00)	(0.98721,1.00)			

5. Conclusion

In this paper, the estimation problem has been considered for the stress-strength reliability of the parallel system when the cold standby redundancy available. It is assumed that stress, strength and standby components come from the exponential distribution. Bayes estimate is approximated by using Lindley's approximation under two different priors, and compared with the maximum likelihood estimates. It is observed that when the prior information is available, the estimated risk of Bayes estimate is smaller than risks of the other estimates. When the prior information is not available, the estimated risks of two maximum likelihood and Bayes estimates are closing to each other as sample size increases.

It is known that the convolution of the independent and non-identical random variables generally has mixed forms. Hence, the total lifetime of the strength component and its corresponding standby component for the other lifetime distributions will be more complicated for this reliability problem. We hope to report our results in this regard in the near future.

References

[1] Akgül, F.G. (2019). Reliability estimation in multicomponent stress-strength model for Topp-Leone distribution. *Journal of Statistical Computation and Simulation*, 89(15), 2914-2929.

[2] Birnbaum, Z. (1956). On a use of the Mann-Whitney statistic. *The Regents of the University of California*, Vol. 1. Washington: Univ. of Calif. Press, 13-17.

[3] Birnbaum, Z.W. and McCarty, B.C. (1958). A distribution-free upper confidence bounds for $Pr(Y < X)$ based on independent samples of X and Y. *The Annals of Mathematical Statistics*, 29(2), 558-562.

[4] Boland, P.J. and El-Neweihi, E. (1995). Component redundancy vs system redundancy in the hazard rate ordering. *IEEE Transactions on Reliability*, 44(4), 614-619.

[5] Chen, J., Zhang, Y., Zhao P. and Zhou S. (2018). Allocation strategies of standby redundancies in series/parallel system. *Communications in Statistics-Theory and Methods*, 47(3), 708-724.

[6] Church, J.D. and Harris, B. (1970). The estimation of reliability from stress-strength relationships. *Technometrics*, 12(1), 49-54.

[7] Dey, S., Mazucheli, J. and Anis, M.Z. (2017). Estimation of reliability of multicomponent stress-strength for a Kumaraswamy distribution. *Communications in Statistics -Theory and Methods*, 46(4), 1560-1572.

[8] Eryilmaz, S. (2008). Multivariate stress-strength reliability model and its evaluation for coherent structures. *Journal of Multivariate Analysis*, 99(9), 1878-1887.

[9] Eryilmaz, S. and Tank, F. (2012). On reliability analysis of a two-dependent-unit series system with a standby unit. *Applied Mathematics and Computation*, 218(15), 7792-7797.

[10] Eryilmaz, S. (2017). The effectiveness of adding cold standby redundancy to a coherent system at system and component levels. *Reliability Engineering and System Safety*, 165, 331-335.

[11] Hasselman, B. (2018). Package 'nleqslv', Version 3.3.2, https://cran.r-project.org/web/packages/nleqslv.

[12] Kızılaslan, F. (2018). Classical and Bayesian estimation of reliability in a multicomponent stress-strength model based on a general class of inverse exponentiated distributions. *Statistical Papers*, 59(3), 1161-1192.

[13] Kotz, S., Lumelskii, Y. and Pensky M. (2003). *The Stres-Strength Model and its Generalizations*. Singapore: World Scientific Press.

[14] Lindley, D.V. (1980). Approximate Bayes method. *Trabajos de Estadistica*, 3, 281-288.

[15] Liu, Y., Shi, Y., Bai, X. and Zhan P. (2018). Reliability estimation of a N-M-cold-standby redundancy system in a multicomponent stress-strength model with generalized half-logistic distribution. *Physica A*, 490, 231-249.

[16] Nelson, W.B.(1972). Graphical analysis of accelerated life test data with the inverse power law model. *IEEE Transactions on Reliability*, 21(1), 2-11.

[17] Nojosa, R. and Rathie, P.N. (2020). Stress-strength reliability models involving generalized gamma and Weibull distributions. *International Journal of Quality & Reliability Management*, 37(4), 538-551.

[18] Pakdaman, Z. and Ahmadi, J. (2013). Stress-strength reliability for $P(X_{r:n_1} < Y_{k:n_2})$ in the exponential case. *Istatistik: Journal of the Turkish Statistical Association*, 6(3), 92-102.

[19] R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria; R Foundation for Statistical Computing.

[20] Rao, C.R. (1965). *Linear Statistical Inference and Its Applications*. Wiley, New York.

[21] Roy, A. and Gupta N. (2020). Reliability function of k-out-of-n system equipped with two cold standby components. *Communications in Statistics - Theory and Methods*, DOI: 10.1080/03610926.2020.1737122

[22] Roy, A. and Gupta, N. (2020). Reliability of a coherent system equipped with two cold standby components. *Metrika*, 83(6), 677-697.

[23] Shen, K. and Xie, M. (1991). The effectiveness of adding standby redundancy at system and component levels. *IEEE Transactions on Reliability*, 40(1), 53-55.

[24] Siju, K.C. and Kumar, M. (2017). Estimation of stress-strength reliability of a parallel system with active, warm and cold standby components. *Journal of Industrial and Production Engineering*, 34(8), 590-610.

[25] Tuncel, A. (2017). Residual lifetime of a system with a cold standby unit. *Istatistik Journal of The Turkish Statistical Association,* 10(1), 24-32.

[26] Yan, R., Lu, B. and Li, X. (2019). On Redundancy allocation to series and parallel systems of two components. *Communications in Statistics - Theory and Methods*, 48(18), 4690-4701.

[27] Zhao, P., Zhang, Y. and Li, L. (2015). Redundancy allocation at component level versus system level. *European Journal of Operational Research*, 241(2), 402-411.

VOLUME/CİLT 13 NUMBER/SAYI 2 JULY/TEMMUZ 2021

İSTATİSTİK

JOURNAL OF THE TURKISH STATISTICAL ASSOCIATION

TÜRK İSTATİSTİK DERNEĞİ DERGİSİ

ISSN: 1300-4077