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 $B \in \mathrm{B}$, 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 $D_l \xrightarrow{a.s.} 0$ 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.

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

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

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

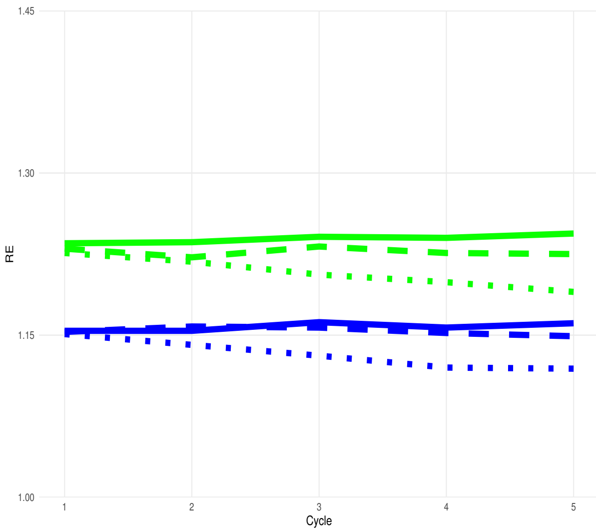
where b and a are upper and lower limits of integral, respectively. L is the number of cut points q_i, $L = \frac{b-a}{w}$ where w is the width of intervals. In the interval $[a,b]$, cut points q_i are obtained by $q_i = a + i(b-a)/L$, $i = 1, \cdots, L$. IMSEs based on sampling designs in RSS and PROS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

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

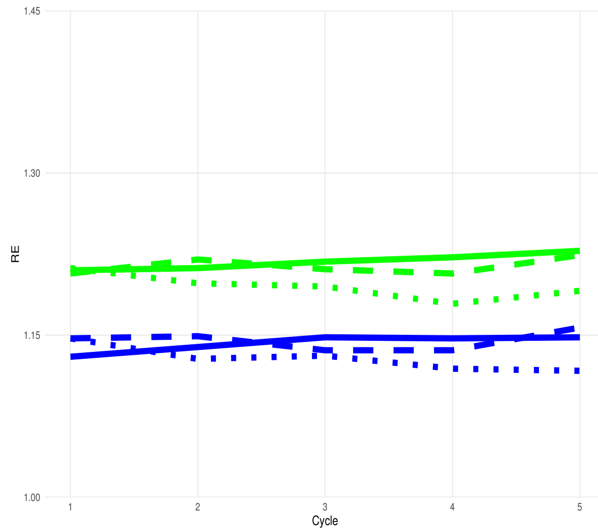
where $\Psi =$ RSS, PROS and $t = 0, 1, 2$. $\hat{F}_{\Psi_{L-t}}(x)$ are calculated by using (3.2) and (3.3), where $x = \{q_1, q_2, \cdots, q_L\}$. In the simulation, the IMSEs of the EDFs based on SRS, RSS and PROS are calculated as taking $w = 0.01$. REs are computed as follow:

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

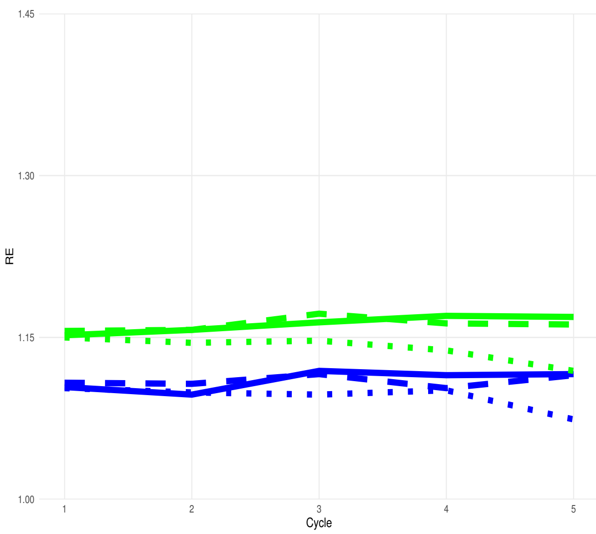
If RE is larger than 1, it can be said that $\hat{F}_{L-t}(x)$ is more efficient than $\hat{F}(x)$. To obtain REs, $10,000$ samples are generated from SRS, RSS and PROS. The REs are illustrated by the Figures 1 and 2 for $\rho = 1$. According to all figures, we can say that the EDFs based on RSS and 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.



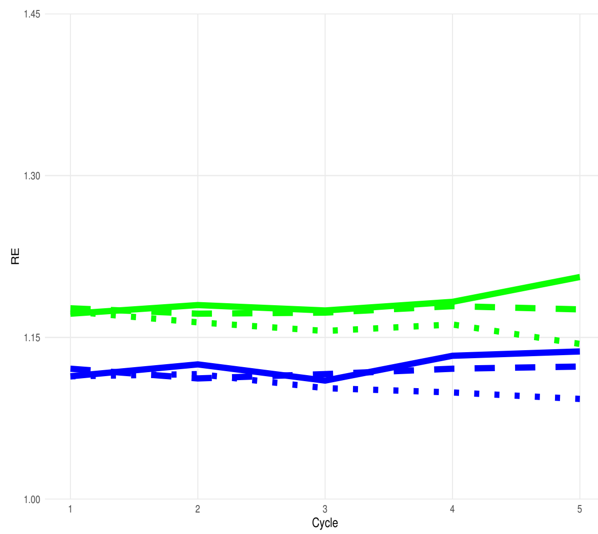
(a) REs for $(g = 0, h = 0)$



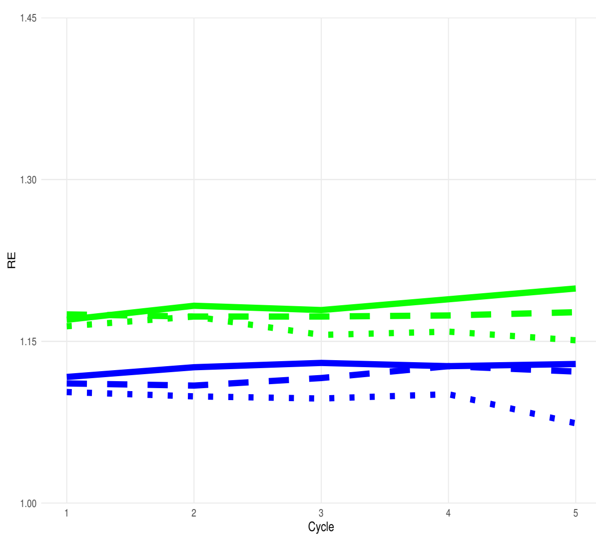
(b) REs for $(g = 0.5, h = 0)$



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

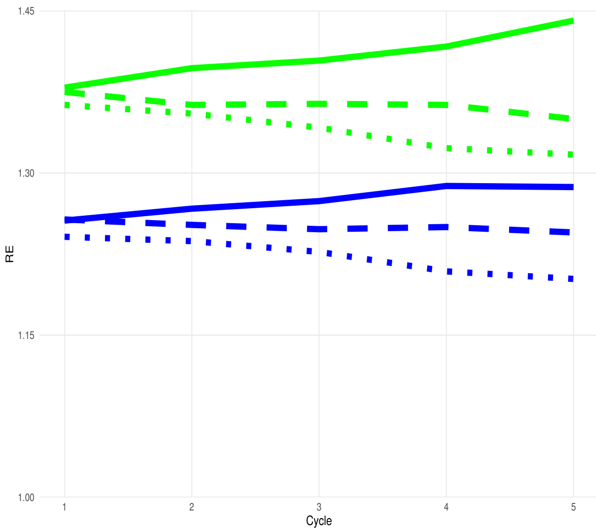


(d) REs for $(g = -0.5, h = 0)$

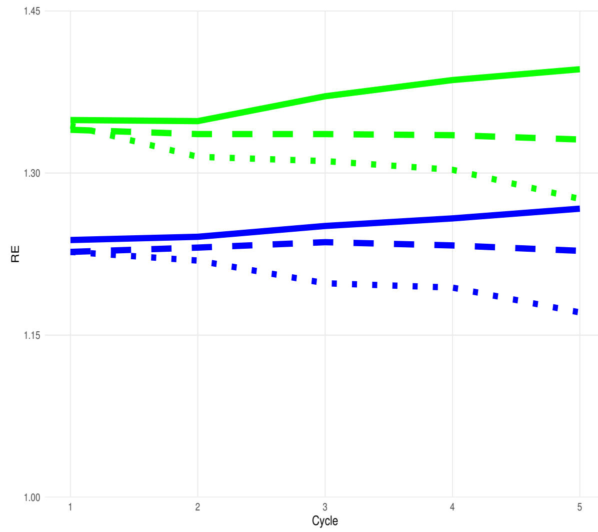


(e) REs for $(g = -1, h = 0)$

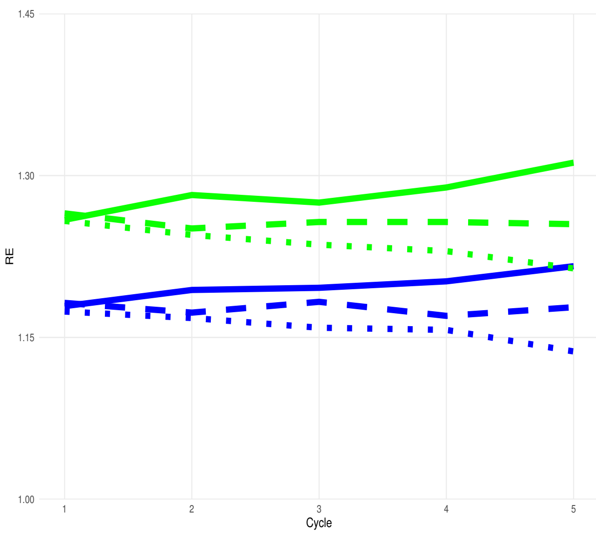
FIGURE 1. REs when $k = H = 3$ (blue: RSS, green: PROS, dotted: Level-0, dashed: Level-1 and solid: Level-2)



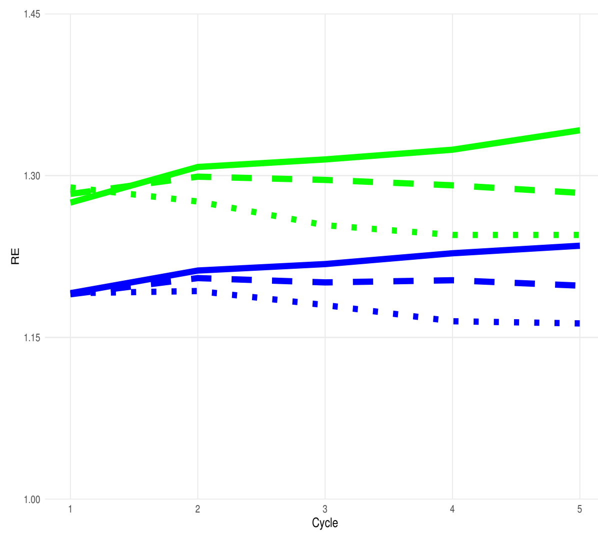
(a) REs for RSS ($g = 0, h = 0$)



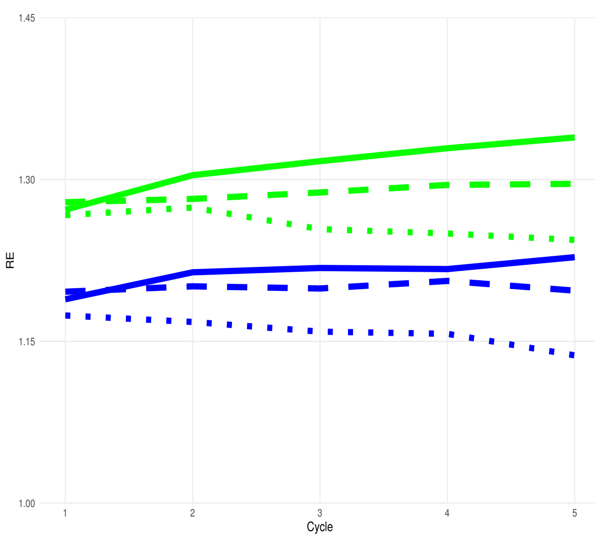
(b) REs for PROS ($g = 0.5, h = 0$)



(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 $\left\{X_{[r:k]i}^{(t)}, r=1,\cdots,k; i=1,\cdots,l\right\}$ and $\left\{X_{[d_v]i}^{(t)}, v=1,\cdots,H; i=1,\cdots,l\right\}$ are Level-t in RSS and PROS, respectively. Moreover, $\zeta_{(1)}^{RSS,t},\cdots,\zeta_{(n)}^{RSS,t}$ and $\zeta_{(1)}^{PROS,t},\cdots,\zeta_{(n)}^{PROS,t}$ 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:

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

- Kuiper test statistic:

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

- Cramér-von Mises test statistic:

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

- Watson test statistic:

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

- Anderson-Darling test statistic:

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

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).

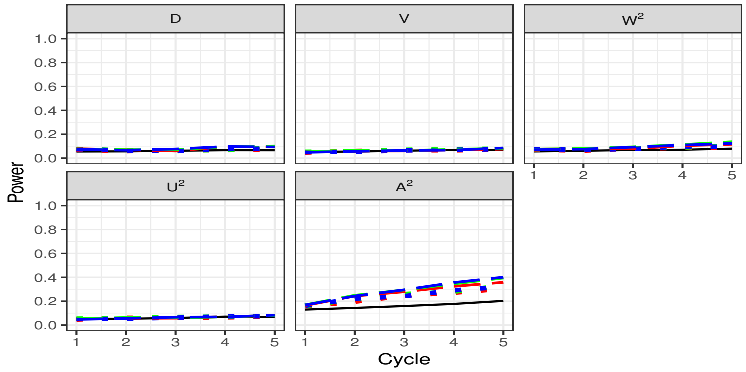
(1) Select a sample using RSS and PROS from an alternative distribution H_1.
(2) Calculate T^{RSS} and T^{PROS} by using the Equations (4.6)-(4.10).
(3) Repeat steps (1)-(2) to get $T_1^{RSS}, \cdots, T_{5,000}^{RSS}$ and $T_1^{PROS}, \cdots, T_{5,000}^{PROS}$.
(4) $Power\, of\, T^{RSS} \approx \frac{1}{5,000} \sum_{t=1}^{5,000} I(T_t^{RSS} > C_{0.05}^{RSS})$ and

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

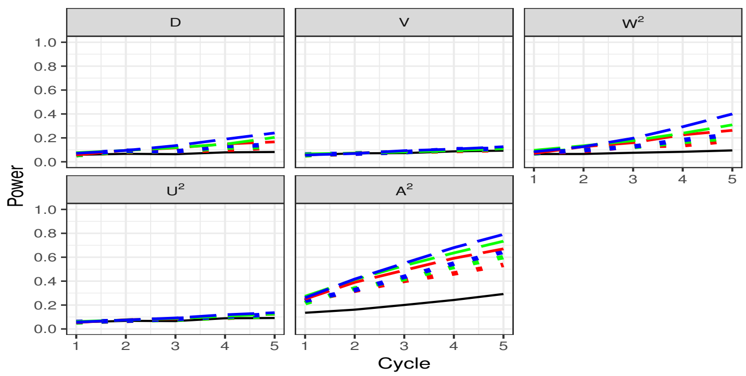
Figure 3-6 include estimated powers of GOF tests based on SRS, RSS and PROS. According to the Figures 3 and 4 (include the powers for $\rho = 1$), it is obviously seen that the best power performance among the GOF tests belong to Anderson-Darling GOF test (A^2) for SRS, RSS and PROS. Also, the highest 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 $\alpha = 0.05$

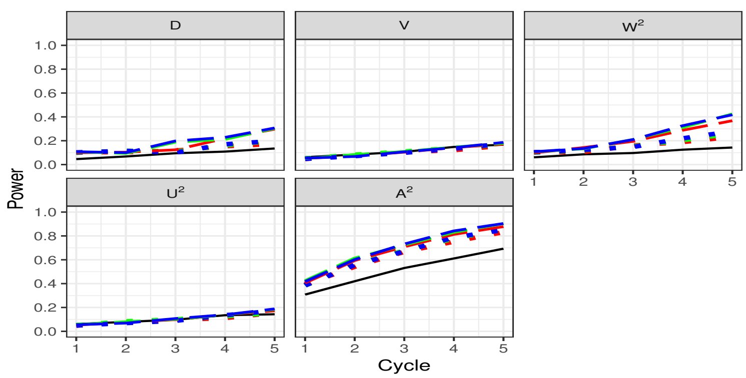
				D		V		W^2		U^2		A^2	
Methods	Designs	k	l	$\rho = 0.25$	$\rho = 1$	$\rho = 0.25$	$\rho = 1$	$\rho = 0.25$	$\rho = 1$	$\rho = 0.25$	$\rho = 1$	$\rho = 0.25$	$\rho = 1$
RSS	Level-0	3	1	0.164	0.052	0.095	0.048	0.174	0.051	0.096	0.048	0.151	0.053
			2	0.132	0.045	0.088	0.051	0.145	0.045	0.090	0.053	0.137	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



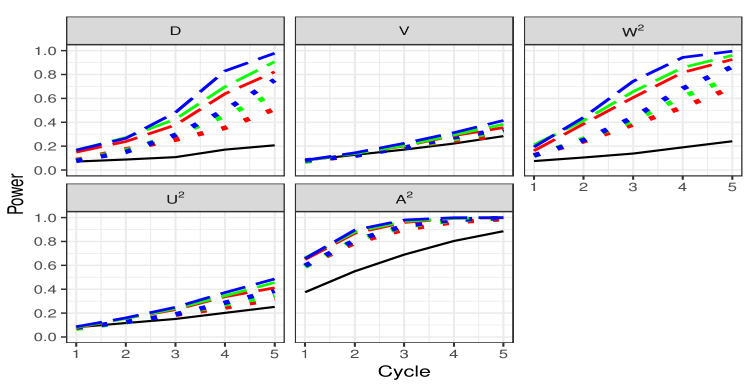
(a) For $k = H = 3$ and $(g = 0.5, h = 0)$



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

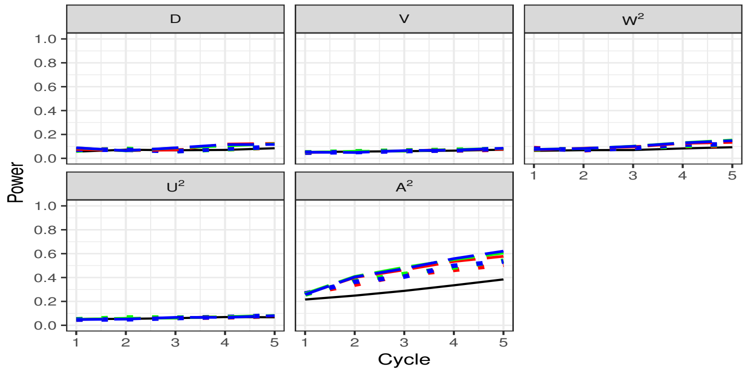


(c) For $k = H = 3$ and $(g = 1, h = 0)$

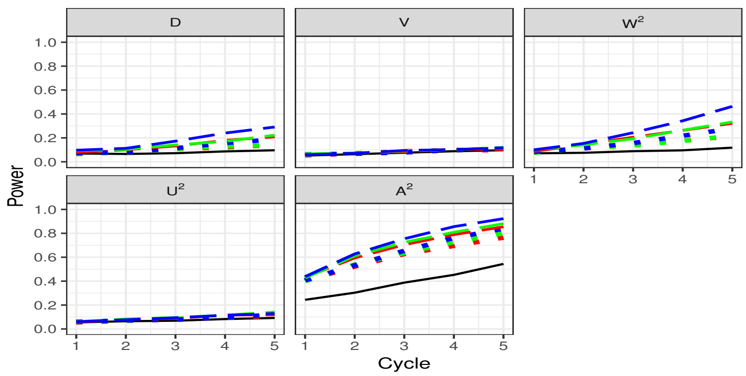


(d) For $k = H = 5$ and $(g = 1, h = 0)$

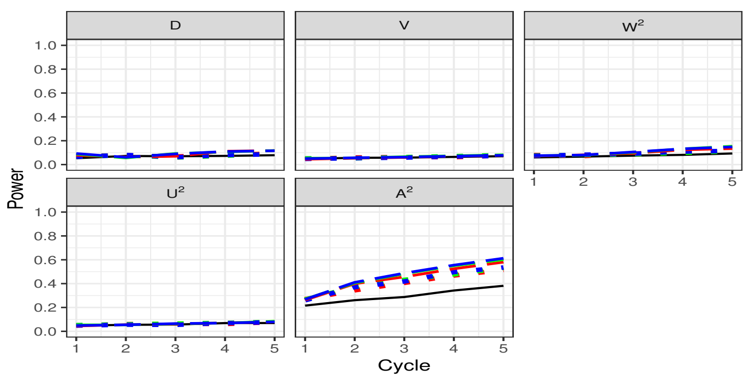
FIGURE 3. The power of GOF tests at $\alpha = 0.05$ for $\rho = 1$, solid: SRS, dotted: RSS and longdash: PROS (For RSS and PROS, red: Level-0, green: Level-1 and blue: Level-2)



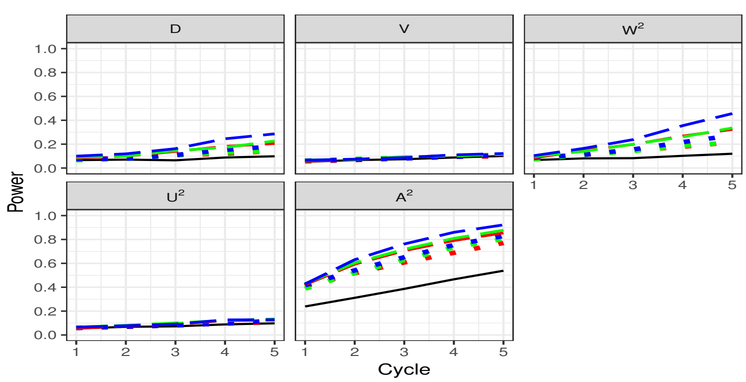
(a) For $k = H = 3$ and $(g = -0.5, h = 0)$



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

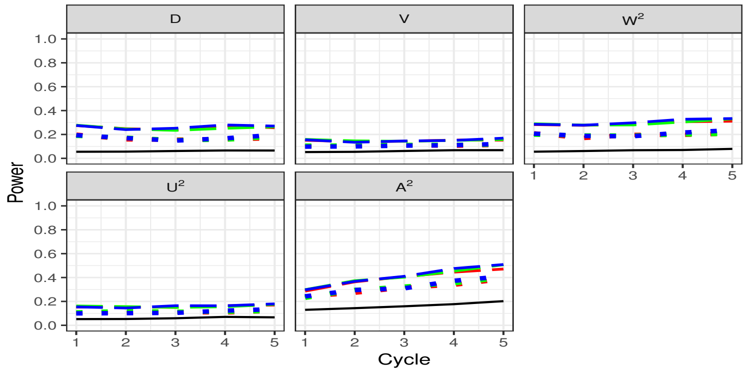


(c) For $k = H = 3$ and $(g = -1, h = 0)$

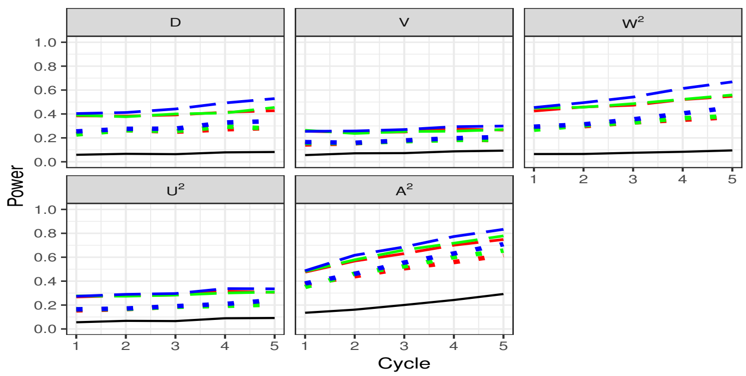


(d) For $k = H = 5$ and $(g = -1, h = 0)$

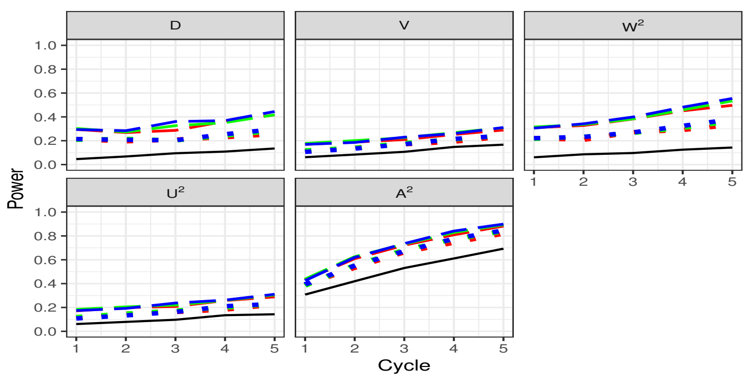
FIGURE 4. The power of GOF tests at $\alpha = 0.05$ for $\rho = 1$, solid: SRS, dotted: RSS and longdash: PROS (For RSS and PROS, red: Level-0, green: Level-1 and blue: Level-2)



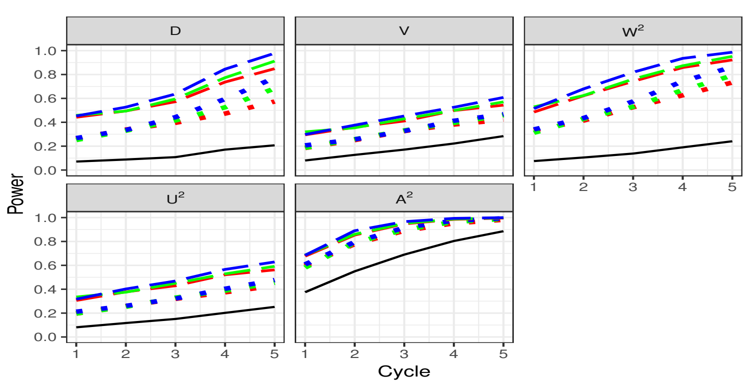
(a) For $k = H = 3$ and $(g = 0.5, h = 0)$



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

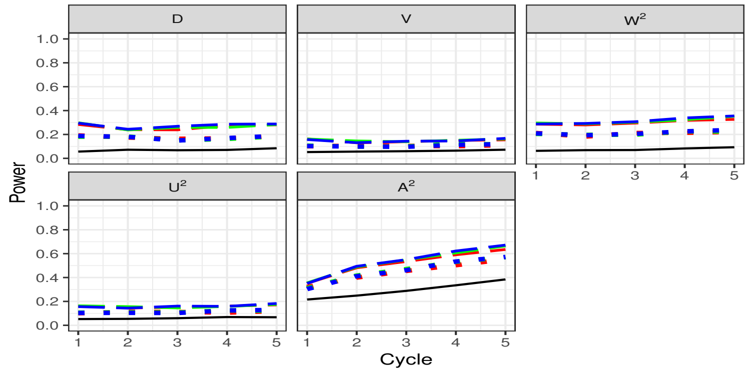


(c) For $k = H = 3$ and $(g = 1, h = 0)$

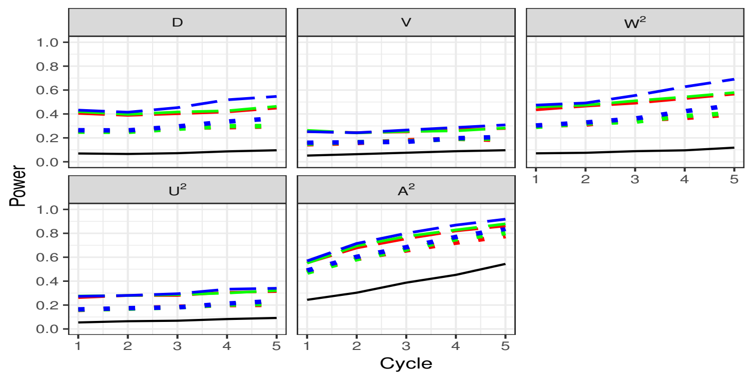


(d) For $k = H = 5$ and $(g = 1, h = 0)$

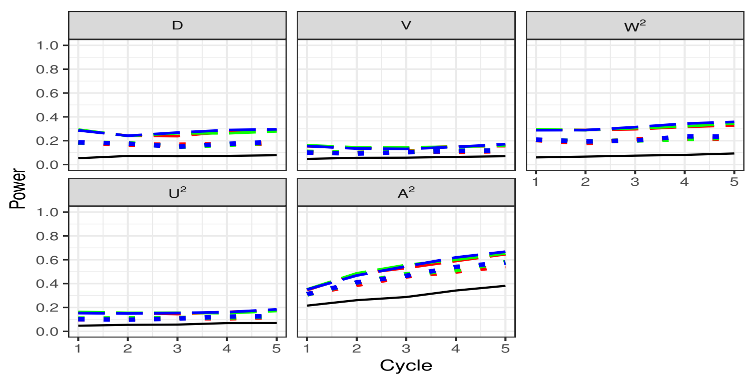
FIGURE 5. The power of GOF tests at $\alpha = 0.05$ for $\rho = 0.25$, solid: SRS, dotted: RSS and longdash: PROS (For RSS and PROS, red: Level-0, green: Level-1 and blue: Level-2)



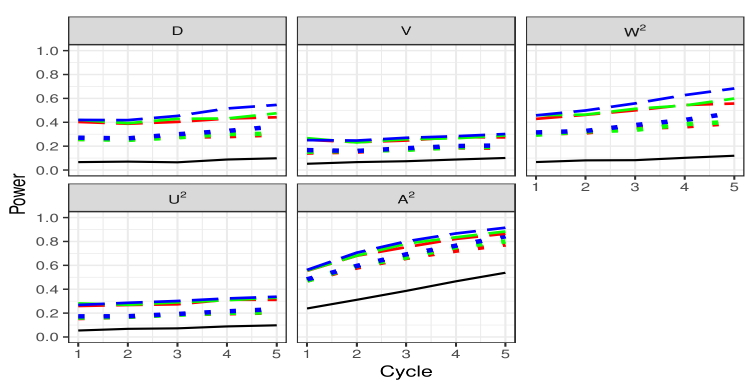
(a) For $k = H = 3$ and $(g = -0.5, h = 0)$



(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)

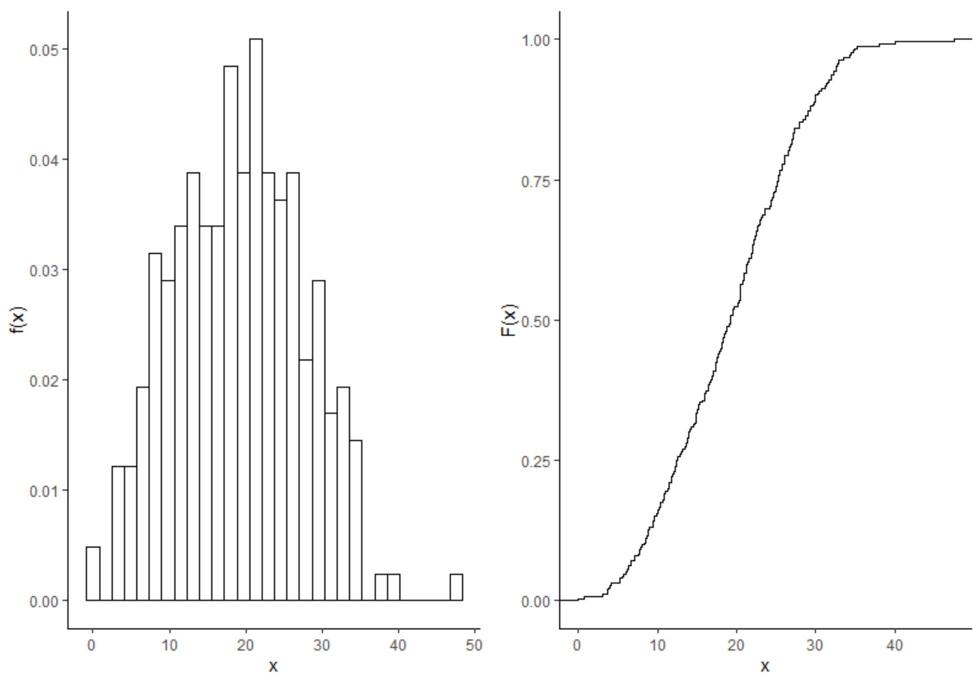


FIGURE 7. The PDF (left) and CDF (right) of the percentage of body fat

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

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