

## Some Modified Mean Estimators in Ranked Set Sampling Using a Covariate

Ehsan Zamanzade\*

*Department of Statistics, University of Isfahan,*

*Isfahan 81746-73441, Iran*

*e.zamnazade@sci.ui.ac.ir; ehsanzamanzadeh@yahoo.com*

Mohammad Mohammadi

*Department of Statistics, University of Isfahan,*

*Isfahan 81746-73441, Iran*

*m.mohammadi@sci.ui.ac.ir*

Received 5 May 2015

Accepted 12 October 2015

### Abstract

In this paper we modify nonparametric regression-based mean estimators in ranked set sampling using a covariate. We will show that our modified mean estimators perform well in comparison with their leading competitors.

**Keywords:** Covariate; Judgment post-stratification; Nonparametric regression; Ranked set sampling.

2000 Mathematics Subject Classification: 62D05; 62G08

### 1. Introduction

Ranked Set Sampling (RSS), which was firstly introduced by McIntyre (1952, 2005), is an appropriate sampling technique which is designed to use auxiliary ranking information. A balanced ranked set sample is selected as follows:

The number of  $k$  samples of size  $k$  is selected, ranked in increasing order without actually measuring them (e.g. by eye inspection or using a covariate), and then the  $i$ th smallest observation from the  $i$ th sample is measured, for  $i=1,2,\dots,k$ . This procedure is repeated  $n$  times to give a ranked set sample of size  $N=nk$ .

Unbalanced RSS differs from balanced RSS in that the number of observations with rank  $i$  varies from one  $i$  to another. Thus, if  $n_i$  be the number of observations with rank  $i$ , then  $N = \sum_{i=1}^k n_i$  is the total sample size. The

standard nonparametric mean estimator in balanced RSS is  $\bar{Y}_{RSS} = \frac{1}{k} \sum_{i=1}^k \bar{Y}_{[i]}$ , where  $\bar{Y}_{[i]}$  is the mean of observations with judgment rank of  $i$ .

Another variation of RSS which is recently proposed by MacEachern et al. (2004) is Judgment Post-Stratification (JPS). To collect a JPS sample of size  $N$ , using set size  $k$ , a simple random sample of size  $N$  is drawn and measured. Then, for each measured unit,  $(k-1)$  additional units are drawn independently to give a set of size  $k$ , and the rank of the measured unit among the  $k$  units is noted. So, JPS data set consists of a simple random sample of size  $N$  and their corresponding ranks. JPS sampling scheme differs from unbalanced RSS in a way that the number of measured samples with rank  $i$  among  $(k-1)$  additional unit, say  $n_i$ , is not

---

\* Corresponding author.

fixed in advance. Indeed if  $(n_1, n_2, \dots, n_k)$  is the vector of the number of observations with rank  $i$ , then one can simply show that  $(n_1, n_2, \dots, n_k) \sim \text{Multinomial}\left(N, \frac{1}{k}, \frac{1}{k}, \dots, \frac{1}{k}\right)$ . The standard JPS mean estimator is

$$\bar{Y}_{JPS} = \frac{1}{\sum_{i=1}^k I(n_i > 0)} \sum_{i=1}^k \bar{Y}_{[i]} I(n_i > 0), \text{ where } \bar{Y}_{[i]} \text{ denotes the mean of observations with rank } i, \text{ and } I(\cdot) \text{ is the indicator function.}$$

Both RSS and JPS are effective sampling strategies over simple random sampling in situations in which obtaining precise values of observations is expensive or time-consuming but ranking them (without actual measurement) is easy and cheap. These situations frequently happen in forestry (Halls and Dell 1996), medicine (Chen et al. 2005), environmental monitoring (Nussbaum and Sinha 1997, Kvam 2003) and entomology (Howard et al. 1982).

The ranking process in RSS and JPS settings can be done based on personal judgment, prior information, or using a covariate. Therefore it is always possible to commit errors in ranking process. Many researchers in RSS and JPS sampling schemes have tried to find the convenient procedures which are robust against errors in rankings. For example, Dell and Clutter (1972) and Stokes and Sager (1988) have showed that empirical mean and distribution function estimators in balanced RSS are unbiased and have less variance than their counterparts in simple random sampling, respectively, even if there are errors in rankings. On the other hand, Bohn and Wolfe (1994) have showed that the Bohn and Wolfe (1992)'s generalization of Mann-Whitney test fails to maintain its level when the assumption of perfect rankings is violated. Frey et al. (2007) and Li and Balakrishnan (2008) developed some tests for assessing the assumption of perfect ranking in RSS, which are followed by Zamanzade et al. (2012) and Zamanzade et al. (2014).

Frey (2011) showed that if ranking observations are done using a covariate, then  $\bar{Y}_{RSS}$ , and  $\bar{Y}_{JPS}$  are not admissible and can be improved by incorporating covariate information. He then proposed some mean estimators, and showed that his proposed mean estimators have less mean square error (MSE) than the standard RSS and JPS mean estimators. Zamanzade and Vock (2015) developed a new variance estimator when RSS and JPS sampling schemes are applied by measuring a covariate.

In this paper we modify some mean estimators proposed by Frey (2011). We then compare our modified mean estimators with their leading competitors via a simulation study.

## 2. Introduction of mean estimators

In this section, we modify some mean estimators proposed by Frey (2011). Let  $Y$ ,  $X$  be the variable of interest and covariate, respectively, and let  $\mathbf{X} = \{X_1, X_2, \dots, X_{kN}\}$  be known vector of covariate. Let  $X_{(1)}^m < X_{(2)}^m < \dots < X_{(N)}^m$  be the  $\mathbf{X}$  values corresponding to measured values of  $Y_1, \dots, Y_N$ .

We can incorporate covariate information into estimating the population mean, by using  $\mu_X = E(Y|X)$ . By identities  $E(Y) = E(\mu_X)$ , and  $Var(Y) = E(Var(Y|X)) + Var(E(Y|X))$ , we can use the estimate of  $E(\mu_X)$  as an estimate for population mean and it is expected that the estimator based on the conditioning on the observed covariate values has less variance than the direct estimator of  $E(Y)$ .

The parameter of  $E(\mu_X)$  can be estimated by the average of  $kN$  estimates of  $E(Y|X_{(i)})$ ,  $i = 1, \dots, kN$ .

Subsequently, we describe different methods for estimating these quantities proposed by Frey (2011), and our modifications one by one.

- ❖ Nonparametric Regression: Suppose we want to estimate  $E(Y|X=x)$ . One technique of estimation is to use the nonparametric kernel regression. Given kernel function  $d(\cdot)$  and bandwidth  $h$ , the

quantity  $E(Y|X=x)$  can be estimated by  $\hat{m}(x, h) = \sum_{j=1}^N \frac{Y_j d\left(\frac{(x - X_{(j)}^m)}{h}\right)}{\sum_{j=1}^N d\left(\frac{(x - X_{(j)}^m)}{h}\right)}$ . Frey (2011) suggested to

use standard normal density as kernel function and  $h_0 = N^{-\frac{1}{4}}$  as the bandwidth and introduced his first mean estimator as  $\bar{Y}_N^{h_0} = \frac{1}{kN} \sum_{i=1}^{kN} \hat{m}_N(X_i, h_0)$ . However, with this choice of kernel function, the

bandwidth  $h_0 = N^{-\frac{1}{4}}$  is not optimum and does not depend on variability in data. So, the resultant estimator may be improved by other choices of bandwidths. Since the regression equation is used for prediction, we propose to choose  $h$  based on Cross-Validation (CV) technique. In CV method, the value of  $h$  is selected such that minimize:  $CV(h) = \frac{1}{N} \sum_{i=1}^N \left( Y_i - \hat{m}^{-i}(X_{(i)}^m, h) \right)^2$ , where  $\hat{m}^{-i}(\cdot)$  is the estimate of regression equation without observation  $(Y_i, X_{(i)}^m)$ . Therefore the value of  $h$  which is selected by CV method can be regarded as the best "predictive" value for  $h$ . We propose to estimate the population mean, by  $\bar{Y}_N^{h_{CV}} = \frac{1}{kN} \sum_{i=1}^{kN} \hat{m}_N(X_i, h_{CV})$ .

If the estimation of  $E(Y|X=x)$  is regarded as density estimation problem, then it can be shown that for  $d(\cdot)$  as the standard normal density, the value of  $h$  which minimizes asymptotic mean square

error is  $h_1 = 1.06\hat{\sigma}n^{-\frac{1}{5}}$ , where  $\hat{\sigma}$  is the sample standard deviation. Details for this choice of bandwidth are given in Silverman (1986) and Hardle (1991). Therefore, we propose our second modification of  $\bar{Y}_N^{h_0}$  as  $\bar{Y}_N^{h_1} = \frac{1}{kN} \sum_{i=1}^{kN} \hat{m}_N(X_i, h_1)$  where  $h_1 = 1.06\hat{\sigma}n^{-\frac{1}{5}}$ , and we take

$\hat{\sigma} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (X_{(i)}^m - \bar{X}^m)^2}$ . It is worth mentioning that although some variance estimators have been proposed by MacEachern et al. (2002), Frey and Feeman (2013), and Zamanzade (2016) for the RSS and the JPS settings, our simulation results indicate that the usage of these estimators instead of the sample variance does not make any improvement in the modified estimators.

- ❖ Monotone Regression: Let  $E(Y|X)$  to be non-decreasing function of  $X$ . Then, it can be shown that  $E(Y|X=X_{(i)}^m)$  is non-decreasing in  $i$  as well. However, the values of  $Y_i$ 's may not be non-decreasing due to sampling noise. One way to resolve this problem is to replace  $Y$  values with its isotonized versions  $Y_i^{iso} = \max_{r \leq i} \min_{s \geq i} \sum_{g=r}^s \frac{Y_g}{s-r+1}$ ,  $i=1,\dots,N$ . In fact,  $Y_i^{iso}$ , which is known as isotonic

regression estimator of  $Y_i$  ( $i = 1, \dots, n$ ), minimizes the mean square  $\sum_{i=1}^N (Y_i - Y_i^{iso})^2$  under constraint

$Y_1^{iso} \leq Y_2^{iso} \leq \dots \leq Y_N^{iso}$ . Based on this idea, Frey (2011) proposed to estimate  $E(Y|X=x)$  by

$$\hat{m}_M(x) = \begin{cases} Y_1^{iso} & x < X_{(1)}^m, \\ \frac{(Y_i^{iso} + Y_{i+1}^{iso})}{2} & X_{(i)}^m < x < X_{(i+1)}^m, i = 1, \dots, N-1, \\ Y_N^{iso} & x > X_{(N)}^m, \\ Y_i^{iso} & x = X_{(i)}^m, \end{cases}$$

and his second estimator of the population mean as  $\bar{Y}_M = \frac{1}{kN} \sum_{i=1}^{kN} \hat{m}_M(X_i)$ . To improve this estimator, we replace  $\hat{m}_M(x)$  by

$$\hat{m}_M^L(x) = \begin{cases} Y_1^{iso} & x < X_{(1)}^m, \\ Y_i^{iso} + \left( \frac{Y_{i+1}^{iso} - Y_i^{iso}}{X_{(i+1)}^m - X_{(i)}^m} \right) (x - X_{(i)}^m) & X_{(i)}^m < x < X_{(i+1)}^m, i = 1, \dots, N-1, \\ Y_N^{iso} & x > X_{(N)}^m, \\ Y_i^{iso} & x = X_{(i)}^m, \end{cases}$$

and propose  $\bar{Y}_M^L = \frac{1}{kN} \sum_{i=1}^{kN} \hat{m}_M^L(X_i)$  as a modification of  $\bar{Y}_M$ .

❖ Monotone Nonparametric Regression: Mukerjee (1988) showed that if we first isotonize the  $Y$  values and then apply nonparametric kernel regression with log concave kernel function on them, the resultant estimator will be monotone. Frey (2011) proposed to use standard normal density as

kernel function and bandwidth  $h_0 = N^{-\frac{1}{4}}$  to estimate population mean via

$$\bar{Y}_{MN}^{h_0} = \frac{1}{kN} \sum_{i=1}^{kN} \hat{m}_{MN}(X_i, h_0), \text{ where } \hat{m}_{MN}(x, h) = \sum_{j=1}^N \frac{Y_j^{iso} d\left(\frac{(x - X_{(j)}^m)}{h}\right)}{\sum_{j=1}^N d\left(\frac{(x - X_{(j)}^m)}{h}\right)}. \text{ The mean estimator can be}$$

modified by determining the value of bandwidth via CV method on  $(Y_i^{iso}, X_{(i)}^m)$ ,  $i = 1, \dots, n$ . The

resultant estimator will be  $\bar{Y}_{MN}^{h_{CV}} = \frac{1}{kN} \sum_{i=1}^{kN} \hat{m}_{MN}(X_i, h_{CV})$ . Furthermore, we propose another

modification of  $\bar{Y}_{MN}^{h_0}$  as  $\bar{Y}_{MN}^{h_1} = \frac{1}{kN} \sum_{i=1}^{kN} \hat{m}_{MN}(X_i, h_1)$ , where  $h_1$  is defined as above.

- ❖ Linear Regression: The mean estimator based on linear regression, which is considered by Yu and Lam (1997), is  $\bar{Y}_L = \frac{1}{kN} \sum_{i=1}^{kN} m_L(X_i)$ , where  $m_L(x) = \bar{Y} + \hat{\beta}(x - \bar{X}_m)$ ,  $\bar{Y}$  is the mean of  $Y_1, \dots, Y_N$  and  $\hat{\beta}$  is the usual slope estimate of linear regression based on  $(Y_i, X_{(i)}^m), i = 1, \dots, N$ .

In the next section, we compare different estimators of mean based on Monte Carlo simulation.

### 3. Simulation study

In this section, we compare the different mean estimators and their modifications by using Monte Carlo simulation with 10000 repetitions. For this purpose, we define the relative efficiency of each estimator as:

$$RE = \frac{\widehat{MSE}(\bar{Y}_{sd})}{\widehat{MSE}(\tilde{Y})},$$

where  $\bar{Y}_{sd}$  is the standard mean estimator, and  $\tilde{Y}$  is one of the aforementioned mean estimators.

We use Dell and Clutter (1972) model of imperfect ranking, assuming that  $(X, Y) \sim BN(0, 0, 1, 1, \rho)$ , and then take  $Y$  and  $e^Y$  as the variable of interest.

The comparisons are made in both balanced RSS and JPS settings for  $(N, k) = (10, 5), (10, 10), (20, 5), (20, 10)$ , and  $\rho = 0, 0.5, 0.75, 0.9, 1$ . These values of  $(N, k)$  are selected such that we can observe the effect of changing sample size while keeping the set size fixed and the effect of changing set size while keeping sample size fixed. The  $\rho$  values are chosen from random rankings ( $\rho = 0$ ) to perfect rankings ( $\rho = 1$ ). The results are given in Tables 1-4.

**Remark 1:** In the simulation study, the value of the bandwidth ( $h$ ) in the CV method is selected from a sequence of values in the interval  $[N^{-\frac{1}{4}}/3, 3N^{-\frac{1}{4}}]$  with steps of 0.01, where  $N^{-\frac{1}{4}}$  is the bandwidth that is used by Frey (2011). It is worth mentioning that  $CV(h)$  can be computed efficiently by using the relation

$CV(h) = \frac{1}{N} \sum_{i=1}^N \left( \frac{Y_i - m(X_i, h)}{1 - H_{ii}} \right)^2$ , where the  $H_{ii}$ 's are the diagonal elements of the hat matrix (see for example, Takezawa, 2006, page 117). Based on an anonymous referee's suggestion, we also report the mean of the predicted values of the bandwidths in the CV method for  $\bar{Y}_N^{h_{cv}}$  and  $\bar{Y}_{MN}^{h_{cv}}$  ( $\hat{h}_{cv}^N, \hat{h}_{cv}^{MN}$ ) in Tables 1-4.

**Table 1.** Estimated relative efficiencies of mean estimators to standard mean estimator in balanced RSS scheme for different set sizes, sample sizes, and correlation between  $X, Y$ , when the target mean is  $E(Y)$ . The values in the parentheses ( $\hat{h}_{CV}^N, \hat{h}_{CV}^{MN}$ ) are the means of the predicted values of the bandwidths in the CV method.

	$\rho$	$\bar{Y}_N^h$	$\bar{Y}_M^h$	$\bar{Y}_M^{h_{CV}}(\hat{h}_{CV}^N)$	$\bar{Y}_M^L$	$\bar{Y}_{MN}^h$	$\bar{Y}_{MN}^{h_{CV}}(\hat{h}_{CV}^{MN})$	$\bar{Y}_L$
$(N, k) = (10, 5)$	0	0.96	0.96	0.95 (1.27)	0.96	0.96	0.98	0.97 (0.55)
	0.50	1.00	1.00	0.98 (0.95)	0.97	0.97	1.01	1.00 (0.36)
	0.75	1.10	1.10	1.07 (0.60)	1.04	1.05	1.11	1.09 (0.33)
	0.90	1.28	1.26	1.24 (0.41)	1.17	1.20	1.28	1.25 (0.31)
	1	1.65	1.60	1.65 (0.26)	1.50	1.64	1.60	1.65 (0.26)
	0	0.96	0.96	0.96 (1.28)	0.96	0.96	0.98	0.98 (0.54)
$(N, k) = (10, 10)$	0.50	1.00	1.01	1.00 (0.96)	0.98	0.98	1.02	1.00 (0.36)
	0.75	1.07	1.07	1.05 (0.60)	1.02	1.03	1.09	1.06 (0.33)
	0.90	1.26	1.24	1.23 (0.41)	1.13	1.18	1.26	1.24 (0.30)
	1	1.84	1.77	1.82 (0.26)	1.50	1.77	1.84	1.77
	0	0.96	0.96	0.95 (1.12)	0.96	0.96	0.98	0.98 (0.36)
$(N, k) = (20, 5)$	0.50	1.01	1.01	0.99 (0.74)	0.98	0.98	1.03	1.01 (0.24)
	0.75	1.12	1.11	1.01 (0.47)	1.05	1.06	1.13	1.10 (0.23)
	0.90	1.32	1.31	1.29 (0.34)	1.23	1.25	1.33	1.31 (0.22)
	1	1.74	1.70	1.74 (0.19)	1.66	1.73	1.74	1.74 (0.19)
$(N, k) = (20, 10)$	0	0.97	0.97	0.97 (1.12)	0.96	0.97	0.98	0.97 (0.36)
	0.50	1.00	1.00	0.98 (0.74)	0.97	0.97	1.01	0.99 (0.25)
	0.75	1.09	1.09	1.07 (0.47)	1.03	1.03	1.10	1.07 (0.23)
	0.90	1.27	1.26	1.25 (0.34)	1.17	1.19	1.27	1.24 (0.22)
	1	1.92	1.90	1.89 (0.19)	1.70	1.86	1.92	1.89 (0.19)

**Table 2.** Estimated relative efficiencies of mean estimators to standard mean estimator in balanced RSS scheme for different set sizes, sample sizes, and correlation between  $X, Y$ , when the target mean is  $E(e^Y)$ . The values in the parentheses ( $\hat{h}_{CV}^N, \hat{h}_{CV}^{MN}$ ) are the means of the predicted values of the bandwidths in the CV method.

		$\rho$	$\bar{Y}_N^h$	$\bar{Y}_N^L$	$\bar{Y}_N^{h_{CV}}(\hat{h}_{CV}^N)$	$\bar{Y}_M^h$	$\bar{Y}_M^L$	$\bar{Y}_{MN}^h$	$\bar{Y}_{MN}^{h_{CV}}(\hat{h}_{CV}^{MN})$	$\bar{Y}_L$
$(N, k) = (10, 5)$	0	0.96	0.96	0.93(1.30)	0.94	0.94	0.99	0.99	0.97(0.56)	0.97
	0.50	1.13	1.13	1.08(1.02)	0.99	1.05	1.18	1.19	1.12(0.37)	1.04
	0.75	1.54	1.55	1.48(0.66)	1.19	1.37	1.57	1.58	1.48(0.34)	1.19
	0.90	2.08	2.11	2.03(0.44)	1.50	1.89	2.10	2.12	2.03(0.32)	1.38
	1	2.67	2.71	2.77(0.27)	1.38	2.44	2.67	2.71	2.77(0.27)	1.56
	0	0.97	0.97	0.97(1.31)	0.94	0.94	1.00	1.00	0.98(0.54)	0.98
$(N, k) = (10, 10)$	0.50	1.18	1.19	1.15(1.02)	1.03	1.10	1.24	1.24	1.18(0.37)	1.10
	0.75	1.51	1.54	1.45(0.66)	1.09	1.31	1.54	1.56	1.44(0.33)	1.31
	0.90	2.14	2.17	2.05(0.44)	1.30	1.83	2.14	2.18	2.05(0.31)	1.39
	1	2.88	2.89	3.11(0.27)	1.53	2.77	2.88	2.88	3.11(0.27)	2.77
	0	0.97	0.97	0.94(1.14)	0.96	0.96	1.00	1.00	0.98(0.35)	0.99
$(N, k) = (20, 5)$	0.50	1.15	1.17	1.12(0.84)	0.99	1.03	1.20	1.22	1.11(0.26)	1.05
	0.75	1.52	1.55	1.48(0.55)	1.18	1.33	1.55	1.58	1.45(0.25)	1.19
	0.90	1.95	1.98	1.93(0.38)	1.49	1.81	1.96	1.99	1.96(0.25)	1.31
	1	2.32	2.33	2.61(0.21)	1.85	2.50	2.32	2.33	2.61(0.21)	1.50
	0	0.97	0.97	0.96(1.14)	0.96	0.97	1.01	1.01	0.99(0.36)	0.99
$(N, k) = (20, 10)$	0.50	1.15	1.16	1.13(0.84)	0.98	1.02	1.19	1.21	1.11(0.26)	1.04
	0.75	1.46	1.49	1.41(0.54)	1.12	1.25	1.49	1.52	1.39(0.26)	1.12
	0.90	1.96	1.97	1.92(0.38)	1.43	1.78	1.97	1.96	1.95(0.25)	1.28
	1	2.43	2.28	2.98(0.22)	1.95	2.90	2.43	2.28	2.98(0.22)	1.52

**Table 3.** Estimated relative efficiencies of mean estimators to standard mean estimator in JPS sampling scheme for different set sizes, sample sizes, and correlation between  $X, Y$ , when the target mean is  $E(Y)$ . The values in the parentheses ( $\hat{h}_{CV}^N, \hat{h}_{CV}^{MN}$ ) are the means of the predicted values of the bandwidths in the CV method.

		$\rho$	$\bar{Y}_N^h$	$\bar{Y}_N^h$	$\bar{Y}_N^{h_{CV}}\left(\hat{h}_{CV}^N\right)$	$\bar{Y}_M^L$	$\bar{Y}_M^L$	$\bar{Y}_M^{h_{CV}}\left(\hat{h}_{CV}^{MN}\right)$	$\bar{Y}_L$
$(N, k) = (10, 5)$	0	0	1.16	1.16	1.16(1.26)	1.10	1.18	1.18	1.14(0.54)
	0.50	0.50	1.26	1.26	1.20(0.96)	1.22	1.32	1.32	1.27(0.35)
	0.75	0.75	1.47	1.44	1.39(0.60)	1.43	1.45	1.53	1.50(0.32)
	0.90	0.90	1.80	1.73	1.82(0.41)	1.83	1.90	1.84	1.78
	1	1	2.23	2.10	2.63(0.25)	2.47	2.68	2.23	2.10
								2.63(0.25)	3.41
$(N, k) = (10, 10)$	0	0	1.09	1.10	1.08(1.26)	1.03	1.03	1.10	1.10
	0.50	0.50	1.25	1.24	1.18(0.96)	1.21	1.21	1.32	1.31
	0.75	0.75	1.60	1.56	1.50(0.60)	1.56	1.58	1.68	1.64
	0.90	0.90	2.17	2.07	2.21(0.41)	2.24	2.34	2.24	2.15
	1	1	3.55	3.21	4.65(0.25)	4.06	4.71	3.55	3.21
								4.65(0.25)	7.39
$(N, k) = (20, 5)$	0	0	1.20	1.21	1.21(1.11)	1.14	1.14	1.21	1.17(0.35)
	0.50	0.50	1.25	1.25	1.21(0.75)	1.20	1.20	1.29	1.30
	0.75	0.75	1.40	1.38	1.36(0.47)	1.35	1.37	1.44	1.42
	0.90	0.90	1.62	1.56	1.63(0.34)	1.60	1.63	1.66	1.59
	1	1	1.99	1.84	2.26(0.19)	2.18	2.28	1.99	1.84
								2.26(0.19)	2.52
$(N, k) = (20, 10)$	0	0	1.22	1.22	1.22(1.11)	1.14	1.15	1.21	1.22
	0.50	0.50	1.33	1.33	1.28(0.75)	1.27	1.28	1.38	1.33(0.24)
	0.75	0.75	1.57	1.54	1.52(0.47)	1.51	1.53	1.63	1.60(0.23)
	0.90	0.90	2.08	1.97	2.09(0.34)	2.02	2.07	2.13	2.03
	1	1	3.45	2.98	4.34(0.19)	3.95	4.36	3.45	2.98
								4.34(0.19)	5.39

**Table 4.** Estimated relative efficiencies of mean estimators to standard mean estimator in JPS sampling scheme for different set sizes, sample sizes, and correlation between  $X, Y$ , when the target mean is  $E(e^Y)$ . The values in the parentheses ( $\hat{h}_{CV}^N, \hat{h}_{CV}^{MN}$ ) are the means of the predicted values of the bandwidths in the CV method.

		$\rho$	$\bar{Y}_N^h$	$\bar{Y}_N^h$	$\bar{Y}_N^{h_{CV}}(\hat{h}_{CV}^N)$	$\bar{Y}_M^L$	$\bar{Y}_M^L$	$\bar{Y}_{MN}^{h_{CV}}(\hat{h}_{CV}^{MN})$	$\bar{Y}_L$
$(N, k) = (10, 5)$	0	1.18	1.17	1.14(1.29)	1.06	1.07	1.20	1.13(0.56)	1.13
	0.50	1.47	1.45	1.36(1.02)	1.21	1.27	1.55	1.41(0.36)	1.36
	0.75	1.91	1.88	1.77(0.67)	1.54	1.71	1.99	1.88(0.33)	1.54
	0.90	2.67	2.63	2.61(0.43)	1.89	2.50	2.71	2.70(0.31)	1.82
	1	3.15	3.06	3.52(0.26)	2.26	3.28	3.15	3.06	3.52(0.26)
	0	1.09	1.08	1.05(1.29)	0.99	0.99	1.10	1.10	1.04(0.53)
$(N, k) = (10, 10)$	0.50	1.44	1.43	1.30(1.03)	1.25	1.32	1.54	1.41(0.36)	1.33
	0.75	2.00	1.96	1.83(0.66)	1.66	1.85	2.11	2.11	1.98(0.33)
	0.90	2.77	2.69	2.73(0.43)	1.99	2.65	2.82	2.75	2.85(0.31)
	1	3.48	3.34	4.06(0.26)	2.69	3.85	3.48	3.34	4.06(0.26)
	0	1.21	1.22	1.21(1.14)	1.13	1.14	1.23	1.23	1.21(0.35)
	0.50	1.49	1.52	1.46(0.84)	1.27	1.32	1.56	1.59	1.42(0.26)
$(N, k) = (20, 5)$	0.75	1.79	1.81	1.69(0.54)	1.45	1.56	1.84	1.88	1.70(0.25)
	0.90	2.14	2.15	2.12(0.37)	1.77	2.05	2.17	2.17	2.18(0.24)
	1	2.70	2.63	3.17(0.21)	2.30	3.07	2.70	2.63	3.17(0.21)
	0	1.19	1.20	1.19(1.13)	1.10	1.10	1.21	1.16(0.34)	1.20
$(N, k) = (20, 10)$	0.50	1.49	1.49	1.41(0.84)	1.24	1.28	1.57	1.58	1.39(0.25)
	0.75	2.00	2.00	1.87(0.54)	1.54	1.73	2.07	2.07	1.91(0.25)
	0.90	2.63	2.58	2.61(0.37)	2.06	2.55	2.67	2.62	2.74(0.24)
	1	3.20	3.04	4.08(0.21)	2.64	3.96	3.20	3.04	4.08(0.21)

Tables 1-2 give the results for balanced RSS. Table 1 gives the results when the parameters of interest is  $E(Y)$ . It is evident from this table that, when  $E(Y)$  is the target mean, then  $\bar{Y}_L$  is the best estimator, although its superiority is negligible for  $\rho \leq 0.5$ . Furthermore, there are no considerable improvements in the modified estimators in this case.

Table 2 presents the results when  $E(e^Y)$  is the target mean. In this case, two estimators  $\bar{Y}_N^{h_{CV}}, \bar{Y}_{MN}^{h_{CV}}$  are the best when the rankings are perfect ( $\rho = 1$ ), but they are overtaken by  $\bar{Y}_N^{h_0}, \bar{Y}_N^{h_1}, \bar{Y}_{MN}^{h_0}$ , and  $\bar{Y}_{MN}^{h_1}$  when  $\rho = 0.5, 0.75, 0.9$ , which shows that the estimators based on the CV method are more sensitive to imperfect rankings than other estimators. Moreover,  $\bar{Y}_N^{h_1}, \bar{Y}_{MN}^{h_1}$ , and  $\bar{Y}_M^L$  improve  $\bar{Y}_N^{h_0}, \bar{Y}_{MN}^{h_0}$ , and  $\bar{Y}_M^L$ , respectively, and improvement of  $\bar{Y}_M^L$  over  $\bar{Y}_M$  for  $\rho \geq 0.75$  is considerable.

We also observe from Tables 1-2 that the means of the predicted values of the bandwidths in the CV technique ( $\hat{h}_{CV}^N$  and  $\hat{h}_{CV}^{MN}$ ) decreases as the value of  $\rho$  increases. Furthermore,  $\hat{h}_{CV}^N$  in the nonparametric regression is larger than its counterpart in the monotone nonparametric regression.

Tables 3-4 present the results for JPS sampling scheme. The pattern in these tables is similar to Tables 1-2. Table 3 gives the results when the target mean is  $E(Y)$ . In the case of  $E(Y)$  being the parameter of interest,  $\bar{Y}_L$  is the best estimator for  $\rho \geq 0.75$ . For other values of  $\rho$ , none of the estimators can dominate the others, but  $\bar{Y}_L$  and  $\bar{Y}_{MN}^{h_1}$  tend to perform better than the others.

It is evident from Table 4 that if  $E(e^Y)$  to be the parameter of interest, then  $\bar{Y}_{MN}^{h_{CV}}$  is the best estimator in case of nearly perfect rankings ( $\rho \geq 0.9$ ). However, this estimator is more sensitive to imperfect rankings and is not the best any more when  $\rho \leq 0.75$ . For  $\rho \leq 0.75$ , two estimators  $\bar{Y}_{MN}^{h_0}$ , and  $\bar{Y}_{MN}^{h_1}$  are almost the best.

The patterns of the mean of the predicted value of the bandwidths in CV technique in JPS setting are similar to RSS setting. The  $\hat{h}_{CV}^N$ , and  $\hat{h}_{CV}^{MN}$  are decreasing in  $\rho$ , and  $\hat{h}_{CV}^{MN}$  is larger than  $\hat{h}_{CV}^N$ , except in the case of perfect ranking, where  $\hat{h}_{CV}^N$  and  $\hat{h}_{CV}^{MN}$  are the same. We also observe that  $\hat{h}_{CV}^N$  and  $\hat{h}_{CV}^{MN}$  in the JPS setting are larger than their counterpart in RSS.

It is interesting to note that, for all considered values of sample size, set size, correlation  $\rho$  and both sampling scheme RSS and JPS,  $\bar{Y}_M^L$  is always superior to  $\bar{Y}_M$ , and this superiority becomes more considerable in the case of perfect rankings. Furthermore, we observe that the estimators based on CV method perform well, when the rankings are nearly perfect, but they are more sensitive to imperfect rankings than the others.

#### 4. Conclusion

We modified some nonparametric mean estimators proposed by Frey (2011). Our simulation results indicate that our modified estimators perform well in comparison with their leading competitors.

**Acknowledgments:** The authors are thankful to Professor Jesse Frey for sharing his R code, and two anonymous referees for their valuable comments and suggestions which improved an earlier version of this paper.

## References

- [1] McIntyre, G. A. A method for unbiased selective sampling using ranked set sampling. *Austral. J. Agricultural Res.* 3, (1952) 385–390.
- [2] McIntyre, G. A. A method for unbiased selective sampling using ranked set sampling. *The American Statistician* 59, (2005), 230–232. Originally appeared in *Australian journal of Agricultural Research* 3, 385-390.
- [3] MacEachern, S.N., Stasny, E.A., and Wolfe, D.A. Judgment Post-Stratification with imprecise rankings. *Biometrics*, 60(1), (2004), 207-215.
- [4] Halls, L. K., Dell, T. R. Trial of ranked-set sampling for forage yields. *Forest Science*, 12, (1996), 22–26.
- [5] Chen, H., Stasny, E. A., Wolfe, D. A.. Ranked set sampling for efficient estimation of a population proportion. *Statistics in Medicine*, 24, (2005) 3319–3329.
- [6] Nussbaum, B.D., and Sinha, B.K. Cost effective gasoline sampling using ranked set sampling, *Proceedings of the Section on Statistics and the Environment*, (1997). 83–87, American Statistical Association.
- [7] Kvam, P. H. Ranked set sampling based on binary water quality data with covariates. *Journal of Agricultural, Biological, and Environmental Statistics*, 8, (2003). 271–279.
- [8] Howard, R. W., Jones, S. C., Mauldin, J. K., Beal, R. H. Abundance, distribution, and colony size estimates for *Reticulitermes* spp. (Isopter: Rhinotermitidae) in Southern Mississippi. *Environmental Entomology*, 11, (1982). 1290–1293.
- [9] Dell, T. R., and Clutter, J. L. Ranked set sampling theory with order statistics background. *Biometrics*, 28(2), (1972). 545-555.
- [10] Stokes, S. L., and Sager, T. W. Characterization of a Ranked-Set sample with application to estimating distribution functions, *Journal of the American Statistical Association*, 83, (1988). 374-381.
- [11] Bohn, L. L., and Wolfe, D. A. The effect of imperfect judgment rankings on properties of procedures based on the Ranked-Set samples analog of the Mann-Whitney-Wilcoxon statistic, *Journal of the American Statistical Association*, 89, (1994). 168-176.
- [12] Bohn, L. L., and Wolfe, D. A. Nonparametric Two-Sample procedures for Ranked-Set samples data. *Journal of the American Statistical Association*, 87, (1992). 552-561.
- [13] Frey, J., Ozturk, O., and Deshpande, J.V. Nonparametric tests for perfect judgment rankings. *Journal of the American Statistical Association* 102, (2007). 708–717.
- [14] Li, T. and Balakrishnan, N. Some simple nonparametric methods to test for perfect ranking in ranked set sampling. *Journal of Statistical Planning and Inference* 138, (2008). 1325–1338.
- [15] Zamanzade, E. , Arghami, N.R., Vock, M. Permutation-based tests of perfect ranking, *Statistics & Probability Letters*, 82, (2012). 3313-2220.
- [16] Zamanzade, E. , Arghami, N.R., Vock, M. A parametric test of perfect ranking in balanced ranked set sampling, *Communications in Statistics-Theory and Methods*. 43:21. (2014). 4589-4611.
- [17] Frey, J. A note on ranked-set sampling using a covariate. *Journal of Statistical Planning and Inference* 141,(2), (2011) 809-816.
- [18] Zamanzade, E. , Vock, M. Variance estimation in ranked set sampling using a concomitant variable. *Statistics & Probability Letters*, 105, (2015) 1-5.
- [19] Silverman, B.W. *Density Estimation for Statistics and Data Analysis*, (1986). Chapman and Hall, London.
- [20] Hardle, W. *Smoothing Techniques. With Implementations in S*. (1991). Springer. New York.
- [21] MacEachern, S.N., Ozturk, O., Wolfe, D.A. and Stark, G.V. A new ranked set sample estimator of variance. *Journal of the Royal Statistical Society: Series B*. 64(2), (2002). 177-188.
- [22] Frey, J., and Feeman, T. G. Variance estimation using judgment post-stratification. *Annals of the Institute of Statistical Mathematics*, 65(3), .(2013). 551-569.
- [23] Zamanzade, E. An isotonized variance estimator for judgement post stratified data, *Journal of the Korean Statistical Society*, 45(1), (2016), 25-32.
- [24] Yu, P.L.H. and Lam, K. Regression estimator in ranked set sampling. *Biometrics* 53 (3), (1997) 1070-1080