

Some nonparametric tests of perfect judgment ranking for judgment post stratification

Ehsan Zamanzade^{a,*} and Michael Vock^b

^a*Department of Statistics, University of Isfahan, Isfahan, 81746-73441, Iran.*

^b*Institute of Mathematical Statistics and Actuarial Science, University of Bern,
Sidlerstrasse 5, CH-3012 Bern, Switzerland.*

March 17, 2017

Abstract

We develop some nonparametric tests of perfect judgment ranking for judgment post stratification sampling scheme. We show that the best proposed test beats the best existing nonparametric test of perfect judgment ranking in ranked set sampling applied to the judgment post stratification case by conditioning on the observed stratum sizes.

Keywords: Judgment ranking; Judgment post stratification; Ranked set sampling; Test.

Mathematics Subject Classifications: 62D05; 62F03.

1 Introduction

In many biological, environmental or ecological problems, there are situations in which ranking population units without actually measuring them (e.g. by eye inspection or using a concomitant

*Corresponding author.

E-mail addresses: e.zamanzade@sci.ui.ac.ir; ehsanzamanzadeh@yahoo.com (E. Zamanzade), michael.vock@stat.unibe.ch (M. Vock)

variable) is easy or cheap as compared with obtaining their precise values. The main motivation behind ranked set sampling (RSS) and judgment post stratification (JPS) sampling designs is to use that ranking information to obtain improved statistical inferences as compared to simple random sampling (SRS).

The RSS procedure, which is first proposed by McIntyre (1952), can be described as follows. One first draws m^2 units from the population and then randomly divides them into m samples, each of size m . One then ranks each sample of size m from the smallest to largest by using any inexpensive methods, such as eye inspection or using a concomitant variable, which does not require measuring the units. Finally, one actually measures the i th judgment ordered unit from the i th sample (for $i = 1, \dots, m$). One repeats this procedure n times to obtain a balanced ranked set sample of size $N = nm$. Here, the term *judgment ranking* indicates that the ranking process is done without actually measuring the sample units, and thus it may contain errors. However, if the judgment ranking is perfect (i.e. without error), then the N units are independent order statistics, otherwise, they are independent judgment order statistics.

To draw an unbalanced ranked set sample, one relaxes the requirement that the number of judgment order statistics with rank i should be the same for $i = 1, \dots, m$. Instead, one first determines the set size m and the vector $\mathbf{n} = (n_1, \dots, n_m)$, where n_i is the number of measured units with judgment rank i (for $i = 1, \dots, m$). Then, one draws n_i random sample(s) of size m and ranks each sample in increasing magnitude without obtaining their precise values. Finally, one measures the i th judgment ordered unit in each sample. The total sample size is $N = \sum_{i=1}^m n_i$.

Although McIntyre (1952) proposed the RSS design to improve the estimation of the mean of the pasture yield, since then, it has been widely developed for many statistical problems. Virtually all of the most important and practical statistical problems are well addressed in the RSS literature. For example, the problem of estimation of a distribution function has been considered by Stokes and Sager (1988), Kvam and Samaniego (1994), and Duembgen and Zamanzade (2013). Stokes (1980), MacEachern et al. (2002), Perron and Sinha (2004) and Zamanzade and Vock (2015) considered the problem of nonparametric estimation of variance. Ratio estimator for the population mean using ranked set sampling has been addressed by Kadilar et al. (2009), and Li et al.

(2012) analysed rounded ranked set sample data.

Another variation of RSS is the judgment post stratification (JPS) sampling scheme, which was recently introduced by MacEachern et al. (2004). To draw a JPS sample of size N , using set size m , one first draws a simple random sample of size N . Then, for each unit in the simple random sample of size N , a supplemental $m - 1$ independent units are drawn to create a set of size m . This set of size m is ranked from smallest to largest without obtaining the precise values of the units (e.g. by eye inspection or using a concomitant variable). Finally, the rank of the unit from the initial simple random sample with respect to the set of size m is noted. Thus, a JPS sample of size N consists of a simple random sample of size N with their corresponding judgment ranks.

There are two main differences between the RSS and the JPS designs. The first difference is about the role of the ranking process in these sampling schemes. In the RSS design, the ranking process, in each set of size m , is done prior to selecting the unit for measurement, and it determines this selection. Therefore, the rank of each observation in the RSS design is essential for the analysis. Thus, statistical methods for simple random samples cannot be applied for ranked set samples, and they can only be analysed by methods that are specially designed for them. However, since in the JPS design, the ranking process is done after measuring all units in the simple random sample, the rank of each unit is additional information that does not necessarily have to be used. Therefore, the researcher can simply ignore the ranks and treat the JPS sample as a simple random sample. This is very useful for situations in which the researcher believes that the ranking process is poor. The second difference between the RSS and the JPS is about the number of observations with judgment rank i , say n_i . In fact, while the vector of $\mathbf{n} = (n_1, \dots, n_m)$ is a fixed and predetermined vector in the RSS setting, it is a random vector in the JPS setting and follows a multinomial distribution with N trials and probability vector $(\frac{1}{m}, \dots, \frac{1}{m})$. Thus, due to random variations of the vector \mathbf{n} , the JPS sampling scheme tends to be less efficient than a balanced RSS scheme with the same set size.

The JPS sampling scheme has drawn attention in recent years. Frey and Feeman (2012) showed that the standard mean estimator in the JPS setting is inadmissible under mean square loss function, they then proposed an alternate mean estimator. Another nonparametric mean esti-

mator is proposed by Wang et al. (2008), by imposing an order constraint on judgment sample means. Frey and Feeman (2013) proposed a conditionally unbiased variance estimator assuming that the post stratum sample sizes are known. Another estimator of variance is proposed by Zamanzade (2016) by using a stochastic order constraint. The problem of estimation of the distribution function has been considered by Frey and Ozturk (2011), Wang et al. (2012), and Duembgen and Zamanzade (2013). Dastbaravarde et al. (2016) provided a theoretical framework for the JPS setting and showed that the empirical estimators in the JPS setting are less efficient than their counterpart in the RSS setting, but more efficient than their corresponding estimators in the SRS setting. Ozturk (2014) developed statistical inference for a population quantile. He then constructed a test as well as a distribution free confidence interval for it.

As mentioned earlier, the ranking in the RSS and the JPS settings is done by using an inexpensive and cheap method which does not require obtaining the precise values of the units, and thus it may be inaccurate and contain errors. In this case, which is called *imperfect ranking*, the statistical procedures are less efficient than in the perfect ranking case. While some statistical methods that were developed under the assumption of perfect ranking are relatively sensitive to deviations from this assumption, there are some alternative procedures for them which are more robust. Although the methods that are developed under the assumption of perfect rankings outperform their robust counterparts if this assumption holds, they may become completely invalid under imperfect rankings. Some examples of such methods were provided by Frey et al. (2007).

Based on the above discussion, it is desirable to test the assumption of perfect rankings, when we want to choose between a procedure that assumes the rankings are perfect and a robust counterpart of it. In the RSS setting, some tests for this purpose have been developed by Frey et al. (2007), Li and Balakrishnan (2008), Zamanzade et al. (2012), Frey and Wang (2013) and Zamanzade et al. (2014).

A test for perfect ranking will also be useful in the JPS setting because the rank of each observation is additional information that does not necessarily have to be taken into account in analyzing the data. Therefore, the test of perfect ranking in the JPS setting can be used for choosing between a JPS procedure requiring perfect ranking and a SRS procedure, which ignores

the rank information. Furthermore, even for JPS-based procedures that are robust to deviations from perfect ranking, the result of the test may help to decide whether it makes sense to use a JPS scheme again in a future, similar study: Even if applying the JPS scheme is assumed to be relatively inexpensive compared to the exact measurement of the quantity of interest, its cost may be unacceptable if the ranking is very far from perfect, such that the gain that can be expected from the JPS scheme will be minimal.

We organize the rest of the paper as follows. In Section 2, we discuss nonparametric tests of perfect judgment ranking in the RSS settings. In Section 3, we develop some nonparametric tests of perfect judgment ranking for the JPS setting. Monte Carlo comparisons of the developed tests in the JPS setting with their counterparts in the RSS setting are provided in Section 4. Some concluding remarks are provided in Section 5.

2 Nonparametric tests of perfect judgment ranking in ranked set sampling

Let $\{Y_{[i]j}, i = 1, \dots, m, j = 1, \dots, n_i\}$ be an unbalanced ranked set sample of size $N = \sum_{i=1}^m n_i$, where $Y_{[i]j}$ is j th observation with judgment rank i . Let $\{R_{[i]j}\}$ be the rank of $Y_{[i]j}$ among all N measured units in the ranked set sample (for $i = 1, \dots, m, j = 1, \dots, n_i$). Let F be the continuous distribution function of the population of interest. Since the rank vector of $\{Y_{[i]j}\}$ is the same as the rank vector of $\{F(Y_{[i]j})\}$, the joint distribution of $\{R_{[i]j}\}$ is distribution free. Note that the $R_{[i]j}$'s are not identically distributed and thus the possible observed values of the $R_{[i]j}$'s have different probabilities under the assumption of perfect ranking. Thus one natural way to construct a test for the assumption of perfect ranking is to reject this assumption when the probability of observing the vector of $\{R_{[i]j}\}$ is too small under the assumption of perfect ranking. This test was first proposed by Frey et al. (2007) and is denoted hereafter by NP .

Let $\bar{\mathbf{R}} = (\bar{R}_{[1]}, \dots, \bar{R}_{[m]})$ be the vector of mean ranks, where $\bar{R}_{[i]} = \frac{1}{n_i} \sum_{j=1}^{n_i} R_{[i]j}$. If the assumption of perfect ranking holds, one then expects the values of $\bar{R}_{[i]}$ to be increasing in i . Frey et al. (2007) showed that the vector $\bar{\mathbf{R}}$ asymptotically follows a multivariate normal distri-

bution. By using the asymptotic distribution of the vector $\bar{\mathbf{R}}$, Frey et al. (2007) suggested two test statistics, a chi-squared statistic K and a concordance test statistic W . The test statistic K is designed to measure the distance between the observed and the expected vector of the mean ranks $\bar{\mathbf{R}}$, whereas the test statistic W is a linear combination of the mean ranks in $\bar{\mathbf{R}}$, in a way that it is stochastically greater under the assumption of perfect ranking than under any alternative of imperfect ranking. Frey et al. (2007)'s simulation results indicate that the test based on K is uniformly less powerful than two other tests, so we do not consider it any further in this paper. A simplified form of the test statistic W is

$$W^* = \sum_{i=1}^m \sum_{j=1}^{n_i} iR_{[i]j},$$

and the assumption of perfect ranking is rejected for large values of W^* .

Independently from Frey et al. (2007), Li and Balakrishnan (2008) developed some nonparametric tests of perfect ranking for balanced RSS. Let

$$\begin{array}{ccc} Y_{[1]1} & \cdots & Y_{[m]1} \\ Y_{[1]2} & \cdots & Y_{[m]2} \\ \vdots & & \vdots \\ Y_{[1]n} & \cdots & Y_{[m]n} \end{array}$$

be a balanced ranked set sample of size $N = nm$. Let R_{ij} be the rank of $Y_{[i]j}$ among all units in the j 'th cycle (for $i = 1, \dots, m; j = 1, \dots, n$). Then, the perfect ranking tests proposed by Li and Balakrishnan (2008) are

$$\begin{array}{ll} N = \sum_{j=1}^n N_j, & N^* = \max(N_1, \dots, N_n), \\ S = \sum_{j=1}^n S_j, & S^* = \max(S_1, \dots, S_n), \\ A = \sum_{j=1}^n A_j, & A^* = \max(A_1, \dots, A_n), \end{array}$$

where

$$\begin{aligned}
N_j &= \sum_{i_1=1}^{m-1} \sum_{i_2=i_1+1}^m I(R_{i_1j} > R_{i_2j}), \\
S_j &= \sum_{i=1}^m (R_{ij} - i)^2, \\
A_j &= \sum_{i=1}^m |R_{ij} - i|,
\end{aligned}$$

for $j = 1, \dots, m$, and the assumption of perfect ranking is rejected when these test statistics are too large. Li and Balakrishnan (2008)'s simulation results indicate that the tests based on N , S and A are uniformly more powerful than the tests based on N^* , S^* and A^* , respectively.

Vock and Balakrishnan (2011) improved the test statistic based on N , by incorporating the information between the cycles. Their proposed test statistic for an unbalanced ranked set sample has the following form:

$$J = \sum_{i_1=1}^{m-1} \sum_{i_2=i_1+1}^m \sum_{j_1=1}^{n_{i_1}} \sum_{j_2=1}^{n_{i_2}} I(R_{i_1j_1} > R_{i_2j_2}).$$

The simulation results of Vock and Balakrishnan (2011) indicate that the tests based on J and W^* are uniformly more powerful than the tests based on N and S , respectively.

Zamanzade et al. (2012) improved three tests of perfect ranking proposed by Li and Balakrishnan (2008) using a permutation approach. Their procedure was based on the idea that, since for $j = 1, \dots, n$, the $Y_{[i]j}$'s are independently and identically distributed, then under the assumption of perfect ranking, each set of $(Y_{[1]l_1}, \dots, Y_{[m]l_m})$ can be regarded as a cycle of a balanced ranked set sample, for all $l_1, \dots, l_m \in \{1, \dots, n\}$. Thus, they proposed the following tests of perfect ranking:

$$\begin{aligned}
PN &= \sum_{h=1}^{n^m} N_h^*, \\
PS &= \sum_{h=1}^{n^m} S_h^*, \\
PA &= \sum_{h=1}^{n^m} A_h^*,
\end{aligned}$$

where the only difference between N_h^*, S_h^*, A_h^* and N_j, S_j, A_j is that the former set is applied to the h th sample out of the n^m samples $(Y_{[1]l_1}, \dots, Y_{[m]l_m})$, $l_1, \dots, l_m \in \{1, \dots, n\}$, instead of the j th cycle.

It turned out that the tests based on PN and PS are equivalent to the tests based on J and W^* , respectively. However, the PA test statistic is uniformly better than the test based on A .

The PA test statistic can be generalized to unbalanced RSS as follows. Let $n = \max_{1 \leq i \leq m} n_i$, $X_{ij} = Y_{[i]j}$ if $j \leq n_i$, and set X_{ij} to “missing” if $j > n_i$, for $i = 1, \dots, m, j = 1, \dots, n$. We apply PA to the X_{ij} s as if we had a balanced RSS. The only difference is in the computation of the A_h^* in the h th permuted sample. Let l_h be the number of non-missing observations in the h th permuted sample, and $(R_1^h, \dots, R_{l_h}^h)$ be the vector of the ranks of the non-missing observations in the h th permuted sample. More precisely, let R_i^h be the rank of the i th non-missing observation from the h th permuted sample among all l_h non-missing observations in the h th permuted sample, where these non-missing observations are renumbered consecutively with increasing judgement order rank. Then A_h^* is computed as $A_h^* = \sum_{i=1}^{l_h} |R_i^h - i|$.

Additional contributions in testing the assumption of perfect ranking were made by Frey and Wang (2013) and Zamanzade et al. (2014), who developed parametric tests of perfect ranking. However, since our interest here is in the nonparametric setting, we do not consider those tests any further.

Remark 1. *Some tests of perfect ranking in the JPS setting can be obtained by exploiting the connection between the JPS and the RSS settings. If we condition on the vector $\mathbf{n} = (n_1, \dots, n_m)$, then we can consider a JPS sample as a balanced or unbalanced ranked set sample. Thus the aforementioned tests of perfect rankings in the RSS setting can be used in the JPS setting, as well. However, we show in this paper that an alternate procedure gives better results in terms of power and ease of use.*

3 Nonparametric tests of perfect ranking for judgment post stratification

To draw a JPS sample of size N , using set size m , we first draw a simple random sample of Y_1, \dots, Y_N from the population. Then for each Y_i , $i \in \{1, \dots, N\}$, we draw an additional simple random sample of size $m - 1$, $\{Y_{i,2}, \dots, Y_{i,m}\}$, to create a set of size m , $\{Y_i, Y_{i,2}, \dots, Y_{i,m}\}$. Then, each set of size m is ranked from smallest to largest without actually measuring the units, and the rank of Y_i in the set is noted. Therefore, a JPS sample of size N can be represented as $(Y_1, R_1), \dots, (Y_N, R_N)$, where R_i is the rank of Y_i in the corresponding set of size m . Let $(Y_{(1)}, R_{(1)}), \dots, (Y_{(N)}, R_{(N)})$ be an ordered JPS sample, which is ordered according to the Y values (i.e. $Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(N)}$).

The following theorem allows us to derive some distributional properties of the vector $\mathbf{R}_o = (R_{(1)}, \dots, R_{(N)})$ under the assumption of perfect ranking.

Theorem 1. *Let $(Y_{(1)}, R_{(1)}), \dots, (Y_{(N)}, R_{(N)})$ be an ordered JPS sample of size N , with set size m , from a population with a continuous distribution function F . Let $\mathbf{R}_o = (R_{(1)}, \dots, R_{(N)})$ be the vector of the ranks in the ordered JPS sample. Then under the assumption of perfect ranking*

(I) *the joint distribution of the vector \mathbf{R}_o is given by*

$$P(R_{(1)} = r_1, \dots, R_{(N)} = r_N) = N! \left(\frac{1}{m}\right)^N \left(\frac{m!}{(mN)!} \prod_{j=2}^N C_j\right) \sum_{i_1} \sum_{i_2} \dots \sum_{i_{N-1}} \prod_{j=1}^{N-1} \frac{(mj - i_j + S_{j+1})! (i_j + r_{j+1} - 1)!}{(mj - i_j)! i_j!},$$

(II) $E(R_{(i)}) = 1 + \frac{i(m-1)}{N+1}$,

where $i_0 = 0$, $S_j = m - r_j$, $C_j = r_j \binom{m}{r_j}$ for $j = 1, \dots, N$ and i_j is indexed from $i_{j-1} + r_j$ to mj , for $j = 1, \dots, N - 1$.

Proof.

(I) Let n_i be the number of observations with judgment rank i , for $i = 1, \dots, m$. Since the vector $\mathbf{n} = (n_1, \dots, n_m)$ follows a multinomial distribution with N trials and probability

vector $(\frac{1}{m}, \dots, \frac{1}{m})$, we can write

$$\begin{aligned}
& P(R_{(1)} = r_1, \dots, R_{(N)} = r_N) \\
&= \left(\frac{N!}{n_1! \dots n_m!} \left(\frac{1}{m} \right)^N \right) (n_1! \dots n_m! \times P(Y'_{r_1} < \dots < Y'_{r_N})) \\
&= N! \left(\frac{1}{m} \right)^N \left(\frac{m!}{(mN)!} \prod_{j=2}^N C_j \right) \sum_{i_1} \sum_{i_2} \dots \sum_{i_{N-1}} \prod_{j=1}^{N-1} \frac{(mj - i_j + S_{j+1})! (i_j + r_{j+1} - 1)!}{(mj - i_j)! i_j!},
\end{aligned}$$

where the Y'_{r_j} s are independent r_j th order statistics from samples of size m , and the last equality follows from Theorem 4 in Kvam and Samaniego (1993), with the correction given in Frey (2007).

(II)

$$\begin{aligned}
E(R_{(i)}) &= 1 + E \left(\sum_{j=2}^m I(Y_{(i)} > Y_{i,j}) \right) \\
&= 1 + (m-1) P(F(Y_{(i)}) > F(Y_{i,2})) \\
&= 1 + (m-1) P(U_{(i)} > U) \\
&= 1 + \frac{i(m-1)}{N+1},
\end{aligned}$$

where $U_{(i)}$ and U are independent random variables and follow $Beta(i, N+1-i)$ and $U(0, 1)$ distributions, respectively. \square

It is clear from the above theorem that the joint distribution of the vector \mathbf{R}_o does not depend on the population distribution, and therefore the tests based on \mathbf{R}_o will be distribution free.

The first test statistic that is introduced in this paper is based on the joint probability of the vector $\mathbf{R}_o = (R_{(1)}, \dots, R_{(N)})$. In fact, if the observed values of the vector \mathbf{R}_o have a small probability under the assumption of perfect ranking, this can be regarded as a symptom of imperfect ranking. Therefore, we propose the following test statistic for assessing the assumption of perfect ranking in the JPS sampling scheme:

$$TP = -\ln [P_0 (R_{(1)} = r_1, \dots, R_{(N)} = r_N)],$$

where (r_1, \dots, r_N) are the actually observed values of \mathbf{R}_o and the probability is calculated under perfect ranking. The TP test statistic can be computed efficiently via the R-code provided by Frey and Wang (2013).

Another test of perfect ranking can be constructed by counting the number of $R_{(i)}$ s that violate their order expectations. Note that in an ordered JPS sample, the units are sorted according to the Y values. Therefore, since $Y_{(i)} < Y_{(j)}$, for $i < j$, then under the assumption of perfect ranking, it is expected that $R_{(i)} < R_{(j)}$, for $i < j$, as well. Thus, we propose the following test statistic for assessing the assumption of perfect ranking:

$$TN = \sum_{i=1}^{N-1} \sum_{j=i+1}^N I(R_{(i)} > R_{(j)})$$

The third and the fourth test statistics we propose in this paper are based on the distance between the vector $\mathbf{R}_o = (R_{(1)}, \dots, R_{(N)})$ and its expectation under the assumption of perfect ranking, i.e. $E(\mathbf{R}_o) = (E(R_{(1)}), \dots, E(R_{(N)}))$. We take the sum of squares of $(R_{(i)} - E(R_{(i)}))$, and the sum of absolute values of $(R_{(i)} - E(R_{(i)}))$ as tests for perfect ranking. Therefore the tests of perfect ranking can be constructed as $TS = \sum_{i=1}^N (R_{(i)} - E(R_{(i)}))^2$ and $TA = \sum_{i=1}^N |R_{(i)} - E(R_{(i)})|$. However, it is easy to see that these two test statistics are one to one functions of

$$TS^* = \sum_{i=1}^N (R_{(i)} - i)^2,$$

$$TA^* = \sum_{i=1}^N |R_{(i)} - i|,$$

respectively, which have the advantage of always being an integer.

All proposed test statistics reject the assumption of perfect ranking when their values are too large. In Table 1, we have given the critical values of the test based on TP for nominal levels close to 0.05 and 0.10. Those entries marked by asterisks were obtained by using Monte Carlo simulation with 1,000,000 repetitions, while the others are exact and were obtained by using Theorem 1. It should be noted that due to the discreteness of the distribution of all considered test statistics in this paper in both the RSS and the JPS settings, it is usually not possible to attain an exact size of α without randomizing. For example, for $N = 5$ and $m = 2$, and under the assumption of perfect ranking, $TP \geq 5.0439$ with null probability 0.0367 and $TP \geq 4.7756$ with null probability

0.0535. Therefore, to obtain the significance level $\alpha = 0.05$ in a randomized test based on TP , the assumption of perfect ranking is rejected with probability one if $TP \geq 5.0439$, and with probability $\frac{0.05-0.0367}{0.0535-0.0367} = 0.791$ if $TP = 4.7756$. However, it is interesting to see that the attainable exact levels of the non-randomized test based on TP are very close to the nominal level of $\alpha = 0.05$ and $\alpha = 0.10$ for small values of N . For moderate and large values of sample size $N \geq 15$, the exact levels of the test based on TP are the same as 0.05 and 0.10 up to four decimal places. This can be regarded as an advantage of the test based on TP as compared to its competitors in the RSS setting, which typically have exact levels not as close to 0.05 and 0.10.

We do not provide tables of critical values for the other proposed test statistics (TN , TS^* , and TA^*), since the simulation results in Section 4 indicate that the test based on TP is uniformly more powerful than its competitors in the JPS setting.

Remark 2. *Each of the tests presented in this section is based on the same idea as one of the tests for the RSS setting presented in Section 2, which can also be applied to the JPS setting as described in Remark 1): The TP test and the NP test both use the probability of observing ranks under the assumption of perfect ranking, TN and PN (equivalent to J) are both based on counting violations of the order of measured values expected according to their judgment ranks, TS^* and PS (equivalent to W^*) both use the squared differences between true ranks among a certain set and judgment ranks, and TA^* and PA are based on the absolute values of these differences.*

However, it is worth mentioning that the tests originally designed for the RSS setting (NP , J , W^ , PA) condition on the vector (n_1, \dots, n_m) , while the tests presented in this section do not. This means that while the tests based on TP , TN , TS^* , and TA^* control the type I error rate only unconditionally, the tests based on NP , J , W^* , and PA control it conditionally on the vector of post stratum sample sizes. This may be considered as an advantage of the tests based on Remark 1 because the vector of post stratum sample sizes (n_1, \dots, n_m) contains no information about whether the rankings are perfect or not. However, the drawback of controlling the type I error rate conditionally will be that the conditional distributions tend to have larger jumps than the unconditional distribution, such that the size of these conditional tests based on NP , J , W^* and PA will typically be smaller for the non-randomized version, or randomization will play a much*

more important role for the randomized version, respectively, in comparison to the unconditional tests based on TP , TN , TS^* , and TA^* .

For the tests based on NP , J , W^* , and PA , the conditional distribution given the vector (n_1, \dots, n_m) of post stratum sample sizes is needed for each of the $\binom{N+m-1}{m-1}$ possible such vectors. This number rapidly grows as N increases, such that providing tables of critical values in advance will hardly ever be feasible except for small JPS samples. In practice, for a single application of the test, the critical values could be determined after collecting the data, for the observed post stratum sample sizes only. However, this would mean that the researcher would have to wait for these time-consuming calculations or simulations to complete before the test can be applied to the observed data.

4 Power Comparisons

In this section, we compare the proposed tests with the best existing tests of perfect ranking in the RSS setting, by conditioning on the vector of post stratum sample sizes $\mathbf{n} = (n_1, \dots, n_m)$ and thinking of a JPS sample as an unbalanced or balanced ranked set sample. The nonparametric tests of perfect ranking in the RSS setting, which are used in our comparison set, are as follows: The tests based on NP and W^* due to Frey et al. (2007), the test based on J due to Vock and Balakrishnan (2011), and the test based on PA due to Zamanzade et al. (2012). For power comparisons, we have considered the following four scenarios of imperfect ranking:

- **Bivariate normal model** due to Dell and Clutter (1972). Let (X, Y) follow a bivariate normal distribution with a correlation of ρ . In each set of size m , the Y values are ranked according to the X values.
- **Fraction of random rankings** due to Frey et al. (2007). In this imperfect ranking scenario, $R_{(i)}$ (the judgment rank of $Y_{(i)}$) is selected correctly with probability λ , and with probability $1 - \lambda$, it is selected randomly from the set $\{1, \dots, m\}$. Thus, the distribution of $R_{(i)}$ is as follows: $R_{(i)} = \lambda R_{0(i)} + (1 - \lambda) R$, where R follows a uniform distribution on the set $\{1, \dots, m\}$ and $R_{0(i)}$ is the rank of $Y_{(i)}$ in the case of perfect ranking.

Table 1: Critical values (CV) of the test based on TP for nominal levels close to 0.05 and 0.10, and the corresponding levels, $P_0(TP \geq CV)$.

N	$m = 2$		$m = 3$		$m = 4$		$m = 5$	
	CV	Level	CV	Level	CV	Level	CV	Level
5	5.0439	0.0367	6.6639	0.0498	7.9671	0.0495	8.9311	0.0499
	4.7756	0.0535	6.5787	0.0526	7.9300	0.0501	8.9265	0.0501
	4.4733	0.0763	6.0803	0.0996	7.2937	0.0999	8.2198	0.0999
	4.3899	0.1011	6.0738	0.1019	7.2937	0.1006	8.2077	0.1005
10	8.1915	0.0499	-	-	-	-	-	-
	8.1914	0.0502	11.7180	0.0500	14.0646	0.0500*	15.8087	0.0500*
	7.4597	0.0997	10.8254	0.0999	13.0857	0.0999*	14.7968	0.0999*
	7.4596	0.1002	10.8253	0.1001	13.0856	0.1001*	14.7967	0.1001*
15	11.3389	0.0500*	16.3931	0.0500*	19.7197	0.0500*	22.2184	0.0500*
	10.4640	0.1000*	15.3147	0.1000*	18.5515	0.1000*	20.9954	0.1000*
20	14.3652	0.0500*	20.8828	0.0500*	25.2115	0.0500*	28.4135	0.0500*
	13.3616	0.1000*	19.6621	0.1000*	23.8792	0.1000*	27.0271	0.1000*
25	17.3244	0.0500*	25.2893	0.0500*	30.5826	0.0500*	34.5289	0.0500*
	16.2189	0.1000*	23.9499	0.1000*	29.1240	0.1000*	32.9997	0.1000*
30	20.2428	0.0500*	29.6265	0.0500*	35.9103	0.0500*	40.5440	0.0500*
	19.0398	0.1000*	28.1865	0.1000*	34.3203	0.1000*	38.8895	0.1000*
35	23.1233	0.0500*	33.9559	0.0500*	41.1477	0.0500*	46.5203	0.0500*
	21.8352	0.1000*	32.4017	0.1000*	39.4567	0.1000*	44.7453	0.1000*
40	25.9648	0.0500*	38.2397	0.0500*	46.3759	0.0500*	52.4407	0.0500*
	24.6040	0.1000*	36.5731	0.1000*	44.5908	0.1000*	50.5462	0.1000*
45	28.7803	0.0500*	42.4482	0.0500*	51.5730	0.0500*	58.3506	0.0500*
	27.3504	0.1000*	40.7107	0.1000*	49.6738	0.1000*	56.3441	0.1000*
50	31.5842	0.0500*	46.6889	0.0500*	56.7021	0.0500*	64.2064	0.0500*
	30.0767	0.1000*	44.8635	0.1000*	54.7328	0.1000*	62.1108	0.1000*

- **Fraction of inverse rankings** due to Frey et al. (2007). Here, $R_{(i)}$ is selected correctly with probability λ , and with probability $1 - \lambda$, it is selected completely inversely. Therefore, the distribution of $R_{(i)}$ is as follows: $R_{(i)} = \lambda R_{0(i)} + (1 - \lambda) (m + 1 - R_{0(i)})$.
- **Fraction of neighbor rankings** due to Vock and Balakrishnan (2011). In this imperfect rankings scenario, $R_{(i)}$ is selected correctly with probability λ , it is one too small with probability $\frac{1-\lambda}{2}$, and it is one too large with probability $\frac{1-\lambda}{2}$. Thus its distribution is: $R_{(i)} = \lambda R_{0(i)} + \frac{1-\lambda}{2} \min(R_{0(i)} + 1, m) + \frac{1-\lambda}{2} \max(R_{0(i)} - 1, 1)$.

We have generated 1,000,000 random samples from the above imperfect rankings scenarios, for $\rho, \lambda \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$ and $(N, m) = (5, 5), (10, 5), (5, 10)$. The (N, m) values are taken so that we can observe the effect of increasing sample size, while the set size is being fixed, and the effect of increasing set size while the sample size is being fixed. The considered values of the sample size (N) in the simulation study are relatively small due to computational limitations of calculating critical values of the competing tests in ranked set sampling. Since a JPS sample of size N with set size m can be considered as a ranked set sample in $\binom{N+m-1}{m-1}$ distinct ways, depending on the observed values of the vector of the ranks \mathbf{R}_o , we determined critical values of the RSS perfect ranking tests for every possible vector of post stratum sample sizes (n_1, \dots, n_m) by exhaustive listing to perform the exact RSS tests of perfect ranking at significance level $\alpha = 0.05$. Thus, we have only compared the powers of the proposed tests with the competitors in the RSS setting for relatively small values of N due to the computational difficulties of calculating the critical values of the tests in the RSS setting. Furthermore, the power of the test based on PA is estimated only for $(N, m) = (5, 5)$ and based on 10,000 repetitions due to its computational difficulties. The comparisons are done at significance level $\alpha = 0.05$, and randomized versions of all tests are used to attain this exact level. The powers of the tests are presented in Tables 2-5.

Remark 3. *In the case that the ranks of all observations in the JPS sample are the same, i.e. $\mathbf{R}_o = (R_{[1]}, \dots, R_{[N]}) = (r, \dots, r)$, for some $r \in \{1, \dots, m\}$, each test statistics puts probability one on a single value, making the distribution degenerate. In this case, we have used the trivial randomized test that always rejects the assumption of perfect ranking with probability $\alpha = 0.05$.*

Table 2: The estimated powers of the different tests of perfect judgment rankings in the JPS sampling scheme at significance level of 0.05 under the bivariate normal model.

(N, m)	ρ	Test Statistic							
		TP	TN	TS^*	TA^*	NP	W^*	J	PA
(5, 5)	0	0.7101	0.6263	0.6299	0.5922	0.6973	0.6924	0.6598	0.6647
	0.2	0.6126	0.5203	0.5220	0.4863	0.6049	0.5992	0.5635	0.5730
	0.4	0.5011	0.4071	0.4039	0.3736	0.4952	0.4900	0.4538	0.456
	0.6	0.3620	0.2862	0.2777	0.2560	0.3601	0.3579	0.3264	0.3449
	0.8	0.2090	0.1631	0.1518	0.1415	0.2066	0.2038	0.1851	0.1955
	1	0.0504	0.0504	0.0506	0.0501	0.0502	0.0501	0.0502	0.0510
(10, 5)	0	0.9593	0.9215	0.6236	0.4980	0.9418	0.9372	0.9269	-
	0.2	0.8994	0.8437	0.4904	0.3787	0.8866	0.8789	0.8625	-
	0.4	0.7966	0.7176	0.3510	0.2643	0.7871	0.7762	0.7530	-
	0.6	0.6158	0.5257	0.2218	0.1667	0.6147	0.6030	0.5747	-
	0.8	0.3470	0.2791	0.1174	0.0926	0.3445	0.3371	0.3151	-
	1	0.0500	0.0499	0.0499	0.0501	0.0499	0.0499	0.0501	-
(5, 10)	0	0.8586	0.7569	0.1946	0.1006	0.8384	0.8336	0.7876	-
	0.2	0.7893	0.6596	0.1550	0.0827	0.7724	0.7666	0.7094	-
	0.4	0.6886	0.5416	0.1198	0.0682	0.6789	0.6718	0.6061	-
	0.6	0.5498	0.3994	0.0909	0.0586	0.5432	0.5362	0.4669	-
	0.8	0.3399	0.2317	0.0666	0.0519	0.3362	0.3312	0.2793	-
	1	0.0501	0.0503	0.0499	0.0499	0.0502	0.0500	0.0501	-

Table 3: The estimated powers of the different tests of perfect judgment rankings in the JPS sampling scheme at significance level of 0.05 under a fraction of random rankings.

(N, m)	λ	Test Statistic							
		TP	TN	TS^*	TA^*	NP	W^*	J	PA
(5, 5)	0	0.7099	0.6260	0.6299	0.5924	0.6970	0.6923	0.6595	0.6672
	0.2	0.6025	0.5120	0.5173	0.4687	0.5942	0.5901	0.5546	0.5612
	0.4	0.4791	0.3928	0.4004	0.3468	0.4758	0.4727	0.4363	0.4454
	0.6	0.3464	0.2728	0.2803	0.2320	0.3433	0.3412	0.3092	0.3096
	0.8	0.2085	0.1564	0.1627	0.1313	0.2004	0.1990	0.1786	0.1806
	1	0.0499	0.0501	0.0499	0.0499	0.0499	0.0501	0.0500	0.0512
(10, 5)	0	0.9515	0.9216	0.6237	0.4979	0.9419	0.9372	0.9270	-
	0.2	0.8796	0.8264	0.4837	0.3719	0.8704	0.8650	0.8454	-
	0.4	0.7565	0.6782	0.3461	0.2586	0.7508	0.7446	0.7136	-
	0.6	0.5681	0.4786	0.2207	0.1653	0.5675	0.5617	0.5211	-
	0.8	0.3256	0.2522	0.1197	0.0961	0.3249	0.3218	0.2845	-
	1	0.0501	0.0499	0.0504	0.0501	0.0500	0.0499	0.0510	-
(5, 10)	0	0.8531	0.7561	0.1947	0.1005	0.8380	0.8331	0.7872	-
	0.2	0.7630	0.6439	0.1562	0.0837	0.7520	0.7477	0.6908	-
	0.4	0.6394	0.5118	0.1225	0.0706	0.6336	0.6298	0.5672	-
	0.6	0.4803	0.3643	0.0938	0.0614	0.4788	0.4759	0.4171	-
	0.8	0.2843	0.2078	0.0696	0.0547	0.2843	0.2830	0.2430	-
	1	0.0499	0.0499	0.0497	0.0497	0.0499	0.0501	0.0499	-

Table 4: The estimated powers of the different tests of perfect judgment rankings in the JPS sampling scheme at significance level of 0.05 under a fraction of inverse rankings.

(N, m)	λ	Test Statistic							
		TP	TN	TS^*	TA^*	NP	W^*	J	PA
(5, 5)	0	0.9778	0.9760	0.9724	0.9609	0.9677	0.9666	0.9642	0.9646
	0.2	0.8998	0.8703	0.8701	0.8166	0.8823	0.8863	0.8734	0.8732
	0.4	0.7661	0.7044	0.7196	0.6285	0.7513	0.7562	0.7291	0.7302
	0.6	0.5788	0.4970	0.5265	0.4212	0.5702	0.5742	0.5368	0.5479
	0.8	0.3391	0.2698	0.3006	0.2191	0.3373	0.3388	0.3060	0.3031
	1	0.0499	0.0501	0.0498	0.0499	0.0500	0.0501	0.0499	0.0512
(10, 5)	0	1.0000	0.9999	0.9852	0.9611	0.9998	0.9998	0.9997	-
	0.2	0.9943	0.9937	0.9042	0.8291	0.9917	0.9934	0.9922	-
	0.4	0.9551	0.9438	0.7308	0.6103	0.9488	0.9500	0.9455	-
	0.6	0.8256	0.7788	0.4850	0.3650	0.8210	0.8216	0.7982	-
	0.8	0.5388	0.4501	0.2308	0.1664	0.5385	0.5385	0.4889	-
	1	0.0501	0.0499	0.0502	0.0502	0.0501	0.0500	0.0498	-
(5, 10)	0	0.9972	0.9966	0.4486	0.2240	0.9919	0.9918	0.9901	-
	0.2	0.9580	0.9388	0.3407	0.1647	0.9469	0.9510	0.9385	-
	0.4	0.8622	0.8060	0.2454	0.1200	0.8493	0.8568	0.8246	-
	0.6	0.6932	0.6041	0.1644	0.0875	0.6817	0.6906	0.6416	-
	0.8	0.4296	0.3447	0.0992	0.0646	0.4233	0.4290	0.3828	-
	1	0.0499	0.0499	0.0498	0.0500	0.0501	0.0499	0.0499	-

Table 5: The estimated powers of the different tests of perfect judgment rankings in the JPS sampling scheme at significance level of 0.05 under a fraction of neighbour rankings.

(N, m)	λ	Test Statistic							
		TP	TN	TS^*	TA^*	NP	W^*	J	PA
(5, 5)	0	0.1849	0.1538	0.1369	0.1390	0.1905	0.1860	0.1800	0.1766
	0.2	0.1544	0.1302	0.1169	0.1172	0.1593	0.1556	0.1502	0.1496
	0.4	0.1254	0.1081	0.0988	0.0979	0.1288	0.1265	0.1223	0.1278
	0.6	0.0984	0.0875	0.0813	0.0803	0.1008	0.0993	0.0966	0.9440
	0.8	0.0729	0.0677	0.0651	0.0640	0.0743	0.0736	0.0720	0.0717
	1	0.0499	0.0501	0.0499	0.0498	0.0498	0.0501	0.0499	0.0512
(10, 5)	0	0.3206	0.2703	0.1059	0.0929	0.3385	0.3095	0.3174	-
	0.2	0.2540	0.2153	0.0930	0.0829	0.2689	0.2475	0.2519	-
	0.4	0.1929	0.1652	0.0809	0.0737	0.2042	0.1902	0.1916	-
	0.6	0.1364	0.1196	0.0689	0.0644	0.1435	0.1357	0.1354	-
	0.8	0.0894	0.0816	0.0591	0.0573	0.0927	0.0895	0.0890	-
	1	0.0501	0.0499	0.0501	0.0501	0.0500	0.0498	0.0502	-
(5, 10)	0	0.1187	0.0942	0.0544	0.0503	0.1264	0.1204	0.1126	-
	0.2	0.1036	0.0845	0.0535	0.0502	0.1095	0.1049	0.0989	-
	0.4	0.0897	0.0759	0.0525	0.0501	0.0938	0.0906	0.0861	-
	0.6	0.0757	0.0668	0.0517	0.0500	0.0786	0.0763	0.0733	-
	0.8	0.0622	0.0583	0.0506	0.0499	0.0638	0.0628	0.0617	-
	1	0.0499	0.0499	0.0500	0.0498	0.0499	0.0501	0.0499	-

It is clear from the Tables 2–5 that the tests based on TP and NP are the best among the tests considered in the JPS and the RSS settings. Furthermore, the tests based on TS^* and TA^* are the worst tests and their powers are considerably lower than those of the other tests. Although the powers of the tests based on TP and NP are competitive in all considered cases, the powers of TP are slightly higher than those of NP in the bivariate normal model, the fraction of random rankings model, and the fraction of inverse rankings model. The NP test is the best in the fraction of neighbour rankings model.

5 Conclusion

In this paper, we developed some nonparametric tests of perfect judgment ranking for the judgment post stratification sampling scheme. We compared the proposed tests with their competitors in ranked set sampling via Monte Carlo simulation. Based on our comparison, we observe that:

- The best proposed test (TP) has competitive powers with the best test of perfect ranking in ranked set sampling (NP), and its powers are slightly higher in most considered cases.
- The critical values of the TP test can be simply calculated or simulated for different values of sample size and set size, while providing critical values of the nonparametric tests in the RSS setting for using them in the JPS setting is a challenging task for large values of (N, m) .
- The exact levels of the test based on TP are very close to the nominal level $\alpha = 0.05$ and $\alpha = 0.10$, when the sample size (N) or set size (m) is not too small.

Therefore, based on the above observations, we generally recommend to use the test based on TP in practice.

If controlling the significance level conditionally on the vector of post stratum sample sizes is considered to be important, the NP test may be more attractive (see Remark 2). According to the simulations, its power is also very high, while for a non-randomized version of the test being applied to small samples, the power can be expected to be slightly below that of the test based on

TP. Determining the critical values for the *NP* test is more complicated than for the test based on *TP*, since this has to be done separately for each possible vector of post stratum sample sizes.

Acknowledgement

The authors are thankful to two anonymous referees for their valuable comments and suggestions which greatly improved the earlier version of this paper.

References

- Dastbaravarde, A., Arghami, N.R., and Sarmad, M. 2016. Some theoretical results concerning non-parametric estimation by using a judgment post-stratification sample. *Communications in Statistics – Theory and Methods*, 45 (8), 2181-2203.
- Dell, T. R., and Clutter, J. L., 1972. Ranked set sampling theory with order statistics background. *Biometrics*, 28, 2, 545-555.
- Duembgen, L., and Zamanzade, E., 2013. Inference on a distribution function from ranked set samples, arXiv:1304.6950v3 [stat.ME] .
- Frey, J., 2007. A note on a probability involving independent order statistics. *Journal of Statistical Computation and Simulation*, 77, 11, 969-975.
- Frey, J., and Feeman, T. G., 2012. An improved mean estimator for judgment post-stratification. *Computational Statistics and Data Analysis*, 56, 2, 418-426.
- Frey, J., and Feeman, T. G., 2013. Variance estimation using judgment post-stratification. *Annals of the Institute of Statistical Mathematics*, 65, 3, 551-569.
- Frey, J., and Ozturk, O., 2011. Constrained estimation using judgment post-stratification. *Annals of the Institute of Statistical Mathematics*, 63, 4, 769-789.

- Frey, J., Ozturk, O., and Deshpande, J.V., 2007. Nonparametric tests for perfect judgment rankings. *Journal of the American Statistical Association*, 102, 478, 708-717.
- Frey, J., and Wang, L., 2013. Most powerful rank tests for perfect rankings. *Computational Statistics and Data Analysis*, 60, 157-168.
- Kadilar, C.; Unyazici, Y., and Cingi, H., 2009. Ratio estimator for the population mean using ranked set sampling. *Statistical Papers*. 50 (2), 301-309.
- Kvam, P. H., and Samaniego, F. J., 1993. On the inadmissibility of empirical averages as estimators in ranked set sampling. *Journal of Statistical Planning and Inference*, 36, 39-55.
- Kvam, P.H., and Samaniego, F.J., 1994. Nonparametric maximum likelihood estimation based on ranked set samples, *Journal of the American Statistical Association*, 89, 426, 526-537.
- Li, T., and Balakrishnan, N., 2008. Some simple nonparametric methods to test for perfect ranking in ranked set sampling. *Journal of Statistical Planning and Inference*, 138, 5, 1325-1338.
- Li, W., Liu, T., and Bai, Z. 2012. Rounded data analysis based on ranked set sample. *Statistical Papers*. 53 (2), 439-455.
- McIntyre, G. A., 1952. A method for unbiased selective sampling, using ranked sets. *Austral. J. Agricultural Res.* 3, 385-390.
- MacEachern, S.N., Ozturk, O., Wolfe, D.A. and Stark, G.V., 2002. A new ranked set sample estimator of variance. *Journal of the Royal Statistical Society: Series B.* 64, 2, 177-188.
- MacEachern, S.N., Stasny, E.A., and Wolfe, D.A., 2004. Judgment Post-Stratification with imprecise rankings. *Biometrics*, 60,1, 207-215.
- Ozturk, O., 2014. Statistical inference for population quantiles and variance in judgment post-stratified samples. *Computational Statistics and Data Analysis*, 77, 188-205.
- Perron, F., and Sinha, B.K., 2004. Estimation of variance based on a ranked set sample. *Journal of Statistical Planning and Inference*, 120, 1-2, 21-28.

- Stokes, S. L., 1980. Estimation of Variance Using Judgment Ordered Ranked Set Samples. *Biometrics*, 36, 1, 35-42.
- 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, 374-381.
- Vock, M., and Balakrishnan, N., 2011. A Jonckheere–Terpstra-type test for perfect ranking in balanced ranked set sampling. *Journal of Statistical Planning and Inference*, 141, 2, 624-630.
- Wang, X., Lim, J., Stokes, L., 2008. A Nonparametric mean estimator for judgment post stratified data. *Biometrics*, 64, 2, 355-363.
- Wang, X., Wang, K., and Lim, J., 2012. Isotonized CDF estimation from judgment post stratification data with empty strata, *Biometrics*, 68, 1, 194-202.
- Zamanzade, E. , Arghami, N.R., Vock, M., 2012. Permutation-based tests of perfect ranking, *Statistics and Probability Letters*, 82, 2213-2220.
- Zamanzade, E. , Arghami, N.R., Vock, M., 2014. A parametric test of perfect ranking in balanced ranked set sampling, *Communications in Statistics – Theory and Methods*, 43, 4589-4611.
- Zamanzade, E., and Vock, M., 2015. Variance estimation in ranked set sampling using a concomitant variable, *Statistics and Probability Letters*, 105, 1-5.
- Zamanzade, E., 2016. An isotonized variance estimator for judgment post stratified data, *Journal of the Korean Statistical Society*. 45 (1), 25-32.