PARAMETRIC TESTS OF PERFECT JUDGMENT RANKING BASED ON ORDERED RANKED SET SAMPLES

Authors: Ehsan Zamanzade
– Department of Statistics, University of Isfahan,
Isfahan 81746-73441, Iran
(E.Zamanzade@sci.ui.ac.ir; Ehsanzamanzadeh@yahoo.com)

Michael Vock
– Institute of Mathematical Statistics and Actuarial Science,
University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland.
(michael.vock@stat.unibe.ch)

Abstract:

- We develop parametric and location-scale free tests of perfect judgment ranking based on ordered ranked set samples. The tests are based on the differences between the elements of the ordered ranked set samples and those of the original ranked set samples. We compare our proposed tests with the best existing tests of perfect judgment ranking in the literature by using Monte Carlo simulation. Our simulation results show that the proposed tests behave favorably in comparison with their leading competitors, especially under the fraction of neighbor rankings model. In comparison to the nonparametric competitors, the proposed tests have the advantage of not needing randomization to attain a specific size.

Key-Words:

- *Ordered ranked set samples; perfect ranking; test.*

AMS Subject Classification:

- 62D05, 62F03.

1. INTRODUCTION

When measuring variables of interest is expensive or time-consuming, but ranking them in small groups without actual measurement is easy and convenient, ranked set sampling (RSS) can be regarded as an efficient technique for collecting more informative samples and therefore having more reliable inferences. This sampling technique, which was firstly introduced by McIntyre (1952, 2005), can be applied in both balanced and unbalanced strategies. In the balanced case, the researcher first draws k random samples of size k and orders them based on his personal judgment (not actual measurement). Then, for $i = 1, \dots, k$, he actually measures the i^{th} judgment ordered observation from the i^{th} sample. Finally, he repeats this procedure n times (cycles) in order to draw a sample of size kn from a Balanced Ranked Set Sampling (BRSS) scheme. In Unbalanced Ranked Set Sampling (UBRSS), the numbers of i^{th} judgment ordered observations are not necessarily the same anymore. A comprehensive review of works on RSS including a comprehensive list of references can be found in Wolfe (2012).

Although many researchers have shown that a ranked set sample may allow for more reliable inferences than a simple random sample of the same size, this reliability decreases as errors in ranking observations based on personal judgment occur. Frey et al. (2007) have exemplified how the ranking error can invalidate the method of inference in both parametric and nonparametric cases. Therefore, it seems to be vital to develop tests for assessing the assumption of perfect judgment ranking for both parametric and nonparametric cases. Surprisingly, this has not been done up to quite recently. Frey et al. (2007) and Li and Balakrishnan (2008) independently proposed some nonparametric tests of perfect judgment ranking, followed by Vock and Balakrishnan (2011), Zamanzade et al. (2012), Vock and Balakrishnan (2013), Frey and Wang (2013), and Zamanzade et al. (2014).

This paper is organized as follows: In Section 2, we propose our tests of perfect judgment ranking for one cycle, then, in the next section, we generalize them to the multi-cycle case. In Section 4, we compare our proposed tests with their leading competitors in the literature. Conclusions and some final remarks are provided in Section 5.

2. INTRODUCTION OF TESTS STATISTICS

Let $X_{[1]}, \dots, X_{[k]}$ be a sample of size k from BRSS with one cycle, where $X_{[i]}$ $(i = 1, \dots, k)$ is the i^{th} judgment ordered observation from the i^{th} sample, which is actually measured. It should be noted that the $X_{[i]}$'s are independent from each other and follow the distribution of an i^{th} order statistic if the assumption of perfect judgment ranking is completely satisfied. Furthermore, due to the independence of the $X_{[i]}$'s, $P\left(X_{[i]} < X_{[j]}\right) < 1$ for $i < j$ and $i, j \in \{1, \dots, k\}$, and this probability decreases as the judgment ranking becomes more and more unreliable. So intuitively, it is expected that the two vectors $\left(X_{[1]},\ldots,X_{[k]}\right)$ and $\left(Z_{(1)},\ldots,Z_{(k)}\right)$ are close to each other provided that the assumption of perfect judgment ranking is completely satisfied, where $\left(Z_{(1)},\ldots,Z_{(k)}\right)$ is the vector of Ordered Ranked Set Samples (ORSS) which is obtained by putting the values of $\left(X_{[1]},\ldots,X_{[k]}\right)$ in order. Therefore if the underlying distribution of population is completely known, then the following tests can be proposed for assessing the assumption of perfect judgment ranking:

$$TA = \sum_{i=1}^{k} \frac{|d_i|}{E|d_i|};$$
$$TS = \sum_{i=1}^{k} \frac{d_i^2}{Ed_i^2};$$

where $d_i = X_{[i]} - Z_{(i)}$, and $E\left(.\right)$ is the expectation operator which is taken under the assumption of perfect judgment ranking.

Intuitively, large values of TA, TS are a symptom of violation of the assumption of perfect judgment ranking and therefore this assumption should be rejected for large enough values of TA, TS.

If the underlying distribution of the population belongs to a location-scale family, then the above test statistics can be simplified as follows:

$$TA = \sum_{i=1}^{k} \frac{|d_i|}{\sigma E_{\mu=0,\sigma=1}|d_i|};$$
$$TS = \sum_{i=1}^{k} \frac{d_i^2}{\sigma^2 E_{\mu=0,\sigma=1} d_i^2};$$

where μ, σ are location and scale parameters, respectively.

Obviously, the above test statistics are location-free, and they will be scale-free if an equivariant estimator is used for the estimation of σ.

3. EXTENSION OF THE PROPOSED TESTS TO THE MULTI-CYCLE CASE

Although several methods have been proposed in the literature for extending tests of perfect judgment ranking from the one-cycle to the multi-cycle case, Zamanzade et al. (2012)'s simulation study has shown that their permutation-based technique provides good results under many scenarios. So we use their method to extend our tests to the multi-cycle case.

Suppose that $\left(X_{[i]j}\right)_{i\leq k, j\leq n}$ is a sample of size of kn, which is drawn by an n-cycle BRSS scheme, where $X_{[i]j}$ is the i^{th} judgment ordered observation from the j^{th} cycle ($i = 1, \ldots, k; j = 1, \ldots, n$). Since all observations are mutually independent and the observations in each column are also identically distributed, it is expected that the vector of ordered observations in each row should be close to the unordered row vector if we permute observations in each column, provided that the assumption of perfect judgment ranking is fully satisfied. In other words, under the assumption of perfect judgment ranking, the vector $\left(X_{[1]l_1}, X_{[2]l_2}, X_{[3]l_3}, \ldots, X_{[k]l_k}\right)$ and the ordered vector of this vector, which is denoted here by $\left(Z_{(1)l_1l_2\ldots l_k}, Z_{(2)l_1l_2\ldots l_k}, Z_{(3)l_1l_2\ldots l_k}, \ldots, Z_{(k)l_1l_2\ldots l_k}\right)$, should be close to each other for all $(l_1, l_2, l_3, \ldots, l_k) \in \{1, 2, 3, \ldots, n\}^k$.

Based on the above arguments, TA, TS can be extended to the multi-cycle case as follows:

$$TPA = \sum_{i=1}^{n^k} TA_i,$$

$$TPS = \sum_{i=1}^{n^k} TS_i,$$

where TA_i, TS_i are the values of TA and TS, respectively, for the i^{th} sample out of all n^k samples of the form $\left(X_{[1]l_1}, X_{[2]l_2}, X_{[3]l_3}, \ldots, X_{[k]l_k}\right)$, $(l_1, l_2, l_3, \ldots, l_k) \in \{1, 2, 3, \ldots, n\}^k$.
We reject the hypothesis of perfect judgment ranking for large enough values of TPA and TPS.

The calculation of TPA or TPS based on all n^k samples of the form mentioned above is too time-consuming for practical application except for very small values of k and n. We therefore propose a less intuitive, but more efficient way of computing these statistics. R-code for the computation of TPA and TPS using the following method is available on request from the authors.

For n cycles, with $E_i = E_{\mu=0,\sigma=1}|d_i|$, TPA can be written as

$$\begin{aligned} TPA &= \sum_{(l_1,\ldots,l_k)\in\{1,\ldots,n\}^k} \sum_{i=1}^{k} \frac{\left|X_{[i]l_i} - Z_{(i)l_1l_2,\ldots,l_k}\right|}{\sigma E_i} \\ &= \sum_{i=1}^{k} \frac{1}{\sigma E_i} \sum_{l_i=1}^{n} \sum_{(l_1,\ldots,l_{i-1},l_{i+1},\ldots,l_k)\in\{1,\ldots,n\}^{k-1}} \left|X_{[i]l_i} - Z_{(i)l_1l_2\ldots l_k}\right| \\ &= \sum_{i=1}^{k} \frac{1}{\sigma E_i} \sum_{l_i=1}^{n} \sum_{j=1}^{k} \sum_{h=1}^{n} m(i, l_i, j, h) \left|X_{[i]l_i} - X_{[j]h}\right| \end{aligned}$$

where $m(i, l_i, j, h)$ is the number of vectors $(l_1, \ldots, l_{i-1}, l_{i+1}, \ldots, l_k) \in \{1, \ldots, n\}^{k-1}$ such that the i^{th} order statistic from $X_{[1]l_1}, \ldots, X_{[k]l_k}$ is the j^{th} judgment ordered observation from the h^{th} cycle. (A similar representation applies to TPS.) Since for each judgment order rank, only one cycle is used, this implies that $l_j = h$, and $m(i, l_i, j, h)$ is actually the number of vectors $(l_q)_{q\in\{1,\ldots,k\}\setminus\{i,j\}} \in \{1, \ldots, n\}^{k-2}$

such that the i^{th} order statistic from $X_{[1]l_1}, \ldots, X_{[k]l_k}$ is the j^{th} judgment ordered observation from the h^{th} cycle.

In the following, we assume that there are no ties. Since (1) $m(i, l_i, j, h)$ is 0 if $i = j$ and $l_i \neq h$ and (2) $X_{[i]l_i} - X_{[j]h} = 0$ if $i = j$ and $l_i = h$, j can be assumed to be different from i:

$$TPA = \sum_{i=1}^{k} \frac{1}{\sigma E_i} \sum_{l_i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{k} \sum_{h=1}^{n} m(i, l_i, j, h) \left| X_{[i]l_i} - X_{[j]h} \right|.$$

We therefore only need the values of $m(i, l_i, j, h)$ for $i \neq j$. For $i = 1, \ldots, k$, let $a(i, j, h)$ be the number of observations in the i^{th} judgment-order stratum that are smaller than $X_{[j]h}$,

$$a(i, j, h) = \# \left\{ l \in \{1, \ldots, n\} : X_{[i]l} < X_{[j]h} \right\}.$$

Then, by using the fact that exactly $i - 1$ observations from $X_{[1]l_1}, \ldots, X_{[k]l_k}$ (of which $X_{[i]l_i}$ may be one or not) have to be smaller than the i^{th} order statistic from $X_{[1]l_1}, \ldots, X_{[k]l_k}$,

$$\begin{aligned} m(i, l_i, j, h) &= \# \Bigg\{ (l_q)_{q \in \{1,\ldots,k\} \setminus \{i,j\}} \in \{1, \ldots, n\}^{k-2} : \\ &\qquad \sum_{q \in \{1,\ldots,k\} \setminus \{i,j\}} I(X_{[q]l_q} < X_{[j]h}) = i - 1 - I(X_{[i]l_i} < X_{[j]h}) \Bigg\} \\ &= \sum_{\substack{Q \subset \{1,\ldots,k\} \setminus \{i,j\} \\ \#Q = i-1-I(X_{[i]l_i} < X_{[j]h})}} \prod_{q \in Q} a(q, j, h) \prod_{q \notin Q} (n - a(q, j, h)) \end{aligned}$$

where $I(.)$ is the indicator function.

The $a(i, j, h)$'s can be calculated efficiently by going through all kn observed values $X_{[j]h}$ in increasing order and using the fact that $a(i, j^*, h^*) = 0$ (for $i = 1, \ldots, k$) if $X_{[j^*]h^*}$ is the smallest value from the sample, as well as the following recursions, where $X_{[j]h}$ and $X_{[j']h'}$ are assumed to be two successive values of the ordered sample:

$$a(i, j', h') = \begin{cases} a(i, j, h) + 1 & \text{if } i = j, \\ a(i, j, h) & \text{if } i \neq j. \end{cases}$$

E.g., for $k = n = 8$, this approach for the computation of TPA and TPS resulted in a reduction of the computation time by a factor of approximately 1200 compared to the original algorithm.

4. POWER COMPARISON

In this section, we compare the power of our proposed tests with their leading competitors in the literature under the assumption that the parent distribution is normal. The competing tests considered here are as follows:

- Nonparametric test based on W^* developed by Frey et al. (2007), which rejects the hypothesis of perfect judgment ranking when $W^* = \sum_{i=1}^{k}\sum_{j=1}^{n} iR_{[i]j}$ is too small, where $R_{[i]j}$ is the rank of $X_{[i]j}$ among all kn observations.

- Nonparametric test based on the null probability (NP) developed by Frey et al. (2007), which rejects the hypothesis of perfect judgment ranking when the null probability of observing set rank $\left\{R_{[i]j}\right\}$ is too small.

- Nonparametric test based on J developed by Vock and Balakrishnan (2011), which rejects the hypothesis of perfect judgment ranking when $J = \sum_{h=1}^{n}\sum_{l=1}^{n}\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} I\left(X_{[i]l} > X_{[j]h}\right)$ is too large.

- Nonparametric test based on PA developed by Zamanzade et al. (2012), which rejects the hypothesis of perfect judgment ranking when $PA = \sum_{h=1}^{n^k}\sum_{i=1}^{k}\left|R^*_{[i]h} - i\right|$ is too large, where $R^*_{[i]h}$ is the rank of the i^{th} judgment ordered observation in the h^{th} permuted sample introduced in Section 3.

- Parametric test based on D developed by Zamanzade et al. (2014), which rejects the hypothesis of perfect judgment ranking when $D = \sum_{h=1}^{n}\sum_{l=1}^{n}\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} \frac{\left(X_{[i]l}-X_{[j]h}\right)I\left(X_{[i]l}>X_{[j]h}\right)}{E_{\mu=0,\sigma=1}\left(\left(X_{[i]l}-X_{[j]h}\right)I\left(X_{[i]l}>X_{[j]h}\right)\right)}$ is too large, where μ, σ are location and scale parameters, respectively.

- Most powerful rank test (MP) developed by Frey and Wang (2013). In this test, it is assumed that the alternative hypothesis of perfect judgment ranking is fully specified, i.e. the underlying distribution of the population, the scenario of imperfect ranking, and the fraction of imperfect ranking are all completely known. Then the null hypothesis is rejected when $r = \frac{P_{H_1}(W_1<W_2<\ldots<W_N)}{P_{H_0}(W_1<W_2<\ldots<W_N)}$ is too large, where W_i has the same distribution as the in-set rank of the observation with rank i among all the $N = kn$ measured values.

We assume that the parent distribution is normal with unknown mean μ and unknown variance σ^2. We don't need to estimate the parameter μ because the proposed tests are location free. The parameter σ is estimated by $\hat{\sigma} = \sqrt{\frac{1}{k-1}\sum_{i=1}^{k}\left(X_{[i]} - \bar{X}\right)^2}$ for $n = 1$ due to Stokes (1980) and by $\hat{\sigma} = \sqrt{\left(\frac{1}{nk} + \frac{1}{nk^2(n-1)}\right)\sum_{i=1}^{k}\sum_{j=1}^{n}\left(X_{[i]j} - \bar{X}_{[i].}\right)^2 + \frac{1}{k}\sum_{i=1}^{k}\left(\bar{X}_{[i].} - \bar{X}\right)^2}$ for $n > 1$ as proposed by MacEachern et al. (2002) and Perron and Sinha (2004), where $\bar{X}_{[i].}$ is the mean of the observations with judgment rank i. Obviously, these estimators of σ are equivariant and the resulting test statistics (denoted by $\widehat{TPA}, \widehat{TPS}$ and $\widehat{D}$) are scale invariant. Therefore, the critical values and powers of tests based on $\widehat{TPA}, \widehat{TPS}$ and $\widehat{D}$ don't depend on the unknown parameters μ and σ. The expected values $E\left(d_i^2\right)$ and $E\left|d_i\right|$ and critical values of the tests based on $\widehat{TPA}$ and $\widehat{TPS}$ under the assumption of normality are available on request from the authors.

In our simulation study, the comparisons are done at a significance level of $\alpha = 0.05$. However, due to the discreteness of the distribution of the non-parametric test statistics, it is not possible to attain an exact size of $\alpha = 0.05$ without randomizing. Therefore, we have used the randomized versions of those tests to make all comparisons at size $\alpha = 0.05$. For example, for $n = 1$ and $k = 5$, under the assumption of perfect ranking, $J \geq 4$ with null probability 0.03345, and $J \geq 3$ with null probability 0.12687. Thus in order to attain the significance level $\alpha = 0.05$ in a randomized test based on J, H_0 is rejected with probability one if $J \geq 4$ and with probability $\frac{0.05-0.03345}{0.12687-0.03345} = 0.177$ if $J = 3$.

We have used two different scenarios of imperfect ranking, which have been used by many researchers in the literature. The first scenario is the bivariate normal model, due to Dell and Clutter (1972), in which the variable of interest X is ordered by using a concomitant variable Y, where (X, Y) has a bivariate normal distribution with correlation coefficient λ.

The second scenario is that of a fraction of neighbor rankings, developed by Vock and Balakrishnan (2011), in which the i^{th} judgment ordered observation is either ranked perfectly with probability λ, or is confused with the $(i+1)^{\text{th}}$ or $(i-1)^{\text{th}}$ ordered observation, both with probability $\frac{\lambda}{2}$, therefore the distribution of the i^{th} judgment ordered observation under this scenario is $F_{[i]} = \frac{\lambda}{2}F_{(i-1)} + (1-\lambda)\,F_{(i)} + \frac{\lambda}{2}F_{(i+1)}$, where $F_{(0)} = F_{(1)}, F_{(k+1)} = F_{(k)}$. This imperfect ranking model could arise when the ranking process is done by using personal judgment of an expert ranker, so he may confuse the true order statistic with an adjacent one.

For power comparisons, we have extended Tables 3 and 6 of Frey and Wang (2013) to all tests introduced above and larger values of (n, k) by using Monte Carlo simulation with 100,000 repetitions. The simulation results are presented in Tables 1–2. It should be noted that in the following tables the powers of the tests based on MP, NP, W^*, J for $(n,k) = (8,2), (4,3), (2,4), (1,5)$ are directly reported from Tables 3 and 6 of Frey and Wang (2013). Furthermore, we haven't estimated the power of the MP test for $(n,k) = (4,5), (5,4)$, since this test is only applicable for small sample sizes and small set sizes because of its computational limitations.

Table 1 gives power results for the bivariate normal model. It is apparent from this table that although the NP test is the most powerful among the non-

k	n	λ	$\widehat{TPS}$ (new)	$\widehat{TPA}$ (new)	$\widehat{D}$ (Z. et al., 2014)	MP (Frey and Wang, 2013)	NP (Frey et al., 2007)	W^* (Frey et al., 2007)	J (V. and Balakr., 2011)	PA (Z. et al., 2012)
2	8	0.9	.1047	.1042	.1060	.1019	.1018	.1000	.1000	.1000
		0.8	.1880	.1815	.1815	.1722	.1720	.1676	.1676	.1676
		0.7	.2715	.2715	.2710	.2566	.2563	.2490	.2490	.2490
		0.6	.3679	.3642	.3671	.3496	.3493	.3394	.3394	.3394
		0.5	.4656	.4694	.4678	.4458	.4456	.4337	.4337	.4337
3	4	0.9	.1265	.1384	.1302	.1317	.1316	.1294	.1289	.1271
		0.8	.2400	.2465	.2487	.2401	.2400	.2346	.2336	.2344
		0.7	.3599	.3729	.3628	.3596	.3594	.3509	.3496	.3515
		0.6	.4678	.4858	.4947	.4777	.4776	.4669	.4655	.4625
		0.5	.5891	.6002	.5959	.5867	.5866	.5748	.5783	.5734
4	2	0.9	.1405	.1425	.1509	.1491	.1420	.1403	.1372	.1398
		0.8	.2599	.2556	.2663	.2555	.2553	.2511	.2448	.2477
		0.7	.3678	.3894	.3882	.3720	.3718	.3653	.3568	.3613
		0.6	.4776	.4811	.4984	.4819	.4818	.4737	.4640	.4697
		0.5	.5763	.5808	.6006	.5808	.5806	.5718	.5617	.5650
5	1	0.9	.1367	.1365	.1485	.1366	.1363	.1358	.1283	.1254
		0.8	.2413	.2355	.2582	.2335	.2332	.2316	.2174	.2113
		0.7	.3428	.3335	.3658	.3287	.3287	.3260	.3071	.2960
		0.6	.4345	.4242	.4615	.4182	.4182	.4147	.3929	.3779
		0.5	.5213	.5103	.5456	.5003	.5003	.4962	.4731	.4532
5	4	0.9	.2697	.2681	.2843	–	.2969	.2866	.2845	.2846
		0.8	.5471	.5437	.5538	–	.5811	.5649	.5625	.5632
		0.7	.7590	.7577	.7541	–	.7820	.7665	.7651	.7648
		0.6	.8841	.8841	.8733	–	.8934	.8818	.8819	.8816
		0.5	.9494	.9395	.9380	–	.9506	.9436	.9434	.9436
4	5	0.9	.2247	.2286	.2181	–	.2261	.2224	.2231	.2182
		0.8	.4508	.4664	.4387	–	.4618	.4520	.4523	.4462
		0.7	.6446	.6633	.6354	–	.6634	.6526	.6536	.6469
		0.6	.7880	.8058	.7831	–	.8022	.7910	.7927	.7884
		0.5	.8943	.8945	.8785	–	.8944	.8846	.8873	.8846

Table 1: Power estimates of different level 0.05 tests, under the concomitant model with correlation coefficient λ.

parametric tests, the powers of the test based on W^* are quite close to it. On the other hand, the proposed tests and the test based on $\widehat{D}$ have the best powers in this scenario, and the differences between their powers are not considerable.

Powers of the tests for the scenario of neighbor rankings are presented in Table 2. This table shows that the test based on PA and the NP test are the best

k	n	λ	$\widehat{TPS}$ (new)	$\widehat{TPA}$ (new)	$\widehat{D}$ (Z. et al., 2014)	MP (Frey and Wang, 2013)	NP (Frey et al., 2007)	W^* (Frey et al., 2007)	J (V. and Balakr., 2011)	PA (Z. et al., 2012)
2	8	0.2	.1943	.1868	.1862	.1775	.1706	.1590	.1590	.1590
		0.4	.3791	.3796	.3731	.3499	.3444	.3243	.3243	.3243
		0.6	.5689	.5694	.5666	.5369	.5345	.5134	.5134	.5134
		0.8	.7328	.7358	.7350	.7066	.7062	.6897	.6897	.6897
		1	.8486	.8587	.8573	.8377	.8377	.8275	.8275	.8275
3	4	0.2	.1349	.1369	.1249	.1285	.1248	.1189	.1201	.1212
		0.4	.2483	.2524	.2164	.2371	.2266	.2122	.2173	.2206
		0.6	.3985	.4057	.3287	.3679	.3475	.3230	.3345	.3450
		0.8	.5157	.5324	.4484	.5085	.4771	.4432	.4622	.4789
		1	.6481	.6684	.5571	.6453	.6048	.5640	.5899	.6092
4	2	0.2	.1064	.1065	.0965	.0976	.0956	.0932	.0928	.0975
		0.4	.1755	.1774	.1508	.1581	.1514	.1447	.1456	.1516
		0.6	.2532	.2593	.2112	.2307	.2159	.2034	.2070	.2223
		0.8	.3416	.3514	.2774	.3136	.2875	.2678	.2757	.2987
		1	.4321	.4495	.3474	.4041	.3643	.3367	.3500	.3822
5	1	0.2	.0854	.0865	.0804	.0785	.0772	.0766	.0757	.0763
		0.4	.1275	.1282	.1156	.1109	.1066	.1051	.1040	.1064
		0.6	.1729	.1759	.1511	.1473	.1381	.1354	.1346	.1387
		0.8	.2210	.2255	.1887	.1878	.1713	.1673	.1673	.1733
		1	.2736	.2818	.2311	.2218	.2061	.2007	.2016	.2119
5	4	0.2	.1454	.1468	.1155	–	.1277	.1166	.1199	.1226
		0.4	.2733	.2799	.1983	–	.2326	.2041	.2163	.2259
		0.6	.4220	.4248	.2885	–	.3620	.3098	.3365	.3563
		0.8	.5664	.5791	.3877	–	.4959	.4266	.4668	.4954
		1	.7003	.7196	.4908	–	.6262	.5441	.5963	.6324
4	5	0.2	.1584	.1544	.1275	–	.1408	.1309	.1347	.1334
		0.4	.3072	.3051	.2260	–	.2661	.2381	.2515	.2562
		0.6	.4673	.4757	.3396	–	.4209	.3741	.4009	.4126
		0.8	.6299	.6441	.4625	–	.5750	.5123	.5537	.5716
		1	.7669	.7868	.5844	–	.7137	.6446	.6937	.7167

Table 2: Power estimates of different level 0.05 tests, under a fraction λ of neighbor rankings under assumption of normality.

nonparametric tests. The test based on $\widehat{TPA}$ is the most powerful test for this imperfect ranking scenario, while the powers of the test based on $\widehat{TPS}$ are quite close. It is worth mentioning that in this scenario, the power difference among the proposed tests and the other tests are considerable in most cases.

The simulation study was also performed for two more imperfect ranking models (fraction of inverse rankings and fraction of random rankings; see, e.g., Zamanzade et al., 2014) as well as under the assumption of an exponential instead of a normal distribution. We do not report these simulation results due to space restrictions. However, they are available on request from the authors.

Remark 4.1. It is important to notice that the proposed tests have the advantage that randomization is not needed to obtain the tests of a specific size. For the nonparametric tests and the MP test, randomization is used in the simulations for a more meaningful power comparison, but when using a non-randomized version of these tests in practice, the power will be lower. For example, in the bivariate normal model, for $n = 1, k = 5$, and $\lambda = 0.5$, the estimated powers (based on 100000 repetitions) of non-randomized nonparametric tests using NP, W^*, J, and PA at a nominal level of $\alpha = 0.05$ are 0.460, 0.460, 0.435, and 0.316, respectively, which are lower that their reported values in Table 1, where the randomized tests are used.

Remark 4.2. It is worth mentioning that although the MP test has reasonably good powers in most cases, the application of this test is too restricted in practice. It should be noted that this test can only be used in practice if the underlying distribution of the population, the scenario of imperfect ranking and the fraction of imperfect rankings (λ) are all completely known. Since these conditions, especially the last one, are hardly conceivable to be satisfied, this test cannot be used in many parametrical situations in practice.

5. CONCLUSION

In this paper, we developed two parametric and location-scale free tests of perfect judgment ranking based on ordered ranked set samples. Our tests are based on the idea that if the assumption of perfect ranking is satisfied, then the difference between ranked set samples and ordered ranked set samples should be small. Then we generalized our proposed tests to the multi-cycle case of BRSS. Finally, we compared our tests with their best known competitors in the literature. Our power comparisons indicate that the proposed tests have good performance in comparison with their leading competitors, especially under the fraction of neighbor rankings model.

It is worth mentioning that although we confine ourselves to the balanced ranked set samples, the proposed tests can straightforwardly be generalized to the unbalanced case.

ACKNOWLEDGMENTS

The authors are thankful to Professor Jesse Frey for his comments on an earlier draft of this paper. The authors are also grateful to an anonymous referee and an associate editor for their valuable comments and suggestions which lead to this improved version.

REFERENCES

[1] Frey, J.; Ozturk, O., and Deshpande, J.V. (2007). Nonparametric tests for perfect judgment rankings, *Journal of the American Statistical Association*, **102**, 708–717.

[2] Frey, J. and Wang, L. (2013). Most powerful rank tests for perfect rankings, *Computational Statistics and Data Analysis*, **60**, 157–168.

[3] 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**, 1325–1338.

[4] 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.

[5] McIntyre, G. A. (1952). A method for unbiased selective sampling using ranked set sampling, *Australian Journal of Agricultural Research*, **3**, 385–390.

[6] McIntyre, G. A. (2005). A method for unbiased selective sampling using ranked set sampling, *The American Statistician*, **59**, 230–232. Originally appeared in Australian Journal of Agricultural Research 3, 385–390.

[7] 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.

[8] Stokes, S. L. (1980). Estimation of variance using judgment ordered ranked set samples, *Biometrics*, **36**, 1, 35–42.

[9] 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**, 624–630.

[10] Vock, M. and Balakrishnan, N. (2013). A connection between two nonparametric tests for perfect ranking in balanced ranked set sampling, *Communications in Statistics – Theory and Methods*, **42**, 191–193.

[11] Wolfe, D. (2012). Ranked set sampling: Its relevance and impact on statistical inference, *ISRN Probability and Statistics*, vol. 2012, Article ID 568385, 32 pages, 2012. doi:10.5402/2012/568385.

[12] Zamanzade, E.; Arghami, N.R. and Vock, M. (2012). Permutation-based tests of perfect ranking, *Statistics & Probability Letters*, **82**, 3313–2220.

[13] Zamanzade, E.; Arghami, N.R. and Vock, M. (2014). A parametric test of perfect ranking in balanced ranked set sampling, *Communications in Statistics – Theory and Methods*, **43**, 4589–4611.