Resampling Unbalanced Ranked Set Samples With Applications in Testing Hypothesis About the Population Mean

Saeid AMIRI, Mohammad JAFARI JOZANI, and Reza MODARRES

Ranked set sampling is a sampling approach that could lead to improved statistical inference when the actual measurement of the variable of interest is difficult or expensive to obtain but sampling units can be easily ordered by some means without actual quantification. In this paper, we consider the problem of bootstrapping an unbalanced ranked set sample (URSS) where the number of observations from each artificially created stratum can be unequal. We discuss resampling a URSS through transforming it into a balanced RSS and extending the existing algorithms. We propose two methods that are designed to obtain resamples from the given URSS. Algorithms are provided and several properties, including asymptotic normality of estimates, are discussed. The proposed methods are compared with the parametric bootstrap using Monte Carlo simulations for the problem of testing a hypothesis about the population mean.

Key Words: Bootstrap transformation; Monte Carlo simulation; Ranked set sample.

1. INTRODUCTION

Ranked set sampling is a sampling approach that could lead to improved statistical inference for many situations where the actual measurement of the variable of interest is difficult or expensive to obtain but sampling units can be easily ordered by some means without actual quantification. Ranked set sampling is a two-stage sampling plan where a number of sampling units are first ordered with respect to a variable without taking actual measurements of the characteristics of interest at a small cost and, in the second stage, measurements are taken from a fraction of the ranked units. Introduced by McIntyre (1952), the study of ranked set sampling has resulted in a substantial literature. The existing results include, but not limited to, works on hypothesis testing, point estimation and interval

Saeid Amiri (✉) is Post Doctoral Associate, Devision of Biostatistics, Department of Epidemiology and Public Health, University of Miami, Miami, USA (E-mail: *saeid.amiri1@gmail.com*). Mohammad Jafari Jozani is Assistant Professor of Statistics, Department of Statistics, University of Manitoba, Winnipeg, MB, Canada, R3T 2N2. Reza Modarres is Professor of Statistics, Department of Statistics, The George Washington University, Washington, DC, USA.

Journal of Agricultural, Biological, and Environmental Statistics, Volume 19, Number 1, Pages 1–17
DOI: 10.1007/s13253-013-0153-y

estimation under both parametric and nonparametric settings. See, for example, Bohn and Wolfe (1992), Wolfe (2004), Fligner and MacEachern (2006), Frey (2007) and references therein. Chen, Bai, and Sinha (2004), henceforth referred to as CBS, provide invaluable information on ranked set sampling, its many variants and applications. There are many applications for ranked set sampling designs in ecological and environmental studies (e.g., Dell and Clutter 1972; Al-Saleh and Zheng 2002), reliability theory (Kvam and Samaniego 1994) and medical studies (Samawi and Al-Sagheer 2001), among others.

In this paper, we consider the problem of resampling an unbalanced ranked set sample (URSS) using the bootstrap method. The bootstrap has become a standard tool in statistical analysis. There have been several good books on the bootstrap, including Efron and Tibshirani (1993), Davison and Hinkley (1997), Shao and Tu (1996) and Hall (1992), each with a different perspective. Since the empirical distribution function (EDF) serves as a good approximation to the population distribution function, the bootstrap can be used to obtain the sampling distribution of a statistic of interest. Bootstrap allows for estimation of the standard error of any well-defined statistic and enables one to draw inferences when the exact or the asymptotic distribution of the statistic of interest is unavailable. Chen (2001) and CBS (2004) describe an algorithm for drawing inferences for trimmed means and Modarres, Hui, and Zhang (2006) explore resampling techniques for balanced RSS. However, as we will show, the EDF of URSS does not converge to the underlying distribution and the algorithms developed for bootstrapping balanced RSS cannot be applied for URSS situation. One possible approach to side-step this difficulty is to transform the URSS to a balanced RSS using a transformation (BTR) that involves an initial step of resampling within strata to create a balanced RSS. This BTR is included in each bootstrap run as part of new bootstrap algorithms that are designed to allow for URSS. Our impetus in development of BTR is to obtain a bootstrap test of the hypothesis $H_0 : \mu = \mu_0$ for the mean of the underlying population (μ) and the associated confidence interval based on a URSS. We pursue this goal through Monte Carlo simulation.

In Section 2, we discuss URSS, define BTR, and investigate its asymptotic properties while the proofs are presented in the Appendix. Section 3 examines two methods of resampling and presents algorithms to bootstrap a URSS. Section 4 describes a simulation study to observe the finite sampling properties of the proposed methods, which are used for testing a hypothesis concerning the population mean. Several test statistics are proposed and their performances are compared to each other and a test based on parametric bootstrap. The comparisons are made based on the observed significance level and power of the tests under location shift and three distributions. In Section 5, we give an application of our results using a real data set. Section 6 provides some concluding comments. Finally, the Appendix is devoted to the proofs and some of the necessary theoretical results.

2. UNBALANCED RSS AND THE BOOTSTRAP TRANSFORMATION

Applications that lead to URSS data include reliability, environmental or medical studies and missing data (CBS 2004). While they have many differences, the structure of a ranked set sample is similar to the classical stratified sampling. In the first stage of the ranked set sampling, a small number of sampling units are identified and ranked, and in the second stage, measurements are taken from a fraction of the ranked units. In this section, we first give a general description of the unbalanced ranked set sampling design and then show how the bootstrap method can be used to introduce a bootstrap transformation for converting a URSS to a balanced RSS.

Suppose a total number of n units are to be measured from the underlying population on the variable of interest. Let n sets of units, each of size k, be randomly chosen from the population using a simple random sampling (SRS) technique. Then the units of each set are ranked by any means other than actual quantification of the variable of interest. Finally, one and only one unit in each ordered set with a pre-specified rank is measured on the variable of interest. Let m_r be the number of measurements on units with rank r, $r = 1, \ldots, k$, such that $n = \sum_{r=1}^{k} m_r$. Let $X_{(r)j}$ denote the measurement on the jth measured unit with rank r. This results in a URSS of size n from the underlying population as $\{X_{(r)j}; r = 1, \ldots, k, j = 1, \ldots, m_r\}$. Note that when $m_r = m$, $r = 1, \ldots, k$, then URSS reduces to the balanced RSS. It is worth mentioning that, in ranked set sampling designs, $X_{(1)j}, \ldots, X_{(k)j}$ are independent order statistics (as they are obtained from independent sets) and each $X_{(r)j}$ provides information about a different stratum of the population. One can represent the structure of a URSS as follows:

$$\begin{aligned}
\mathcal{X}_1 &= \{X_{(1)1}, X_{(1)2}, \ldots, X_{(1)m_1}\} \overset{\text{i.i.d.}}{\sim} F_{(1)}, \\
\mathcal{X}_2 &= \{X_{(2)1}, X_{(2)2}, \ldots, X_{(2)m_2}\} \overset{\text{i.i.d.}}{\sim} F_{(2)}, \\
&\vdots \\
\mathcal{X}_k &= \{X_{(k)1}, X_{(k)2}, \ldots, X_{(k)m_k}\} \overset{\text{i.i.d.}}{\sim} F_{(k)},
\end{aligned} \tag{2.1}$$

where $F_{(r)}$ is the distribution function (DF) of the rth-order statistic. Also, the EDF of a URSS is defined as (Chen 2001):

$$\widehat{F}_{q_n}(t) = \frac{1}{n} \sum_{r=1}^{k} \sum_{j=1}^{m_r} I(X_{(r)j} \leq t) = \sum_{r=1}^{k} q_{m_r} \widehat{F}_{(r)}(t), \tag{2.2}$$

where $n = \sum m_r$ and $q_{m_r} = m_r/n$. Similarly, the EDF of a balanced RSS is

$$\widehat{F}_n(t) = \frac{1}{km} \sum_{r=1}^{k} \sum_{j=1}^{m} I(X_{(r)j} \leq t) = \frac{1}{k} \sum_{r=1}^{k} \widehat{F}_{(r)}(t). \tag{2.3}$$

We will show that if $n \longrightarrow \infty$, and $q_{m_r} \longrightarrow q_r$, $r = 1, \ldots, k$, then $\widehat{F}_{q_n}(t) \longrightarrow F_q(t)$, where

$$F_q(t) = \sum_{r=1}^{k} q_r F_{(r)}(t). \tag{2.4}$$

It is easy to see that $F_q(t) = F(t)$, for all $t \in \mathbb{R}$, whenever $q_r = 1/k$, $r = 1, \ldots, k$. Let Γ_p be the set of all distribution functions F with $\int |x|^p \, dF(x) < \infty$ and define the metric d_p on Γ_p as the infimum of $\sqrt{E(|X - Y|^p)}$ over all pairs of random variables X and Y with marginal distributions F and G, respectively. We pursue with the following results, whose proofs are presented in the Appendix.

Proposition 1. *If F_q has a continuous density function and $q_{m_r} \longrightarrow q_r$, then $\widehat{F}_{(r)}(t)$ converges to $F_{(r)}(t)$ for all t almost surely (a.s.), $r = 1, \ldots, k$. Furthermore, $\widehat{F}_{q_n}(t) - F_q(t)$ converges to zero a.s. for all t as $\min(m_r)$ approaches ∞ where $\widehat{F}_{q_n}(t)$ is the EDF of URSS given in (2.3) and F_q is the mixture distribution given in (2.4).*

Proposition 2. *If $F_q \in \Gamma_2$ and $\widehat{F}_{q_n}$ defined as in (2.2) then $d_2(\widehat{F}_{q_n}, F_q) \longrightarrow 0$ a.s.*

In many cases, the statistical procedures for SRS can be directly extended to the balanced ranked set sampling design (e.g., Chen 2001 and CBS 2004). A fundamental equation for this to happen is the equality $F(t) = \frac{1}{k}\sum_{r=1}^{k} F_{(r)}(t)$. In unbalanced ranked set sampling the process is more complicated and since the balanced structure of the design is destroyed, the statistical procedures for SRS cannot directly be extended to this case. As previously mentioned, one possible approach to side-step this difficulty is to transform the URSS to a balanced RSS before applying algorithms that are designed for RSS.

2.1. BOOTSTRAP TRANSFORMATION

We now introduce the bootstrap transformation RSS = BTR(URSS, N), which is used to convert a URSS to a balanced RSS. Let $\mathcal{X} = \{\mathcal{X}_1, \ldots, \mathcal{X}_k\}$ be the URSS as in (2.1) and $\mathcal{X}_i^\diamond$ denote a sample of size N from $\mathcal{X}_i$. The bootstrap is used to obtain a resample of size N within each stratum in order to create a balanced RSS as follows:

$$\begin{aligned}
\mathcal{X}_1^\diamond &= \left\{X_{(1)1}^\diamond, X_{(1)2}^\diamond, \ldots X_{(1)N}^\diamond\right\} \sim \widehat{F}_{(1)},\\
\mathcal{X}_2^\diamond &= \left\{X_{(2)1}^\diamond, X_{(2)2}^\diamond, \ldots X_{(2)N}^\diamond\right\} \sim \widehat{F}_{(2)},\\
&\cdots\\
\mathcal{X}_k^\diamond &= \left\{X_{(k)1}^\diamond, X_{(k)2}^\diamond, \ldots X_{(k)N}^\diamond\right\} \sim \widehat{F}_{(k)}.
\end{aligned} \tag{2.5}$$

Let $\mathcal{X}^\diamond = \{\mathcal{X}_1^\diamond, \ldots, \mathcal{X}_k^\diamond\}$ and denote its EDF by

$$\widehat{F}_N^\diamond(t) = \frac{1}{Nk}\sum_{r=1}^{k}\sum_{j=1}^{N} I\left(X_{(r)j}^\diamond \le t\right) = \frac{1}{k}\sum_{r=1}^{k}\widehat{F}_{(r)}^\diamond(t), \tag{2.6}$$

where $\widehat{F}_{(r)}^\diamond(t)$ is the EDF of $\mathcal{X}_r^\diamond$, $r = 1, \ldots, k$. There are many possible choices for N. We consider three of them, including $N = \text{mean}\{m_r, r = 1, \ldots, k\}$, in Section 4. If N is not an integer then the closest integer less than N is used. As we show in the Appendix (see Lemma A.2), when N approaches ∞, $\sup_{t\in\mathbb{R}} |\widehat{F}_{(r)}^\diamond(t) - \widehat{F}_{(r)}(t)|$ converges to zero for all $t \in \mathbb{R}$ and all $r = 1, \ldots, k$. This, along with the results in Proposition A.1 and Corollaries A.1 and A.2 (see the Appendix), validates the use of the bootstrap transformation RSS = BTR(URSS, N) to transform a URSS to a balanced RSS.

3. BOOTSTRAPPING UNBALANCED RANKED SET SAMPLES

Once the bootstrap transformation is performed and the URSS is transformed to a balanced RSS via RSS $=$ BTR(URSS, N), one can use the bootstrapping techniques available for balanced RSS to resample from the transformed data. Two algorithms, both of which are modifications of Modarres, Hui, and Zhang (2006) algorithms for balanced RSS, are presented below. Note that the bootstrap transformation is included in each bootstrap run of the following algorithms. The first algorithm resamples within each row separately while the second algorithm resamples from the pooled data.

3.1. ALGORITHM BTR1

1. Apply RSS $=$ BTR(URSS, N) to obtain $\mathcal{X}_r^\diamond = \{X_{(r)j}^\diamond, r = 1, \ldots, k, j = 1, \ldots, N\}$.
2. Assign probability $1/N$ to each element of $\mathcal{X}_r^\diamond$.
3. Select N elements with replacement and denote by $X_{(r)1}^{*\diamond}, \ldots, X_{(r)N}^{*\diamond}$.
4. Perform Step 3 for $r = 1, \ldots, k$ to generate a bootstrap RSS $\{X_{(r)j}^{*\diamond}\}$.
5. Repeat the above steps B times.

We will refer to this algorithm as BTR1. This algorithm, without the transformation, is referred to as B-RSS-R by Modarres, Hui, and Zhang (2006). We need the following result which is proved in Modarres, Hui, and Zhang (2006).

Proposition 3. *Suppose $F_\theta \in \Gamma_2$ where θ is a location parameter and $\widehat{F}_{(r)}$ is the EDF of the rth row where the resampling plan is BTR1. If $\upsilon_i^* = \sqrt{m}_i(\theta(\widehat{F}_{(i)}^*) - \theta(\widehat{F}_{(i)}))$, then $(\upsilon_1^*, \ldots, \upsilon_k^*)$ converges in distribution to a multivariate normal distribution with mean vector zero and covariance matrix* $\operatorname{diag}(\sigma(F_{(1)}), \ldots, \sigma(F_{(k)}))$ *where $\sigma(F_{(i)}) = \int (X - \mu_{(i)})^2\, dF_{(i)}$ and $\mu_{(i)} = \int x\, dF_{(i)}(x)$.*

To test $H_0 : \mu = \mu_0$, using URSS, Proposition 3 suggests the statistic $\frac{1}{\sqrt{k}} \sum_{r=1}^k \sqrt{m_r}(\bar{X}_{(r)} - \mu_{(r)})$ where $\mu_{(r)}$ is the expected value of the rth-order statistic under H_0. Using Corollary 1, we propose the test statistic $T_2(X) = \frac{1}{k} \sum_{r=1}^k (\frac{\bar{X}_{(r)} - \mu_0}{S_1})$, where $S_1^2 = \frac{1}{k^2} \sum_{r=1}^k \frac{S^2(X_{(r)})}{m_r}$. The corollary shows how the BTR1 algorithm can be used to carry out the a test based on T_2 using URSS data.

Corollary 1. *Suppose $F_\theta \in \Gamma_2$ where θ is a location parameter and $\widehat{F}_{(r)}^\diamond$ is the EDF of the rth row where the resampling plan is BTR1. If $\upsilon_i^* = \sqrt{N}(\theta(\widehat{F}_{(i)}^*) - \theta(\widehat{F}_{(i)}^\diamond))$, then $(\frac{\upsilon_1^*}{\sigma(\widehat{F}_{(1)})}, \ldots, \frac{\upsilon_k^*}{\sigma(\widehat{F}_{(k)})})$ converges in distribution to a multivariate normal distribution with mean vector zero and identity covariance matrix, where $\sigma(\widehat{F}_{(i)}) = \int (X - \bar{X}_{(i)})^2\, d\widehat{F}_{(i)}$.*

Using Corollary 1, $\frac{1}{k} \sum_{r=1}^k \sqrt{N}(\bar{X}_{(r)}^{*\diamond} - \bar{X}_{(r)}^\diamond) \sim AN(0, \frac{1}{k} \sum_{r=1}^k \sigma^2(\bar{X}_{(r)}^{*\diamond}))$. One can consider the test statistic

$$T_{\text{BTR1-T1}}^*\big(X_b^{*\diamond}, X^\diamond\big) = \frac{1}{k} \sum_{r=1}^k \sqrt{N}\big(\bar{X}_{(r)}^{*\diamond} - \bar{X}_{(r)}^\diamond\big), \tag{3.1}$$

where $\bar{X}^{*\diamond}_{(r)} = \frac{1}{N}\sum_{j=1}^{N} X^{*\diamond}_{(r)j}$ and $\bar{X}^{\diamond}_{(r)} = \frac{1}{N}\sum_{j=1}^{N} X^{\diamond}_{(r)j}$. However, $T^{*}_{\text{BTR1-T1}}$ is not a pivotal quantity and the appropriate test statistic for testing $H_0 : \mu = \mu_0$ is

$$T^{*}_{\text{BTR1-T2}}\big(X^{*\diamond}_{b}, X^{\diamond}\big) = \frac{1}{k}\sum_{r=1}^{k}\sqrt{N}\left(\frac{\bar{X}^{*\diamond}_{(r)} - \bar{X}^{\diamond}_{(r)}}{S^{*\diamond}_{1}}\right), \tag{3.2}$$

where $S^{2*\diamond}_{1} = \frac{1}{k^2}\sum_{r=1}^{k} S^2(\bar{X}^{*\diamond}_{(r)})$. Later in Section 4 these test statistics are compared using simulation studies.

3.2. ALGORITHM BTR2

A second algorithm that works well under RSS combines and resamples all the observations as explained in Modarres, Hui, and Zhang (2006). Here again the bootstrap transformation is included in each bootstrap run of the algorithm. To this end,

1. Apply RSS = BTR(URSS, N) to obtain $\{X^{\diamond}_{(r)j}, r = 1, \dots, k, j = 1, \dots, N\}$.
2. Combine all the observations to form $\mathcal{X}^{\diamond}$ and assign the probability of $1/kN$ to each element of $\mathcal{X}^{*}$.
3. Randomly draw $Y_1, \dots, Y_k$ from $\mathcal{X}^{\diamond}$, order them as $Y_{(1)} \le \dots \le Y_{(k)}$ and retain $X^{*\diamond}_{(r)1} = Y_{(r)}$.
4. Perform Step 3 for $r = 1, \dots, k$.
5. Repeat Steps 1–4, N times to obtain $\{X^{*\diamond}_{(r)j},\ j = 1, \dots, N\}$.
6. Repeat all steps B times to obtain the bootstrap samples.

Consider $X_1, \dots, X_n \sim F_1$ and $Y_1, \dots, Y_m \sim F_2$ with EDFs $\widehat{F}_n$ and $\widehat{F}_m$, respectively. Resampling $n + m$ observations from the combined two samples, $\{X_1, \dots, X_n, Y_1, \dots, Y_m\}$, is equivalent to resampling from the mixture distribution $\widehat{F}(x) = \frac{n}{n+m}\widehat{F}_n(x) + \frac{m}{n+m}\widehat{F}_m(x)$. The following proposition due to Boos, Janssen, and Veraverbeke (1989) will aid us in obtaining a suitable bootstrap test statistic for testing the null hypothesis $H_0 : \mu = \mu_0$ based on the algorithm BTR2.

Proposition 4. *Let $X_1, \dots, X_n$ and $Y_1, \dots, Y_m$ be independent random samples from distributions F_1 and F_2 with means μ_1 and μ_2, respectively. Consider V-statistics of the form $V_1 = \sqrt{n}(\theta(\widehat{F}_1) - \theta(F_1))$ and $V_2 = \sqrt{m}(\theta(\widehat{F}_2) - \theta(F_2))$, where $\theta(\cdot)$ is a von Mises function of the form $\theta(K) = \int\int h(x, y)\,dK(x)\,dK(y)$, $h(\cdot,\cdot)$ is a symmetric kernel and K is a DF with $\psi(x, K) = 2[\int h(x, y)\,dK(y) - \theta(K)]$. If $n/(n + m) \longrightarrow \lambda$ as $n, m \longrightarrow \infty$, then (V^{*}_1, V^{*}_2) converges in distribution to a bivariate normal $N(0, \mathrm{diag}(\sigma^2(H), \sigma^2(H)))$ where $\sigma^2(H) = \int \psi^2(x, H)\,dH(x)$ and $H(x) = \lambda F_1(x) + (1 - \lambda)F_2(x)$.*

The following proposition is obtained by setting $h(x, y) = \frac{x+y}{2}$ in Proposition 4 so that $\theta(K) = \mu$, and $\psi(X, K) = X$.

Proposition 5. *Suppose $F \in \Gamma_2$ and $\widehat{F}_{(r)}$ is the EDF of the rth row where the resampling plan is BTR2. Let $\upsilon^{*}_{i} = \sqrt{m}_i(\theta(\widehat{F}^{*}_{(i)}) - \theta(\widehat{F}_{(i)}))$. Then $(\upsilon^{*}_1, \dots, \upsilon^{*}_k)$ converges* *in distribution to a multivariate normal distribution* $N(0, \operatorname{diag}(\sigma^2(\widehat{F}_{(1)q}), \ldots, \sigma^2(\widehat{F}_{(k)q})))$, *where* $\sigma^2(\widehat{F}_{(r)q}) = \int (X - \mu_q)^2 \, d\widehat{F}_{(r)q}(X)$ *and* $\widehat{F}_q(x) = \sum_{r=1}^{k} q_r \widehat{F}_{(r)}(x)$.

The above result, when $m_i = m$, reduces to the case of balanced RSS. For testing the null hypothesis $H_0 : \mu = \mu_0$, the test statistic based on the balanced RSS of size mk is $\frac{1}{\sqrt{k}} \sum_{r=1}^{k} \sqrt{m}(\bar{X}_{(r)} - \mu_{(r)}) = \sqrt{mk}(\bar{X} - \mu)$. However, the test statistic cannot directly be used for URSS to test H_0 because $\mu_{(r)}$ must be determined under the null hypothesis. To sidestep this difficulty one propose to use either the test statistic $T_2(X)$ as defined earlier o the test statistic $T_3(X) = \frac{1}{k} \sum_{r=1}^{k} (\frac{\bar{X}_{(r)} - \mu_0}{S_2})$, where S_2^2 is the variance of the pooled sample. Using Corollary 2, one can easily apply the BTR2 algorithm to carry out a test based on T_3 for testing H_0 against H_1.

Corollary 2. *Suppose* $F \in \Gamma_2$ *and* $\widehat{F}_{(r)}$ *is the EDF of the* r*th row where the sampling plan is BTR2. Let* $\upsilon_i^* = \sqrt{N}(\theta(\widehat{F}_{(i)}^{*\diamond}) - \theta(\widehat{F}_{(i)}^{\diamond}))$. *Then* $(\upsilon_1^*, \ldots, \upsilon_k^*)$ *converges in distribution to a multivariate normal distribution* $N(0, \operatorname{diag}(\sigma^2(\widehat{F}_q), \ldots, \sigma^2(\widehat{F}_q)))$, *where* $\sigma^2(\widehat{F}) = \int (X - \bar{X}_q)^2 \, d\widehat{F}(X)$, $\widehat{F}_q(x) = \frac{1}{k} \sum_{r=1}^{k} \widehat{F}_{(r)}(x)$, $\bar{X}_q = \frac{1}{k} \sum_{r=1}^{k} \bar{X}_{(r)} = \bar{X}$ *and* $\bar{X}_{(r)}$ *is the sample mean of* r*th ordered sample.*

According to Corollary 2, the test statistic

$$T^*_{\text{BTR2-T3}}(X_b^{*\diamond}, X^{\diamond}) = \frac{1}{k} \sum_{r=1}^{k} \sqrt{N} \left(\frac{\bar{X}_{(r)}^{*\diamond} - \bar{X}_{(r)}^{\diamond}}{S_2^{*\diamond}} \right), \tag{3.3}$$

where $S_2^{2*\diamond}$ is the variance of the pooled samples, is the appropriate test statistic for testing $H_0 : \mu = \mu_0$. Another possible test statistic is $T^*_{\text{BTR2-T3}}$ without the standard deviation in the denominator as $T^*_{\text{BTR2-T1}} = \frac{1}{k} \sum_{r=1}^{k} \sqrt{N}(\bar{X}_{(r)}^{*\diamond} - \bar{X}_{(r)}^{\diamond})$.

4. SIMULATION STUDY

We compare the finite sample performance of BTR1 and BTR2, both of which are nonparametric resampling methods, with a parametric bootstrap (PB) procedure to test the hypothesis $H_0 : \mu = \mu_0$, where μ is the unknown parameter of interest and μ_0 is a known constant. In order to carry out the bootstrap test, we follow the guidelines provided by Hall and Wilson (1991). The first guideline recommends to perform resampling in a way that reflects the null hypothesis, while the second recommends using $T^* = (\theta^* - \widehat{\theta})/S(\theta^*)$ as a suitable test statistic. These guidelines are considered to carry out the tests for the hypothesis $H_0 : \mu = \mu_0$. We use the following test statistics:

$$T_1(X) = \frac{1}{k}\sum_{r=1}^{k}(\bar{X}_{(r)} - \mu_0), \tag{4.1}$$

$$T_2(X) = \frac{1}{k}\sum_{r=1}^{k}\left(\frac{\bar{X}_{(r)} - \mu_0}{S_1}\right), \tag{4.2}$$

$$T_3(X) = \frac{1}{k}\sum_{r=1}^{k}\left(\frac{\bar{X}_{(r)} - \mu_0}{S_2}\right), \tag{4.3}$$

where $S_1^2 = \frac{1}{k^2}\sum_{r=1}^{k}\frac{S^2(X_{(r)})}{m_r}$ and $S_2^2 = \frac{1}{mk-1}\sum_{r=1}^{k}\sum_{j=1}^{m}(X_{(r)j} - \bar{X})^2$, as in Corollaries 1 and 2, respectively. Also, the existing tests developed by Modarres, Hui, and Zhang (2006), which are denoted by B-RSS-R and B-RSS, are studied simultaneously in comparison with the proposed tests. To perform these methods on URSS data, we obtain $T_i(X)$ using $\bar{X}_{(r)}$ of the original URSS/RSS data. Next, we perform B-RSS-R (B-RSS) to obtain resamples and calculate $T_i^*(X^*, X)$. This process is performed B times to obtain approximate p-values. As simulations show, the straightforward application of B-RSS-R and B-RSS does not produce reliable results and the bootstrap transformation to balance the sample is needed. The nonparametric bootstrap tests using BTR1 and BTR2 for testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$ are conducted based on the following steps:

1. Let X be a URSS sample of F.
2. Calculate $T_i = T_i(X)$, $i = 1, 2, 3$.
3. Apply RSS $=$ BTR(URSS, N) and denote the resulting RSS sample by $X^\diamond = \{X^\diamond_{(r)j}\}$.
4. Carry out Steps 2–5 of BTR1 or BTR2 on $X^\diamond$ to obtain $X_b^{*\diamond} = \{X^{*\diamond}_{(r)j}\}_b$.
5. Repeat Steps 3–4 for $b = 1, \ldots, B$ and calculate $T^*_{i,b} = T(X_b^{*\diamond}, X^\diamond)$, $b = 1, \ldots, B$, as in (3.1), (3.2) and (3.3).
6. Obtain the proportion of rejections via $\frac{\#\{T^*_{i,b} > T_i\}}{B}$, that approximates the p-value.

We obtained samples from distributions (a) Normal(0, 1), (b) Logistic(1, 1) and (c) Exponential(1) to use in the simulation study which compares our developed methods in terms of the observed significance level and power of the test against location shift, $\mu_0 + \delta$. We also performed the desired testing hypothesis using PB by generating URSS samples from Normal(0, 1), Logistic(1, 1) and Exponential(1) distributions. To perform PB test we use the following steps (for more details on PB method see Efron and Tibshirani (1993)):

1. Let X be a URSS sample from a distribution F_θ where θ is the unknown parameter and let $\mu = E_\theta(X)$.
2. Calculate $T_i = T_i(X)$, $i = 1, 2, 3$.
3. Estimate θ from X and take a URSS of $F_{\widehat{\theta}}$, $X_b^* = \{X^*_{(r)j}\}_b$.

4. Calculate $T^*_{i,b} = T^*_{i,b}(X^*_b, X)$, $i = 1, 2, 3$, as in (3.1), (3.2) and (3.3).

5. Obtain the proportion of rejections via $\frac{\#\{T^*_{i,b} > T_i\}}{B}$, that approximates the p-value.

Parametric bootstrap requires estimation of all unknown parameters. We estimated the population mean using $\bar{x}$, the mean of the URSS, and used $\sigma = 1$. In particular, we generated samples from $N(\bar{x}, 1)$, Logistic$(\bar{x}, 1)$ and Exponential$(\bar{x})$. We used $B = 500$ resamples from a given sample or simulated $B = 500$ parametric bootstrap replications with $\alpha = 0.05$. The simulation experiment is then replicated 2000 times. We used several RSS and URSS designs with different sample sizes in our simulation study when $k = 5$. Each design can be shown by a $D = (m_1, m_2, \ldots, m_5)$ with $n_D = \sum_{r=1}^{5} m_r$. For example, the first design is balanced with $k = 5$ and $m_r = 5$ observations per stratum, which is denoted by

$$D_1 = (5, 5, 5, 5, 5) \quad \text{with } n_{D_1} = 25.$$

Similarly, we define the following designs:

$$D_2 = (3, 8, 5, 3, 8) \quad \text{with } n_{D_2} = 27,$$

$$D_3 = (8, 3, 5, 8, 3) \quad \text{with } n_{D_3} = 27,$$

$$D_4 = (5, 3, 6, 3, 2) \quad \text{with } n_{D_4} = 19,$$

$$D_5 = (3, 8, 6, 3, 4) \quad \text{with } n_{D_5} = 24.$$

We also examined several ways of selecting N for use in algorithms BTR1 and BTR2. When the design is balanced, e.g. D_1, we use $N = m$. For unbalanced designs, since m_i's are unequal, we considered $N_1 = \min\{m_i,\ i = 1, \ldots, k\}$, $N_2 = \text{mean}\{m_i,\ i = 1, \ldots, k\}$ and $N_3 = \max\{m_i,\ i = 1, \ldots, k\}$. If N_2 is not an integer then the closest integer to N_2 is used. The reported results are for N_3. In general, using larger N results is better performance in the simulation study.

Table 1 displays the observed α levels. The parametric bootstrap (PB) method is accurate and the estimated α levels are close to the nominal level 0.05. However, BTR1-T1, BTR1-T3, BTR2-T1 and BTR2-T3 cannot maintain their significance levels very well. The observed that α levels for BTR2-T2 follow the PB method closely. Hence, it is appropriate to compare the power of this procedure with that of PB. The simulation results support the use of the average variances, S_1^2, under BTR1 as indicated Corollary 1 and the use of the total variance, S_2^2, under BTR2 as indicated in Corollary 2. The simulation results also show that the existing B-RSS-R and B-RSS methods are not suitable for URSS and the methods can lead to quite conservative or liberal behavior.

Remark 1. As suggested by a referee, we also conducted the test procedure introduced in CBS (2004), which is based on the asymptotic normal distribution. Let $\widehat{F}_n(x) = \sum_{i=1}^{n} p_i I\{Z_{n:i \le x}\}$ where $Z_{n:i}$s are the URSS order statistics and p_is are defined such that

Table 1. Observed α-levels for the Z-procedure, T_i, $i = 1, 2, 3$, and parametric bootstrap (PB) for testing $H_0 : \mu = 0$ in the Normal distribution and $H_0 : \mu = 1$ for the Exponential and Logistic distributions.

		CBS	B-RSS-R	B-RSS	BTR1			BTR2			PB		
D	AT	Z	T_1	T_1	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
						Normal distribution							
D_1	0.056	0.001	0.075	0.050	0.102	0.044	0.094	0.052	0.048	0.045	0.051	0.050	0.038
D_2	0.068	0.004	0.190	0.029	0.158	0.061	0.149	0.117	0.066	0.109	0.049	0.053	0.047
D_3	0.062	0.002	0.026	0.238	0.168	0.057	0.160	0.106	0.059	0.104	0.046	0.049	0.050
D_4	0.080	0.002	0.028	0.300	0.192	0.062	0.195	0.111	0.072	0.122	0.050	0.056	0.048
D_5	0.061	0.004	0.062	0.185	0.188	0.054	0.198	0.131	0.059	0.141	0.046	0.046	0.046
						Exponential distribution							
D_1	0.033	0.007	0.047	0.019	0.068	0.019	0.092	0.020	0.032	0.041	0.019	0.046	0.051
D_2	0.034	0.000	0.118	0.010	0.098	0.029	0.103	0.060	0.037	0.063	0.020	0.043	0.051
D_3	0.051	0.000	0.028	0.215	0.159	0.046	0.201	0.085	0.052	0.145	0.015	0.053	0.041
D_4	0.064	0.000	0.028	0.315	0.209	0.057	0.255	0.102	0.064	0.178	0.009	0.049	0.056
D_5	0.045	0.000	0.036	0.164	0.150	0.038	0.188	0.087	0.046	0.133	0.016	0.054	0.054
						Logistic distribution							
D_1	0.062	0.004	0.079	0.058	0.104	0.044	0.104	0.056	0.054	0.054	0.051	0.051	0.54
D_2	0.077	0.005	0.191	0.032	0.159	0.069	0.157	0.120	0.074	0.117	0.048	0.051	0.58
D_3	0.056	0.003	0.025	0.237	0.153	0.048	0.152	0.099	0.049	0.103	0.041	0.048	0.55
D_4	0.066	0.002	0.026	0.293	0.189	0.055	0.191	0.108	0.058	0.129	0.046	0.047	0.45
D_5	0.065	0.004	0.063	0.183	0.180	0.055	0.191	0.127	0.061	0.144	0.046	0.045	0.43

$\widehat{F}_n(x) \xrightarrow{\text{a.s}} F(x)$ (see, CBS 2004, Chapter 4, for more details). Then, under suitable regularity conditions, the following test statistic,

$$Z = \frac{\sqrt{n}(\mu(\widehat{F}_n) - \mu_0)}{\sigma(\widehat{F}_n)} \longrightarrow N(0, 1),$$

can be used to test $H_0 : \mu = \mu_0$, where $\mu(\widehat{F}_n) = \sum_{i=1}^{n} p_i Z_{n:i}$ and $\sigma^2(\widehat{F}_n) = \sum_{i=1}^{n} p_i (Z_{n:i} - \mu(\widehat{F}_n))^2$. The results of our simulation study, presented in Table 1, show that this procedure underestimates the p-values and is very conservative. We also observe that $\mu(\widehat{F}_n)$ provides an accurate estimate of μ under RSS while $\sigma^2(\widehat{F}_n)$ does not provide an accurate estimate of the variance of the numerator. This may explain the poor performance of the Z test. In the remainder of the article, we did not consider this test. Another competitor for the bootstrap tests developed in this paper is to simply evaluate $T_2(X)$ as defined in Equation (4.2) and compare it to t-distribution cut-off values. We refer to this method as AT and compare it with the other methods developed in this section. For this method, one could also compare the values of $T_2(X)$ in (4.2) to the standard normal cut-off values. However, this leads to slightly elevated type I error rates. Our simulation studies show that AT (especially for small sample sizes) is a conservative for the normal and liberal for the exponential distributions.

It is important to study the proposed resampling under imperfect ranking. Following Dell and Clutter (1972), to produce imperfect URSS/RSS samples, let $X_{[i]j}$ and $X_{(i)j}$

Table 2. Observed α-levels for T_i, $i = 1, 2, 3$, for testing $H_0 : \mu = 0$, under imperfect ranking in the Normal distribution.

	D	AT	BTR1			BTR2		
			T_1	T_2	T_3	T_1	T_2	T_3
$\sigma_\epsilon = 0.5$	D_1	0.053	0.095	0.043	0.087	0.082	0.046	0.072
	D_2	0.067	0.169	0.059	0.153	0.160	0.061	0.149
	D_3	0.063	0.156	0.059	0.147	0.139	0.057	0.126
	D_4	0.069	0.183	0.053	0.177	0.148	0.060	0.148
	D_5	0.061	0.172	0.054	0.179	0.162	0.058	0.167
$\sigma_\epsilon = 1$	D_1	0.063	0.104	0.046	0.093	0.123	0.053	0.112
	D_2	0.067	0.153	0.063	0.139	0.185	0.063	0.170
	D_3	0.062	0.152	0.053	0.142	0.180	0.056	0.167
	D_4	0.073	0.181	0.059	0.173	0.190	0.063	0.180
	D_5	0.060	0.169	0.053	0.165	0.195	0.053	0.194
$\sigma_\epsilon = 2$	D_1	0.056	0.103	0.038	0.082	0.160	0.043	0.141
	D_2	0.060	0.155	0.053	0.141	0.215	0.054	0.208
	D_3	0.058	0.153	0.051	0.138	0.211	0.053	0.198
	D_4	0.072	0.181	0.057	0.167	0.218	0.061	0.206
	D_5	0.060	0.166	0.052	0.152	0.221	0.054	0.213

denote the judgment and true order statistics, respectively. Suppose

$$X_{[i]j} = X_{(i)j} + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma_\epsilon),$$

where $X_{(i)j}$ and ϵ_{ij} are independent. Using imperfect URSS with $\sigma_\epsilon = 0.5$, 1 and 2, we report the observed significance levels in Tables 2, 3, 4. For a normal distribution, the choices of σ_ϵ result in the correlation coefficients of 0.81, 0.71 and 0.45 between the ranking variable and the variable of interest, respectively. As compared with Table 1, the proposed method seems to be robust with respect to imperfect ranking.

Tables 5, 6, 7 display the estimated power values under location shift $H_1 : \mu = \mu_0 + \delta$ with $\delta \neq 0$. We used 95% percentile bootstrap confidence intervals for μ, using T_i, $i = 1, 2, 3$, under BTR1 and BTR2 to obtain the power of the test statistics at $\alpha = 0.05$. The entries of these tables are the proportion of times, while the bootstrap confidence intervals do not cover zero. We observe that the tests result in high powers, and considering the observed α levels for these tests, BTR2-T2 can be nominated to conduct appropriate tests. The results of other simulation studies (not presented here) show similar behavior for other values of k such as $k \in \{2, 3, 10\}$ as well as different sample sizes.

5. APPLICATION

To examine the proposed methods on real data, we consider the yield data (dray content g/m^2) from Swedish Spring barley. The data set includes 1278 observations from trials performed during the years 1998–2006, with a total of 59 varieties. Descriptive statistics on yields are presented in Table 8. The trials are performed in southern and central parts of Sweden and presented every year in summary tables on the Internet (*www.ffe.slu.se*).

Table 3. Observed α-levels for $T_i,\ i=1,2,3,$ for testing $H_0:\mu=1$, under imperfect ranking for the Exponential distribution.

	D	AT	BTR1 T_1	BTR1 T_2	BTR1 T_3	BTR2 T_1	BTR2 T_2	BTR2 T_3
$\sigma_\epsilon=0.5$	D_1	0.042	0.089	0.028	0.101	0.057	0.042	0.085
	D_2	0.053	0.123	0.052	0.121	0.110	0.058	0.114
	D_3	0.056	0.165	0.051	0.194	0.142	0.055	0.185
	D_4	0.065	0.206	0.052	0.234	0.147	0.065	0.193
	D_5	0.050	0.170	0.044	0.194	0.145	0.050	0.177
$\sigma_\epsilon=1$	D_1	0.051	0.094	0.037	0.092	0.108	0.047	0.116
	D_2	0.059	0.138	0.056	0.134	0.168	0.063	0.169
	D_3	0.057	0.160	0.050	0.174	0.182	0.052	0.195
	D_4	0.073	0.195	0.059	0.199	0.191	0.069	0.207
	D_5	0.059	0.174	0.052	0.174	0.199	0.062	0.212
$\sigma_\epsilon=2$	D_1	0.059	0.104	0.044	0.093	0.154	0.051	0.146
	D_2	0.063	0.161	0.056	0.148	0.216	0.059	0.211
	D_3	0.062	0.159	0.054	0.151	0.215	0.057	0.215
	D_4	0.074	0.184	0.059	0.172	0.218	0.068	0.211
	D_5	0.070	0.178	0.060	0.165	0.230	0.063	0.223

Table 4. Observed α-levels for $T_i,\ i=1,2,3,$ for testing $H_0:\mu=1$, under imperfect ranking for the Logistic distribution.

	D	AT	BTR1 T_1	BTR1 T_2	BTR1 T_3	BTR2 T_1	BTR2 T_2	BTR2 T_3
$\sigma_\epsilon=0.5$	D_1	0.061	0.098	0.043	0.092	0.065	0.051	0.064
	D_2	0.067	0.170	0.061	0.157	0.141	0.067	0.137
	D_3	0.063	0.173	0.054	0.166	0.1304	0.056	0.128
	D_4	0.064	0.187	0.048	0.179	0.126	0.056	0.130
	D_5	0.073	0.177	0.059	0.190	0.144	0.069	0.161
$\sigma_\epsilon=1$	D_1	0.065	0.104	0.048	0.098	0.094	0.052	0.090
	D_2	0.071	0.164	0.065	0.159	0.163	0.065	0.159
	D_3	0.068	0.177	0.058	0.167	0.164	0.061	0.161
	D_4	0.062	0.179	0.052	0.173	0.149	0.055	0.149
	D_5	0.059	0.182	0.054	0.184	0.173	0.062	0.184
$\sigma_\epsilon=2$	D_1	0.063	0.109	0.047	0.093	0.128	0.052	0.121
	D_2	0.073	0.169	0.066	0.157	0.205	0.066	0.197
	D_3	0.066	0.170	0.059	0.158	0.199	0.061	0.190
	D_4	0.071	0.173	0.056	0.162	0.179	0.057	0.175
	D_5	0.062	0.170	0.055	0.167	0.199	0.058	0.197

Sweden is divided into seven agricultural regions (production areas) and the trials are performed randomly inside the regions. Due to several factors, including barley varieties and soil, this data are right-skewed. Forkman, Amiri, and von Rosen (2012) evaluated the importance of the region used in the Swedish variety trail program. It is of interest to study the proposed method with these data. We consider the yield data as a population and produce

Table 5. Power comparison for T_i, $i = 1, 2, 3$, and parametric bootstrap (PB) under location shift $H_1 : \mu = \delta$, for the Normal distribution.

			BTR1			BTR2			PB		
	D	AT	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
$\delta = 0.1$	D_1	0.146	0.247	0.104	0.226	0.137	0.124	0.128	0.134	0.126	0.134
	D_2	0.150	0.274	0.102	0.253	0.144	0.126	0.136	0.121	0.116	0.126
	D_3	0.162	0.278	0.121	0.259	0.168	0.145	0.153	0.141	0.120	0.131
	D_4	0.152	0.294	0.026	0.300	0.115	0.103	0.119	0.110	0.096	0.102
	D_5	0.146	0.260	0.082	0.269	0.128	0.122	0.150	0.123	0.115	0.119
$\delta = 0.2$	D_1	0.406	0.564	0.316	0.539	0.406	0.356	0.368	0.400	0.370	0.395
	D_2	0.363	0.544	0.291	0.493	0.383	0.329	0.351	0.336	0.316	0.348
	D_3	0.398	0.562	0.311	0.541	0.414	0.361	0.386	0.372	0.326	0.356
	D_4	0.333	0.501	0.088	0.501	0.278	0.251	0.275	0.281	0.233	0.251
	D_5	0.364	0.523	0.224	0.522	0.342	0.303	0.367	0.329	0.288	0.325
$\delta = 0.3$	D_1	0.709	0.820	0.615	0.794	0.714	0.653	0.671	0.708	0.671	0.706
	D_2	0.650	0.812	0.547	0.771	0.692	0.601	0.639	0.633	0.585	0.648
	D_3	0.667	0.812	0.577	0.788	0.681	0.623	0.660	0.654	0.589	0.631
	D_4	0.549	0.734	0.194	0.7403	0.491	0.446	0.487	0.525	0.436	0.471
	D_5	0.605	0.758	0.431	0.758	0.597	0.540	0.593	0.598	0.529	0.574

Table 6. Power comparison for T_i, $i = 1, 2, 3$, and parametric bootstrap (PB) under location shift $H_1 : \mu = 1 + \delta$ for the Exponential distribution.

			BTR1			BTR2			PB		
	D	AT	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
$\delta = 0.1$	D_1	0.221	0.317	0.125	0.224	0.243	0.152	0.150	0.208	0.111	0.106
	D_2	0.224	0.291	0.102	0.146	0.212	0.123	0.078	0.218	0.107	0.110
	D_3	0.255	0.394	0.153	0.338	0.299	0.193	0.243	0.203	0.090	0.099
	D_4	0.281	0.446	0.069	0.427	0.280	0.191	0.236	0.194	0.075	0.077
	D_5	0.233	0.356	0.104	0.271	0.247	0.149	0.182	0.218	0.090	0.096
$\delta = 0.2$	D_1	0.439	0.541	0.261	0.392	0.461	0.305	0.301	0.433	0.234	0.241
	D_2	0.427	0.516	0.235	0.296	0.407	0.269	0.200	0.436	0.253	0.263
	D_3	0.417	0.568	0.291	0.499	0.477	0.339	0.391	0.393	0.190	0.218
	D_4	0.401	0.597	0.127	0.530	0.420	0.284	0.342	0.350	0.136	0.137
	D_5	0.375	0.525	0.192	0.393	0.398	0.247	0.280	0.381	0.170	0.187
$\delta = 0.3$	D_1	0.639	0.744	0.423	0.583	0.663	0.481	0.471	0.676	0.422	0.429
	D_2	0.659	0.724	0.402	0.488	0.647	0.469	0.376	0.696	0.448	0.487
	D_3	0.576	0.736	0.418	0.660	0.658	0.476	0.555	0.594	0.314	0.374
	D_4	0.533	0.716	0.197	0.643	0.559	0.400	0.461	0.496	0.215	0.255
	D_5	0.557	0.691	0.328	0.561	0.585	0.415	0.439	0.582	0.324	0.342

URSS/RSS with replacement samples from this data using the balanced RSS design D_1 as well as the URSS designs D_i, $i = 1, \ldots, 4$, as explained earlier in Section 4. We then carried out the proposed methods using these designs on our data set. Table 9 shows the observed α-levels for testing $H_0 : \mu = 430.4$. Here again, we observe that BTR2-T2 per-

Table 7. Power comparison for T_i, $i = 1, 2, 3$, and parametric bootstrap (PB) under location shift $H_1 : \mu = 1 + \delta$ for the Logistic distribution.

			BTR1			BTR2			PB		
	D	AT	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
$\delta = 0.1$	D_1	0.104	0.177	0.061	0.170	0.090	0.088	0.094	0.068	0.085	0.086
	D_2	0.087	0.190	0.048	0.182	0.084	0.070	0.091	0.064	0.065	0.069
	D_3	0.121	0.212	0.079	0.205	0.109	0.107	0.110	0.072	0.088	0.087
	D_4	0.117	0.239	0.017	0.255	0.076	0.077	0.095	0.070	0.068	0.069
	D_5	0.094	0.206	0.041	0.237	0.080	0.069	0.113	0.066	0.059	0.065
$\delta = 0.2$	D_1	0.171	0.284	0.121	0.261	0.163	0.148	0.157	0.145	0.1547	0.154
	D_2	0.158	0.304	0.107	0.283	0.170	0.136	0.168	0.143	0.132	0.142
	D_3	0.198	0.306	0.139	0.296	0.194	0.177	0.193	0.137	0.138	0.147
	D_4	0.176	0.316	0.030	0.329	0.135	0.121	0.148	0.124	0.110	0.117
	D_5	0.171	0.307	0.081	0.324	0.151	0.130	0.195	0.120	0.129	0.139
$\delta = 0.3$	D_1	0.292	0.427	0.205	0.407	0.286	0.249	0.268	0.274	0.262	0.275
	D_2	0.279	0.445	0.190	0.417	0.298	0.234	0.281	0.222	0.240	0.254
	D_3	0.307	0.457	0.241	0.440	0.330	0.284	0.315	0.266	0.238	0.271
	D_4	0.271	0.424	0.053	0.441	0.219	0.198	0.233	0.216	0.175	0.195
	D_5	0.264	0.419	0.151	0.435	0.256	0.210	0.296	0.216	0.205	0.230

Table 8. Summary statistics for the values of Spring barley (dry content g/m^2).

Min.	Q1.	Median	Mean	Q3.	Max.
145.8	352.6	408.4	430.4	492.8	810.1

Table 9. Observed α-levels for T_i, $i = 1, 2, 3$, for testing $H_0 : \mu = 430.4$ for the Spring barley.

		BTR1			BTR2		
D	AT	T_1	T_2	T_3	T_1	T_2	T_3
D_1	0.056	0.093	0.039	0.099	0.044	0.050	0.047
D_2	0.061	0.152	0.054	0.148	0.103	0.052	0.104
D_3	0.070	0.160	0.063	0.170	0.093	0.063	0.110
D_4	0.091	0.218	0.072	0.238	0.107	0.078	0.136
D_5	0.061	0.167	0.056	0.187	0.108	0.057	0.134

forms better than the other methods in terms of maintaining its α-level for different URSS designs.

6. CONCLUDING REMARKS

While CBS (2004) and Modarres, Hui, and Zhang (2006) provide resampling methods for RSS, they do not offer algorithms for URSS. This article explores two methods of resampling an unbalanced ranked set sample. We discuss resampling a URSS by transforming it to a balanced RSS and extending the existing algorithms to obtain resamples from the resulting RSS. We describe the two algorithms and investigate their properties. A simulation study is conducted to compare the two methods against parametric bootstrap. The results show that test statistic T2 under algorithm BTR2 maintains the significance level and provides power values that are comparable with the power of the parametric bootstrap test, which uses more information.

APPENDIX

A.1. PROOF OF PROPOSITION 1

Applying the Glivenko–Cantelli theorem to $\mathcal{X}_r$, we can show that $\sup_{t\in\mathbb{R}}\{\widehat{F}_{(r)}(t) - F_{(r)}(t)\} \longrightarrow 0$ a.s. $\forall t$, $m_r \longrightarrow \infty$. Since $q_{m_r} - q_r$ and $q_{m_r}\widehat{F}_{(r)}(t) - q_r F_{(r)}(t)$ converge to zero a.s. as m_r approaches ∞, $\forall t$, one can show that

$$\begin{aligned}
\left|\widehat{F}_{q_n}(t) - F_q(t)\right| &= \left|\sum_{r=1}^{k} q_{m_r}\widehat{F}_{(r)}(t) - \sum_{r=1}^{k} q_r F_{(r)}(t)\right| \\
&\leq \sum_{r=1}^{k}\left|q_{m_r}\widehat{F}_{(r)}(t) - q_r F_{(r)}(t)\right|.
\end{aligned}$$

Therefore, we establish $|\widehat{F}_{q_n}(t) - F_q(t)| \longrightarrow 0$, a.s. $\forall t$.

A.2. PROOF OF PROPOSITION 2

To show the result we use the following lemma, the proof of which can be found in Bickel and Freedman (1981) and Belyaev (1995).

Lemma A.1. *Let $G_n, G \in \Gamma_p$. Then $d_p(G_n, G)$ converges to zero if and only if G_n converges to G in distribution and $\int |x|^p\, dG_n(x)$ converges to $\int |x|^p\, dG(x)$ as n goes to ∞.*

Proposition 1 proves the first condition of Lemma A.1 while the second condition follows from

$$\begin{aligned}
\int t^2\, d\widehat{F}_{(r)}(t) &= \sum q_{m_r} \sum_{j=1}^{m_r} \frac{X_{(r)j}^2}{m_r} \\
&\rightarrow \sum q_r \int t^2\, dF_{(r)}(t) \\
&= \int t^2\, d\sum q_r F_{(r)}(t) \\
&= \int t^2\, dF_q(t).
\end{aligned}$$

Therefore, $d_2(\widehat{F}_{q_n}, F_q) \longrightarrow 0$ a.s.

A.3. VALIDATING THE BOOTSTRAP TRANSFORMATION

To validate the use of the bootstrap transformation RSS $=$ BTR(URSS, N) to transform a URSS to a balanced RSS we use the following lemma which can be easily proved using the Glivenko-Cantelli theorem (see Bickel and Freedman 1981).

Lemma A.2. *Suppose $\mathcal{X}_r^\diamond = \{X_{(r)1}^\diamond, \ldots, X_{(r)N}^\diamond\}$ are i.i.d. samples from $\widehat{F}_{(r)}(x)$ and let $\widehat{F}_{(r)}^\diamond(t)$ be the EDF of $X_{(r)}^\diamond$. Then $\|\widehat{F}_{(r)}^\diamond - \widehat{F}_{(r)}\|_\infty = \sup_{t\in\mathbb{R}} |\widehat{F}_{(r)}^\diamond(t) - \widehat{F}_{(r)}(t)|$ converges to zero a.s. for all $r = 1, \ldots, k$, as N approaches ∞.*

We pursue with the following result.

Proposition A.1. *If $F \in \Gamma_2$ and $\widehat{F}_N^\diamond$ is the EDF defined in (2.6), then $d_2(\widehat{F}_N^\diamond, F)$ converges to zero a.s. as N approaches ∞.*

Proof: Lemma A.2 shows $\widehat{F}_{(r)}^\diamond(t)$ converges to $\widehat{F}_{(r)}(t)$ $\forall t$ a.s. as m_r approaches ∞. It follows that

$$\begin{aligned}
\sup_{t\in\mathbb{R}}\left\{\widehat{F}_{(r)}^\diamond(t) - F_{(r)}(t)\right\} &= \sup_{t\in\mathbb{R}}\left\{\widehat{F}_{(r)}^\diamond(t) - \widehat{F}_{(r)}(t) + \widehat{F}_{(r)}(t) - F_{(r)}(t)\right\} \\
&= \sup_{t\in\mathbb{R}}\left\{\widehat{F}_{(r)}^\diamond(t) - \widehat{F}_{(r)}(t)\right\} + \sup_{t\in\mathbb{R}}\left\{\widehat{F}_{(r)}(t) - F_{(r)}(t)\right\} \\
&= A_1 + A_2.
\end{aligned}$$

Since A_1 converges to zero (a.s.) under Lemma A.2 and A_2 converges to zero (a.s.) under Glivenko–Cantelli theorem, we have $\sup_{t\in\mathbb{R}}\{\widehat{F}_N^\diamond(t) - F(t)\} \longrightarrow 0$ as $m_r \longrightarrow \infty$ (a.s.). The second condition is satisfied because

$$\begin{aligned}
\int t^2\, dF^\diamond &= \sum_{r=1}^{k} \frac{1}{k} \sum_{j=1}^{m_r} \frac{x_{(r)j}^{\diamond 2}}{N} \\
&\xrightarrow{\text{a.s}} \sum_{r=1}^{k} \frac{1}{k} \sum_{j=1}^{m_r} \frac{x_{(r)j}^{2}}{m_r} \\
&\xrightarrow{\text{a.s}} \sum_{r=1}^{k} \frac{1}{k} \int t^2\, dF_{(r)} \\
&= \int t^2\, d\sum_{r=1}^{k} \frac{1}{k} F_{(r)} = \int t^2\, dF.
\end{aligned}$$

□

The next corollaries follow from Propositions 2 and A.1, respectively.

Corollary A.1. *If $F \in \Gamma_2$ and $\widehat{F}_n$ is the EDF in (2.3), then $d_2(\widehat{F}_n, F)$ converges to zero (a.s.) as n approaches ∞.*

Corollary A.2. *If $F \in \Gamma_2$ and $\widehat{F}_N^\diamond$ is the EDF in (2.6), then $\bar{X}^{*\diamond}$ converges to μ (a.s.) as N approaches ∞, where $\bar{X}^{*\diamond} = \frac{1}{Nk}\sum_{i=1}^{Nk} X_i^{*\diamond}$ and $X_i^{*\diamond} \overset{\text{i.i.d.}}{\sim} F_N^\diamond(\cdot)$.*

ACKNOWLEDGEMENTS

The research of S. Amiri was supported by a Grant No. NR66853W from NSF. Part of this work was done while S. Amiri was a visiting Ph.D. student at the University of Manitoba and he gratefully acknowledges the additional financial support of M. Jafari Jozani (NSERC Grant). M. Jafari Jozani gratefully acknowledges the research support of the NSERC of Canada. The research of R. Modarres was supported in part by the National Institute of Health, under the Grant No. 1R01GM092963-01A1. The authors would also like to thank Silvelyn Zwanzig and three anonymous reviewers for their constructive comments.

[Received June 2012. Accepted July 2013. Published Online August 2013.]

REFERENCES

Al-Saleh, M. F., and Zheng, G. (2002), "Estimation of Bivariate Characteristics Using Ranked Set Sampling," *Australian & New Zealand Journal of Statistics*, 44, 221–232.

Belyaev, Y. K. (1995), "Bootstrap, Resampling and Mallows Metric," Lecture Notes, No. 1, Institute of Mathematical Statistics, Umeå University.

Bickel, P. J., and Freedman, D. A. (1981), "Some Asymptotic Theory for the Bootstrap," *The Annals of Statistics*, 9, 1196–1217.

Bohn, L. L., and Wolfe, D. A. (1992), "Nonparametric Two-Sample Procedures for Ranked Set Samples Data," *Journal of the American Statistical Association*, 87, 552–561.

Boos, D. D., Janssen, P., and Veraverbeke, N. (1989), "Resampling From Centered Data in the Two-Sample Problem," *Journal of Statistical Planning and Inference*, 21, 327–345.

Chen, Z. (2001), "Nonparametric Inferences Based on General Unbalanced Ranked-Set Samples," *Journal of Nonparametric Statistics*, 13, 291–310.

Chen, Z., Bai, Z., and Sinha, B. K. (2004), *Ranked Set Sampling: Theory and Applications*, New York: Springer.

Davison, A. C., and Hinkley, D. V. (1997), *Bootstrap Methods and Their Application*, Cambridge: Cambridge University Press.

Dell, T. R., and Clutter, J. L. (1972), "Ranked Set Sampling Theory With Order Statistics Background," *Biometrics*, 28, 545–555.

Efron, B., and Tibshirani, R. (1993), *An Introduction to Bootstrap*, New York: Chapman & Hall.

Fligner, M. A., and MacEachern, S. N. (2006), "Nonparametric Two-Sample Methods for Ranked Set Sample Data," *Journal of the American Statistical Association*, 101, 1107–1118.

Forkman, J., Amiri, S., and von Rosen, D. (2012), "Effect of Region on the Uncertainty in Crop Variety Trial Programs With a Reduced Number of Trials," *Euphytica*, 186, 489–500.

Frey, J. (2007), "Distribution-Free Statistical Intervals Via Ranked Set Sampling," *Canadian Journal of Statistics*, 35, 585–596.

Hall, P. (1992), *The Bootstrap and Edgeworth Expansion*, New York: Springer.

Hall, P., and Wilson, S. R. (1991), "Two Guidelines for Bootstrap Hypothesis Testing," *Biometrics*, 47, 757–762.

Kvam, P. H., and Samaniego, F. J. (1994), "Nonparametric Maximum Likelihood Estimation Based on Ranked Set Samples," *Journal of the American Statistical Association*, 89, 526–537.

McIntyre, G. A. (1952), "A Method for Unbiased Selective Sampling, Using Ranked Sets," *Australian Journal of Agriculture Research*, 3, 385–390.

Modarres, R., Hui, T. P., and Zhang, G. (2006), "Resampling Methods for Ranked Set Samples," *Computational Statistics & Data Analysis*, 51, 1039–1050.

Samawi, H. M., and Al-Sagheer, O. A. M. (2001), "On the Estimation of the Distribution Function Using Extreme and Median Ranked Set Sampling," *Biometrical Journal*, 43, 357–373.

Shao, J., and Tu, D. (1996), "*The Jackknife and Bootstrap*," New York: Springer.

Wolfe, D. A. (2004), "Ranked Set Sampling: An Approach to More Efficient Data Collection," *Statistical Science*, 19, 636–643.