Ann Inst Stat Math (2017) 69:1029–1057
DOI 10.1007/s10463-016-0572-y

Statistical inference with empty strata in judgment post stratified samples

Omer Ozturk[1]

Received: 21 July 2015 / Revised: 20 May 2016 / Published online: 11 July 2016

Abstract This article develops estimators for certain population characteristics using a judgment post stratified (JPS) sample. The paper first constructs a conditional JPS sample with a reduced set size K by conditioning on the ranks of the measured observations of the original JPS sample of set size $H \geq K$. The paper shows that the estimators of the population mean, median and distribution function based on this conditional JPS sample are consistent and have limiting normal distributions. It is shown that the proposed estimators, unlike the ratio and regression estimators, where they require a strong linearity assumption, only need a monotonic relationship between the response and auxiliary variable. For moderate sample sizes, the paper provides a bootstrap distribution to draw statistical inference. A small-scale simulation study shows that the proposed estimators based on a reduced set JPS sample perform better than the corresponding estimators based on a regular JPS sample.

Keywords Reduced set size · Empty cell · Ranking error · Ranked set sample · Stratified sample

1 Introduction

In setting, where an abundance of auxiliary information is available, information content of a simple random sample can be increased by inducing some additional structures in the data. One can start with a simple random sample and use available auxiliary information from additional sample units through a ranking process to determine

✉ Omer Ozturk
omer@stat.osu.edu

[1] Department of Statistics, The Ohio State University, 1958 Neil Avenue, Columbus, OH 43210, USA

the relative position of each measured unit in a small set. This position information provides a mechanism to create a judgment post stratified (JPS) sample by putting homogeneous observations together in the same group. A JPS sample in this setting can be considered a post stratified sample since grouping is performed after a simple random sample has been collected. The increased efficiency of a JPS sample is then anticipated from the general theory of the stratified sample in survey sampling designs.

Let X_i, $i = 1, \ldots, n$, be a simple random sample (SRS) from a distribution F. For the construction of a JPS sample from this SRS, we select additional $H - 1$ units for each measured value X_i to construct a set of size H. Units in this set are ranked without measurement with no additional cost to determine the rank of X_i, R_i. The ranking process is performed without measurement, such as using visual inspection, auxiliary variables, etc. If ranking is performed based on a visual inspection, the ranker should be blinded to the actual value of X_i to avoid possible biases in the ranking process. Throughout this paper, the set of observations X_i, $i = 1, \ldots, n$, with the judgment rank h is called judgment class h or stratum h, $h = 1 \ldots, H$. The ranking scheme is called consistent if the equality

$$F(y) = \frac{1}{H} \sum_{h=1}^{H} F_{[h:H]}(y)$$

holds, where $F_{[h:H]}(y)$ is the CDF of the h-th judgment order statistics. If the ranking procedure is perfect, assigning the true ranks to the units, we replace the judgment order statistics with the usual order statistics in a simple random sample of size H and write $F_{[h:H]}(y) = F_{(h:H)}(y)$, $h = 1, \ldots, H$. If the ranking procedure is random, assigning the ranks at a completely random fashion, the judgment order statistics become independent identically distributed random variables and we write $F_{[h:H]}(y) = F(y)$, $h = 1, \ldots, H$.

JPS sampling has a wide range of applications. A particular example can be found in allometry study in MacEachern et al. (2004) and Wang et al. (2008). The main goal in these studies is to estimate the mean of log-adjusted brain weight. A JPS sample first selects a simple random sample of n mammals and measures the brain weights. Each one of the mammals in this SRS is matched with $H - 1$ randomly selected mammals to form a set of size H. The rank of the measured brain weight in this set is then estimated by ranking the H mammals based on their intelligence level. Another example is given in the estimation of the percentage-area-coverage by a chemical spray on apple tree leaves in Ozturk (2013). In this case, the percentage-area-coverage is computed with a chemical analysis on the selected simple random sample of n leaves. Each one of the leaves in this SRS is again matched with $H - 1$ randomly selected additional leaves to form a set of size H. The rank of each measured leaf in its set is then estimated under ultra-violet lights to construct the JPS sample. The third example for a special version of JPS is given in clinical trials to compare the efficacy of three treatments used for the reversal of anesthesia (Du and MacEachern 2008). In this study, the response variable is the elapsed time from the treatment administration to completion of the anesthesia reversal. In this case, the ranks of the subjects in a set of size H are constructed through the nonlinear covariate variable depth of neuromuscular block at the time of reversal.

In these examples, the regression estimator and covariance analysis may not be used since there may not exist a standard measurement for predictor (or auxiliary) variables. For example, the intelligence level of mammals and light intensity may not be converted to a numerical measure very easily. JPS sampling, unlike a regression estimator, does not require a standard measurement for the predictors. It only requires the relative position of the measured unit in a set which can be obtained from visual inspection or some other form of ranking process. JPS sampling requires some sort of monotonic relationship between the ranking variable and response, which is much weaker than the strong linearity assumption of regression and ratio estimators.

A JPS sample consists of a simple random sample, $SRS = (X_1, \ldots, X_n)$, and a rank vector, $\boldsymbol{R} = (R_1, \ldots, R_n)$, associated with X_i, $i = 1, \ldots, n$, in SRS. Since the rank vector $\boldsymbol{R}$ is loosely related to SRS, it can be ignored, and a JPS sample can be treated as a SRS sample. A JPS sample can be considered, conditionally on the observed rank vector $\boldsymbol{R}$, as an unbalanced ranked set sample, where the ranks are strongly attached to the measured observations. A ranked set sample (RSS) of size n contains n independent judgment order statistics selected from n different sets. The construction of a RSS sample requires selecting n sets, each of size H. Units in each set are ranked without measurement. Again the ranking process, as in JPS sampling, is performed by using a visual inspection or auxiliary variable without requiring a full measurement. For $h = 1, \ldots, H$, the judgment ranked unit h is selected for measurement in n_h sets so that $\sum_{h=1}^{H} n_h = n$. The measured random variables, $X_{[h]j}$, $j = 1, \ldots, n_h$; $h = 1, \ldots, H$, are then called an unbalanced RSS sample. Up-to-date references for RSS sampling designs can be found in a recent review paper in Wolfe (2012) and in a book by (Hollander et al. 2014, Chapter 15).

In recent years, research in JPS sampling drew considerable attention in literature. Wang et al. (2006) developed a class of estimators for the population mean based on the concomitant of multivariate order statistics. Wang et al. (2008) used stochastic ordering constraint to construct estimators for the population mean. Frey and Ozturk (2011) relaxed the assumption of the stochastic ordering constraint. They showed that under a consistent ranking scheme, the judgment class cumulative distribution functions (CDF) can be no more extreme than the CDF of true order statistics. This constraint is weaker than the one used in the stochastic ordering of the judgment order statistics. In a sequel paper, Frey (2012) combined the constraint in Frey and Ozturk (2011) with the stochastic constraint of order statistics to produce more efficient estimators for the population mean and variance. Frey and Feeman (2012, 2013) developed the optimal estimators within a class of unbiased estimators for the population mean and variance. Chen et al. (2014), Stokes et al. (2007) and Ozturk (2013) combined ranking information from different sources to develop statistical inference for population characteristics under a JPS sample.

The research in RSS and JPS indicates that the procedures based on an RSS sample yield higher efficiencies than the same procedures based on a JPS sample due to the strong data structure in an RSS sample. Even though a JPS sample is less efficient than a RSS sample, it still yields higher efficiency than a SRS sample. The loss of efficiency in a JPS sample can be attributed to the empty strata and unequal judgment class sample sizes. Let N_h be the number of observations having ranks in the judgment class h, $h = 1, \ldots, H$. The sample size vector $\boldsymbol{N} = (N_1, \ldots, N_H)$ has a multinomial distribution with parameters n and $(1/H, \ldots, 1/H)$. For small sample size n, it is highly possible that some of the judgment class sample sizes would be zero or very close to zero, creating a highly unbalanced JPS sample. The empty judgment classes not only reduce efficiency, but may also introduce a bias in the estimator. Dastbaravarde et al. (2016) constructed classes of unbiased estimators for population moment and variance by putting weight to each judgment class to account for the empty strata in a JPS sample. Ozturk (2014) constructed a similar class of estimators for the p-th quantile of a distribution F. Ozturk (2015) developed distribution free two-sample methods based on JPS samples. Wang et al. (2012) used the stochastic ordering constraint among judgment class CDFs to fill-in the empty judgment classes to estimate the CDF of the underlying distribution.

In this paper, to reduce the impact of the empty judgment classes in a JPS sample, we create another JPS sample with a reduced set size $K \leq H$ conditionally on the rank vector $\boldsymbol{R}$. In the proposed JPS sample, the ranks of the measured observations X_i, $i = 1, \ldots, n$, are recomputed in a set of size K by conditioning on the rank R_i in the original JPS sample of the set size H. Use of smaller set size K makes sure that the conditional JPS sample does not have any empty judgment class and hence improves the efficiency. Section 2 provides a detailed description for the construction of the JPS sample with a reduced set size K. This section also constructs a nonparametric estimator for the parameter $\theta_g = E(g(X - b))$ for a wide range of choices of real-valued function g. We show that the estimator is consistent and has limiting normal distribution. Section 3 considers a special case of the function g to draw inference for the p-th quantile of F. Section 4 discusses the selection of the optimal reduced set size K to improve the efficiency of the proposed estimators. Section 5 investigates the bootstrap variance estimate of estimators and constructs a percentile confidence interval for θ_g. Section 6 provides some empirical evidence on the finite sample behavior of the estimators. Section 7 applies the proposed estimator to 2012 United States Department of Agriculture (USDA) census data to estimate the mean corn production at county levels. Section 8 provides concluding remarks.

2 JPS sample with reduced set size

We consider a judgment post-stratified sample, (X_i, R_i), $i = 1, \ldots, n$, where n could be small and the set size H could be relatively large to produce some empty judgment classes. We note that the relative size of H depends on its potential to generate empty strata. The set sizes $H = 3$ and $H = 5$ could be relatively large for a sample size of $n = 5$, but may not be large for sample sizes $n = 50$ and $n = 100$. For relatively large H, sample size(s) of some judgment classes may be zero ($N_h = 0$) for some h. In this case, the estimators could be biased and very unstable. To minimize the impact of the empty judgment classes, we re-evaluate the rank of the measured unit X_i, conditionally on the original rank R_i, in a smaller set of size $K \leq H$ so that none of the sample sizes N_k, $k = 1, \ldots, K$, in the reduced JPS sample is zero.

Let

$$S_{i,H} = \{X_i, Y_1, \ldots, Y_{H-1}\}$$

be a set of size H in which X_i has rank R_i. We consider the probability distribution of the rank of X_i if we remove the $(H-K)$ unmeasured Y- observations at random from this set. Let $S_{i,K} = \{X_i, Y_{t_1}, \ldots, Y_{t_{K-1}}\}$ be the reduced set and $C_{K|R_i}$ be the rank of X_i in this reduced set. We assume that the ranking process of the units in the reduced set is not affected by the absence of some units in the original set. This assumption basically states that ranking process is consistent for all reduced set sizes $K \leq H$, and it holds for ranking processes based on concomitant variables and perceived size ranking model of Dell and Clutter (1972). Under a consistent ranking model, we compute

$$a_{k:K|h} = P(C_{K|R_i=h} = k) = \frac{\binom{h-1}{k-1}\binom{H-h}{K-k}}{\binom{H-1}{K-1}}, \quad k = 1, \ldots, K, \tag{1}$$

where $a_{k:K|h}$ is the probability that the unit with rank h $(R_i = h)$ in the set $S_{i,H}$ is assigned the rank k in the reduced set $S_{i,K}$.

Remark 1 For $K = H$, $P(C_{K|R_i=h} = k) = a_{k:H|h} = 1$ if $k = h$ and zero otherwise. In this case, $C_{K|R_i}$ is a degenerate random variable at h.

For any bounded real-valued function g, we define $W_{k,R_i,g}(b) = a_{k:K|R_i} g(X_i - b)$, $-\infty < b < \infty$. It is clear that $W_{k,R_i,g(b)}$ prorates the original measurement $X_i - b$ to the k-th strata in the reduced set. By using Eq. (1), under a consistent ranking scheme, one can establish

$$\begin{aligned}
E(W_{k,R_i,g}(b)) &= E(E(W_{k,R_i,g}(b)|R_i)) = E(a_{k:K|R_i})E\left\{(g(X_i - b)|R_i)\right\} \\
&= \sum_{h=1}^{H} E\left\{I(R_i = h)\right\} a_{k:K|h} E\{g(X_{[h:H]} - b)\} \\
&= \sum_{h=1}^{H} \frac{\binom{h-1}{k-1}\binom{H-h}{K-k}}{H\binom{H-1}{K-1}} Eg(X_{[h:H]} - b) \\
&= \frac{1}{K}\sum_{h=1}^{H} \frac{\binom{h-1}{k-1}\binom{H-h}{K-k}}{\binom{H}{K}} E\{g(X_{[h:H]} - b)\} \\
&= \frac{E\{g(X_{[k:K]} - b)\}}{K}.
\end{aligned} \tag{2}$$

The last equality in the above expression follows from the fact that one of the judgment rank in a set of size H must be assigned to a unit having a judgment rank k in the reduced set of size K. In a similar fashion, the expected value of $a_{k:K|R_i}$ reduces to $E(a_{k:K|R_i}) = 1/K$ under a consistent ranking scheme

$$E\{a_{k:K|R_i}\} = \sum_{h=1}^{H} E\{I(R_i = h)\} \frac{\binom{h-1}{k-1}\binom{H-h}{K-k}}{\binom{H-1}{K-1}}$$

$$= \sum_{h=1}^{H} \frac{\binom{h-1}{k-1}\binom{H-h}{K-k}}{H\binom{H-1}{K-1}} = 1/K. \tag{3}$$

We now propose a new estimator for the parameter $\theta_g(b) = Eg(X - b)$ for a wide range of choices of the real-valued function g. A natural estimator for $\theta_g(b)$ is given by

$$\hat{\theta}_{g,K}(b) = \frac{1}{d_n} \sum_{k=1}^{K} I_k \frac{\sum_{i=1}^{n} a_{k:K|R_i} g(X_i - b)}{\sum_{i=1}^{n} a_{k:K|R_i}}, \tag{4}$$

where $I_k = 1$ if $D_k = \sum_{i=1}^{n} a_{k:K|R_i} > 0$, and zero otherwise, and $d_n = \sum_{k=1}^{K} I_k$.

Theorem 1 *Let $g(X - b)$ be a bounded real-valued function on the real line and let b be a real number.*

1. *The estimator $\hat{\theta}_{g,K}(b)$ in Eq. (4) is a consistent estimator for $\theta_g(b) = Eg(X - b)$ under any consistent ranking scheme.*
2. *If $\left\{E[g(X_{[h:H]} - b)] + E[g(X_{[H+1-h:H]} - b)]\right\}/2 = E[g(X - b)]$ for all $h \leq H$, then $\hat{\theta}_{g,K}(b)$ is an unbiased estimator for $\theta_g(b) = E[g(X - b)]$.*

Even though the estimator is not unbiased for asymmetric distributions, the simulation study (not reported in the paper) under a wide range of simulation settings suggests that the estimator $\hat{\theta}_{g,K}(b)$ has a very small bias for moderately skewed distributions.

We now consider the limiting distribution of $\hat{\theta}_{g,K}(b)$. It is clear that this estimator is not a sum of independent observations since X_i is prorated to K different classes. For the asymptotic distribution, we first construct a $2K - 1$ dimensional mean vector of independent random variables and show that it converges to a multivariate normal distribution with certain covariance structure. We then use a continuous function to map this vector to our estimator $\hat{\theta}_{g,K}(b)$. Let $\bar{\boldsymbol{T}}_g^\top(b) = (\bar{T}_{g,1}(b), \ldots, \bar{T}_{g,K}(b), \bar{T}_{g,K+1}(b), \ldots, \bar{T}_{g,2K-1}(b))$, where

$$\bar{T}_{g,k}(b) = \begin{cases} \frac{1}{n}\sum_{i=1}^{n} a_{k:K|R_i} g(X_i - b) & 1 \leq k \leq K \\ \frac{1}{n}\sum_{i=1}^{n} a_{k-K:K|R_i} & K < k \leq 2K - 1. \end{cases} \tag{5}$$

It is clear that $\bar{\boldsymbol{T}}_g(b)$ is the sum (mean) of n independent random vectors, each of which is $2K - 1$ dimensional.

Lemma 1 *Let (X_i, R_i), $i = 1, \ldots, n$, be a JPS sample constructed under a consistent ranking scheme from a distribution F having mean μ and variance σ^2. For any bounded real-valued function g defined on the real line and $\bar{\boldsymbol{T}}_g$ in Eq. (5), as n increases, $\sqrt{n}\left\{\bar{\boldsymbol{T}}_g(b) - E(\bar{\boldsymbol{T}}_g(b))\right\}$ converges to a $2K-1$ dimensional multivariate normal distribution with a mean vector of zero and covariance matrix of $\boldsymbol{\Sigma}$,*

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{1,1} & \boldsymbol{\Sigma}_{1,2} \\ \boldsymbol{\Sigma}_{2,1} & \boldsymbol{\Sigma}_{2,2} \end{bmatrix},$$

where $\boldsymbol{\Sigma}_{1,1} = \left(\sigma_{r,s}\right)_{K \times K}$, $\boldsymbol{\Sigma}_{1,2} = \left(\tau_{r,s}\right)_{K \times (K-1)}$, $\boldsymbol{\Sigma}_{2,2} = \left(\gamma_{r,s}\right)_{(K-1) \times (K-1)}$,

$$\begin{aligned} \sigma_{r,s} &= \frac{1}{H} \sum_{h=1}^{H} a_{r:K|h} a_{s:K|h} E(g^2(X_{[h:H]} - b)) \\ &\quad - \frac{1}{K^2} E(g(X_{[r:K]} - b)) E(g(X_{[s:K]} - b)), \\ \tau_{r,s} &= \frac{1}{H} \sum_{h=1}^{H} a_{r:K|h} a_{s:K|h} E(g(X_{[h:H]} - b)) - \frac{1}{K^2} E(g(X_{[r:K]} - b)), \\ \gamma_{r,s} &= \frac{1}{H} \sum_{h=1}^{H} a_{r:K|h} a_{s:K|h} - \frac{1}{K^2}, \end{aligned}$$

and $E(\bar{\boldsymbol{T}}_g^{\top}(b)) = \frac{1}{K}\left(Eg(X_{[1:K]} - b), \ldots, Eg(X_{[K:K]} - b), 1, \ldots, 1\right)$.

Theorem 2 *Let (X_i, R_i), $i = 1, \ldots, n$, be a JPS sample constructed under a consistent ranking scheme from a distribution F having mean μ and variance σ^2. For any bounded real-valued function g defined on the real line and $\hat{\theta}_{g,K}(b)$ in Eq. (4), as n goes to infinity, the distribution of $\sqrt{n}\{\hat{\theta}_{g,K}(b) - \theta_g(b)\}$ converges to a normal distribution with mean zero and variance $\sigma^2_{\hat{\theta}_{g,K}}$,*

$$\begin{aligned} \sigma^2_{\hat{\theta}_{g,K}} &= \sigma_g^2 - 2 \sum_{r=1}^{K} (Eg(X_{[r:K]} - b) - Eg(X - b))^2 / K \\ &\quad + \sum_{r=1}^{K} \sum_{s=1}^{K} \frac{1}{H} \sum_{h=1}^{H} a_{r:K|h} a_{s:K|h} Eg(X_{[r:K]} - b) Eg(X_{[s:K]} - b) \\ &\quad - \{E(g(X - b))\}^2, \end{aligned}$$

where $\sigma_g^2 = \mathrm{var}(g(X - b))$ and $\theta_g(b) = E(g(X - b))$.

If $H = K$, $\sigma^2_{\hat{\theta}_{g,K}}$ reduces to the asymptotic variance of the ranked set sample estimator of θ_g

$$\sigma^2_{\hat{\theta}_{g,H}} = \sigma^2_{\hat{\theta}_{g,\mathrm{RSS}(H)}} = \sigma_g^2 - \sum_{r=1}^{H} (Eg(X_{[r:H]} - b) - Eg(X - b))^2 / H,$$

where $\sigma^2_{\hat{\theta}_{g,\mathrm{RSS}(H)}}$ is the asymptotic variance of $\hat{\theta}_g$ based on a ranked set sample of set size H.

Corollary 1 *For* $1 \le K \le H$, $\sigma^2_{\hat{\theta}_{g,K}} \le \sigma^2_{\hat{\theta}_{g,\mathrm{RSS}(K)}}$.

The Corollary 1 indicates that the asymptotic variance of $\hat{\theta}_{g,K}$ based on a JPS sample estimator of set size K is always smaller than the asymptotic variance of $\hat{\theta}_{g,\mathrm{RSS}(K)}$ based on a ranked set sample of set size K.

Corollary 2 *If* $g(x-b) = x$, *then* $\theta_g(b) = \mu$, $\hat{\theta}_{g:K}(b) = \hat{\mu}_K$ *and* $\sigma^2_{\hat{\mu}_K}$ *reduces to*

$$\begin{aligned}\sigma^2_{\hat{\mu}_K} &= \sigma^2 - 2\sum_{r=1}^{K}(\mu_{[r:K]}-\mu)^2/K + \sum_{r=1}^{K}\sum_{s=1}^{K}\frac{1}{H}\sum_{h=1}^{H} a_{r:K|h}a_{s:K|h}\mu_{[r:K]}\mu_{[s:K]} - \mu^2\\ &= \sigma^2_{\mathrm{RSS}(K)} - \left\{\sum_{r=1}^{K}\mu^2_{[r:K]}/K - \sum_{r=1}^{K}\sum_{s=1}^{K}\frac{1}{H}\sum_{h=1}^{H} a_{r:K|h}a_{s:K|h}\mu_{[r:K]}\mu_{[s:K]}\right\},\end{aligned}$$

where $\sigma^2_{\mathrm{RSS}(K)}$ *is the variance of the sample mean based on a ranked set sample of set size* K.

The asymptotic variance of the proposed estimator of a population mean is decomposed into two pieces. The first piece, $\sigma^2_{\mathrm{RSS}(K)}$, is the variance of the sample mean of a ranked set sample of set size K. The Corollary 1 indicates that the proposed estimator is asymptotically more efficient than an RSS sample mean estimator of set size K. Similar result also holds for the point estimator of a population distribution function, $F(t)$.

Corollary 3 *If* $g(x-b) = I(x \le t)$, *then* $\theta_{g,K}(b) = F(t)$, $\hat{\theta}_{g,K}(b) = \hat{F}_K(t)$ *and* $\sigma^2_{\hat{F}_K}$ *reduces to*

$$\begin{aligned}\sigma^2_{\hat{F}_K} &= F(t)\{1-F(t)\} - 2\sum_{r=1}^{K}(F_{[r:K]} - F(t))^2/K\\ &\quad + \sum_{r=1}^{K}\sum_{s=1}^{K}\frac{1}{H}\sum_{h=1}^{H} a_{r:K|h}a_{s:K|h}F_{[r:K]}(t)F_{[s:K]}(t) - F^2(t)\\ &= \mathrm{var}(\hat{F}_{\mathrm{RSS}(K)}(t)) - \left\{\frac{1}{K}\sum_{r=1}^{K}F^2_{[r:K]}(t)\right.\\ &\quad \left. - \sum_{r=1}^{K}\sum_{s=1}^{K}\frac{1}{H}\sum_{h=1}^{H} a_{r:K|h}a_{s:K|h}F_{[r:K]}(t)F_{[s:K]}(t)\right\}\\ &\le \mathrm{var}(\hat{F}_{\mathrm{RSS}(K)}(t)),\end{aligned}$$

where $\mathrm{var}(\hat{F}_{\mathrm{RSS}(K)}(t))$ *is the variance of the CDF estimator of a ranked set sample of set size* K.

3 Quantile inference

In this section, we consider developing inference for the p-th order quantile (η_p) of distribution F, $\eta_p = \inf\{x : F(x) \geq p\}$. We first consider testing $H_0 : \eta_p = \eta_0$ against $H_A : \eta_p \neq \eta_0$. It is also possible to construct a test for appropriate one-sided alternatives. Let $g(x-b) = I(X > b)$ and

$$S_n(\eta_p) = \sum_{i=1}^{n} \frac{1}{d_n} \sum_{k=1}^{K} \frac{I_k a_{k:K|R_i}}{\sum_{j=1}^{n} a_{k:K|R_j}} \left\{I(X_i > \eta_p) - p\right\}$$
$$= \sum_{i=1}^{n} w_{R_i}(\boldsymbol{R}) \left\{I(X_i \geq \eta_p) - p\right\}, \tag{6}$$

where

$$w_{R_i}(\boldsymbol{R}) = \frac{1}{d_n} \sum_{k=1}^{K} \frac{I_k a_{k:K|R_i}}{\sum_{j=1}^{n} a_{k:K|R_j}}.$$

It is clear that $S_n(\eta_p)$ is a non-increasing function of η_p; we then reject the null hypothesis H_0 in favor of the alternative hypothesis for the extreme values of $S_n(\eta_p)$. The exact null distribution of $S_n(\eta_0)$ is a linear combination of weighted binomial random variables given the rank vector $\boldsymbol{R} = \boldsymbol{r}$,

$$S_n(\eta_0)|\boldsymbol{R} = \boldsymbol{r} \overset{D}{=} \sum_{h=1} w_{[h]}(\boldsymbol{r}) Y_{[h]},$$

where $Y_{[h]} = \sum_{i=1}^{n} I(R_i = h)\{I(X_i > \eta_0) - p\}$ and $Y_{[h]}$ has a binomial distribution, conditionally on the given ranks $R_i = h$, $i = 1, \ldots, n$, with parameters $n_h = N_h$ and $1 - F_{[h:H]}(\eta_0)$. Under perfect and random ranking procedures, the exact null distribution of $S_n(\eta_0)$ is distribution free, but the construction of its probability mass function is computationally intensive for large sample size n. Under any other ranking procedure, the exact null distribution of $S_n(\eta_0)$ may depend on the ranking methods and the underlying distribution.

For large sample sizes, the asymptotic null distribution of $S_n(\eta_p)$ has a normal distribution with mean 0 and variance $\sigma^2_{S_n} = \mathrm{var}(\sqrt{n} S_n(\eta_p))$. This can be observed easily from Theorem 2 and Corollary 3 with $g(x-p) = I(x > \eta_p) - p$, $\hat{\theta}_{g,K}(\eta_p) = S_n(\eta_p)$,

$$\sigma^2_{S_n} = \mathrm{var}\left\{\sqrt{n} S_n(\eta_p)\right\} = \mathrm{var}\left\{\hat{F}_{\mathrm{RSS}(K)}(\eta_p)\right\} - \frac{1}{K} \sum_{r=1}^{K} F^2_{[r:K]}(\eta_p)$$
$$+ \sum_{r=1}^{K} \sum_{s=1}^{K} \frac{1}{H} \sum_{h=1}^{H} a_{r:K|h} a_{s:K|h} \{1 - F_{[r:K]}(\eta_p)\}\{1 - F_{[s:K]}(\eta_p)\}. \tag{7}$$

The point estimate of η_p can be obtained from the solution of $S_n(\hat{\eta}_p) = 0$. Since $S_n(b) = 0$ is a non-increasing function of b, the solution always exists, but may not be unique for small sample sizes. If $\hat{\eta}_p$ is not unique, one may modify the definition of the estimator to force it to be unique, such as $\hat{\eta}_p = \inf\{\eta_p : S_n(\eta_p)) > 0\}$. The asymptotic distribution of the estimator based on these two definitions are the same. To investigate asymptotic behavior of $\hat{\eta}_p$, we first provide a linear approximation for $S_n(b)$ in a compact set.

Theorem 3 *Let (X_i, R_i), $i = 1, \ldots, n$, be a JPS sample from a continuous distribution F having a density function $f(y) = \frac{\mathrm{d}}{\mathrm{d}y}F(y)$. Assume that $f(\eta_p) > 0$. For any $B > 0$,*

$$\sup_{|b| \le B} |\sqrt{n}S_n(\eta_p + b/\sqrt{n}) - \sqrt{n}S_n(\eta_p) - f(\eta_p)b| = o_p(1), \tag{8}$$

where $S_n(.)$ is defined in Eq. (6).

Using Lemma 3 in Ozturk (2012) with appropriate notation, one can show that $\hat{\eta_p}$ is bounded in probability. Hence, in Eq. (8), we insert $\hat{\eta}_p = \eta_p + b/\sqrt{n}$ and write

$$\sqrt{n}(\hat{\eta}_p - \eta_p) = \sqrt{n}\frac{S_n(\eta_p)}{f(\eta_p)} + o_p(1).$$

It is now easy to observe that $\sqrt{n}(\hat{\eta}_p - \eta_p)$ has a limiting normal distribution with mean zero and variance $\sigma^2_{\hat{\eta}_p} = \sigma^2_{S_n}/f^2(\eta_p)$, where $\sigma^2_{S_n}$ is given in Eq. (7).

4 Selection of subset size K

The construction of the reduced sets needs an integer K, $K \le H$. The selection of K depends on the sample size n. If the sample size n is large, the probability of having empty strata would be small. In this case, we select K to be large provided that $K \le H$. Since the probability of having empty strata is a function of sample size n, we provide a guidance to determine the selection of K. Let $W_n(q)$ be the number of judgment classes having more than q observations in each class, $q = 0, 1$. In other words, $W_n(q) = \sum_{h=1}^{H} I(N_h > q)$. In this notation, $W_n(0)$ gives the number of non-empty judgment classes and $W_n(1)$ gives the number of judgment classes having at least two observations in each class. Since the judgment class sample sizes N_h, $h = 1, \ldots, H$, are random variables, $W_n(q)$ has a discrete probability mass function.

Theorem 4 *Let (X_i, R_i), $i = 1, \ldots, n$, be a judgment post stratified sample. The probability mass function of $W_n(q)$ for $q = 0, 1$ is given by*

$$P(W_n(q) = K) = \frac{n!}{H^n}\binom{H}{K}\sum_{j=0}^{q(H-K)}\frac{K^{n-qj}\binom{q(H-K)}{j}}{(n-qj)!} \times \left[1 - \sum_{i=1}^{K}\binom{K}{i}(-1)^{i-1}\sum_{s=0}^{iq}\frac{\binom{iq}{s}\binom{n-qj}{s}s!(k-i)^{n-qj-s}}{k^{n-qj}}\right]$$

for $q = 0, 1$ and $K = 1, \ldots, H$.

We use Theorem 4 to determine K. For small sample size n, selecting large K leads to empty judgment classes. Hence, it may inflate the variance and introduce bias. Selecting a small K is also not desirable, since it reduces the data structure from a set size H to a set size K and increases the variance of the estimator. To find a reasonable K, we maximize $P(W_n(q) = K)$ with respect to K. Let K_q^* be the maximizer of $P(W_n(q) = K)$,

$$K_q^* = \max_{1 \leq K \leq H} P(W_n(q) = K). \tag{9}$$

The integer K_0^* gives a reduced set size yielding non-empty judgment classes with the highest probability, but it may still create highly unbalanced JPS sample. For example, some classes may have just one observation, and the other classes may have a large number of observations. The integer K_1^* gives a reduced set size yielding judgment classes having at least two observations with highest probability. Hence, it creates a more balanced sample.

5 Bootstrap inference

We established that the estimators have a normal distribution for large sample sizes. The rate of convergence is usually very slow due to highly unbalanced nature of the sample size vector. To draw inference for small sample sizes, the results of Sect. 3 may not be very useful. For finite samples, we construct bootstrap distributions of the proposed estimators to draw statistical inference. The bootstrap estimators can be obtained from a plug-in method. Let θ be a statistical functional, $\theta = T(F)$. The bootstrap estimate of θ can be obtained from $\hat{\theta} = T(\hat{F})$, where $\hat{F}$ is the empirical CDF. Let

$$\boldsymbol{Y}_i^\top = (X_i, a_{1:K|R_i}, \ldots, a_{K:K|R_i}), \quad i = 1, \ldots, n,$$

be the data set that contains the reduced set structure in a JPS sample. Let $\mathcal{F}$ be the empirical CDF of $\boldsymbol{Y}_i$, $i = 1, \ldots, n$. We generate a bootstrap re-sample, $\boldsymbol{Y}_i^*$, $i = 1, \ldots, n$, from $\mathcal{F}$ with replacement. To construct the bootstrap distribution of the estimator $\hat{\theta}_{g:K}(b)$, we generate re-samples $\boldsymbol{Y}_i^{*c}$, $i = 1, \ldots, n$, and compute

$$\hat{\theta}_{g:K}^{*c}(b) = \sum_{k=1}^{K} \frac{I_k^{*c}}{d_n^*} \frac{\sum_{i=1}^{n} a_{k:K|R_i^{*c}}^{*c} g(X_i^{*c} - b)}{\sum_{j=1}^{n} a_{k:K|R_j^{*c}}^{*c}}, \quad c = 1, \ldots, C.$$

The bootstrap variance estimate of $\hat{\theta}_{g:K}(b)$ is then obtained from

$$B(\hat{\theta}_{g:K}(b)) = \frac{1}{C-1} \sum_{c=1}^{C} \left(\hat{\theta}_{g:K}^{*c}(b) - \bar{\theta}_{g:K}^{*}(b) \right)^2,$$

where $\bar{\theta}_{g:K}^{*}(b)$ is the mean of $\hat{\theta}_{g:K}^{*c}(b)$, $c = 1, \ldots, C$.

A $(1-\alpha)100\,\%$ bootstrap percentile confidence interval is constructed by $(L^{\alpha/2}, L^{1-\alpha/2})$, where L^a is the ath quantile of $\hat{\theta}^*_{g:K}(b)$ satisfying $a = P(\hat{\theta}^*_{g:K}(b) \leq L^a|\mathcal{F})$ for $0 < a < 1$. The bootstrap variance estimate of $\hat{\eta}_p$ and confidence interval of η_p is constructed in a similar fashion.

6 Finite sample properties of the estimators

In this section, we perform a simulation study to investigate the finite sample properties of the estimator. The simulation study investigates the convergence rate and efficiency of the estimators. JPS samples are generated from standard normal and log-normal distributions with set size $H = 5$, and sample sizes $n = 15, 30$ and 100. The quality of ranking information is controlled by the additive perceptual error model in Dell and Clutter (1972). To determine the judgment rank of an observation X_i in set $S_{i,H} = \{X_i, Y_1, \ldots, Y_{H-1}\}$, we generate H independent normal random variable, $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_H)$, with mean zero and variance τ_ϵ. Let $\mathbf{Z}_i = (X_i, Y_1, \ldots, Y_{H-1})$ be an H dimensional independent random vector from distribution F with mean μ and variance σ^2. We add the vectors $\mathbf{Z}_i$ and $\boldsymbol{\epsilon}$ to construct the additive model

$$\boldsymbol{V}_i = (V_1, V_2, \ldots, V_H) = (X_i, Y_1, \ldots, Y_{H-1}) + (\epsilon_1, \ldots, \epsilon_H).$$

In this model, the rank of V_1 is taken as the judgment rank of X_i in set $S_{i,H}$. The quality of ranking information is controlled by the correlation coefficient between X and ϵ, $\rho = \text{corr}(X, Y) = \frac{1}{\sqrt{1+\tau_\epsilon^2/\sigma^2}}$. In the simulation study, we used $\rho = 1$ for perfect ranking and $\rho = 0.75$ for imperfect ranking.

The simulation study considered the estimation of two parameters, mean μ and median $\eta_{0.5}$ of standard normal and log-normal distributions. Tables 1 and 2 present the simulation results for estimator $\hat{\mu}_K$ for normal and log-normal distributions, respectively. The probability distribution of $W_n(q)$ in Theorem 4 is given by $P_0(K) = P(W_n(0) = K)$ for $q = 0$ and $P_1(K) = P(W_n(1) = K)$ for $q = 1$. The simulation size is taken to be 5000. The heading $C_{\hat{\mu}}(K)$ gives the coverage probability of the bootstrap percentile interval of μ. The simulated and bootstrap variance estimates of the estimators are given by $V(\hat{\mu}_K)$ and $B(\hat{\mu}_K)$, respectively. In these simulations, the bootstrap simulation size is taken to be 500. The relative efficiency of $\hat{\mu}_K$ with respect to $\hat{\mu}_{K_O^*}$ is given by the ratio of mean square errors (MSE)

$$\text{RE}_K(\hat{\mu}) = \frac{\text{MSE}(\hat{\mu}_K)}{\text{MSE}(\hat{\mu}_{K_O^*})}, \quad K = 1, \ldots, H, \quad K_O^* = \begin{cases} K_1^* & \text{if } n > 30 \\ K_1^* - 1 & \text{if } n \leq 30, \end{cases} \qquad (10)$$

where K_1^* is given in Eq. (9). The asymptotic variance of the estimator $\hat{\mu}_K$ for perfect ranking is given in the last column under the heading $A(\hat{\mu}_K)$.

Tables 1 and 2 reveal that the estimators $\hat{\mu}_K$ are essentially unbiased for the population means ($\mu = 0$ for the standard normal distribution and $\mu = exp(1/2) = 1.649$ for the standard log-normal distribution) for all K and sample sizes $n = 15, 30$, and 100. The efficiency of the estimator $\hat{\mu}_K$ with respect to $\hat{\mu}_{K_1^*}$ depends on the sample size n and the quality of ranking information. Even though the estimator $\hat{\mu}_{K_1^*}$ always

Table 1 Mean, variance and efficiency of the mean estimator $\hat{\mu}_K$ from a standard normal distribution, $H = 5$, simulation and bootstrap simulation sizes are 5000 and 500, respectively

n	ρ	K	$P_0(K)$	$P_1(K)$	$\hat{\mu}_K$	$C_{\hat{\mu}}(K)$	$V(\hat{\mu}_K)$	$\mathrm{RE}_K(\hat{\mu})$	$B(\hat{\mu}_K)$	$A(\hat{\mu}_K)$
15	1.00	1	0.000	0.000	−0.002	0.916	1.010	2.332	0.933	1.000
15	1.00	2	0.000	0.006	−0.001	0.917	0.525	1.212	0.508	0.522
15	1.00	3	0.005	0.145	−0.004	0.899	0.455	1.051	0.399	0.403
15	1.00	4	0.167	0.530	−0.002	0.896	0.433	1.000	0.399	0.368
15	1.00	5	0.829	0.320	−0.003	0.901	0.589	1.360	0.580	0.361
15	0.75	1	0.000	0.000	−0.001	0.919	1.004	1.375	0.934	
15	0.75	2	0.000	0.006	−0.002	0.919	0.720	0.986	0.691	
15	0.75	3	0.005	0.145	0.004	0.910	0.691	0.947	0.631	
15	0.75	4	0.167	0.530	−0.001	0.900	0.730	1.000	0.644	
15	0.75	5	0.829	0.320	−0.003	0.898	0.884	1.211	0.748	
30	1.00	1	0.000	0.000	0.001	0.931	0.985	2.234	0.963	1.000
30	1.00	2	0.000	0.000	−0.000	0.929	0.534	1.211	0.515	0.522
30	1.00	3	0.000	0.000	−0.003	0.927	0.426	0.966	0.399	0.403
30	1.00	4	0.006	0.052	0.002	0.924	0.396	0.898	0.370	0.368
30	1.00	5	0.994	0.948	0.001	0.933	0.441	1.000	0.471	0.361
30	0.75	1	0.000	0.000	−0.002	0.938	0.999	1.282	0.964	
30	0.75	2	0.000	0.000	0.002	0.930	0.742	0.952	0.713	
30	0.75	3	0.000	0.000	0.002	0.933	0.654	0.840	0.646	
30	0.75	4	0.006	0.052	0.001	0.926	0.697	0.895	0.638	
30	0.75	5	0.994	0.948	−0.002	0.931	0.779	1.000	0.761	
100	1.00	1	0.000	0.000	0.001	0.945	1.000	2.663	0.991	1.000
100	1.00	2	0.000	0.000	−0.000	0.941	0.531	1.414	0.521	0.522
100	1.00	3	0.000	0.000	−0.001	0.943	0.407	1.083	0.402	0.403
100	1.00	4	0.000	0.000	−0.001	0.941	0.384	1.023	0.367	0.368
100	1.00	5	1.000	1.000	−0.000	0.943	0.376	1.000	0.374	0.361
100	0.75	1	0.000	0.000	−0.001	0.943	1.027	1.501	0.994	
100	0.75	2	0.000	0.000	0.001	0.944	0.720	1.053	0.723	
100	0.75	3	0.000	0.000	−0.001	0.938	0.673	0.984	0.659	
100	0.75	4	0.000	0.000	−0.001	0.944	0.646	0.944	0.646	
100	0.75	5	1.000	1.000	−0.003	0.938	0.684	1.000	0.663	

improves $\hat{\mu}_H$ and $\hat{\mu}_1$, it is not uniformly better than $\hat{\mu}_K$ for all K. For example, for normal distribution when $n = 15$, $\rho = 0.75$, the reduced set size $K_1^* = 4$ maximizes $P(W_n(1) = K)$, but $\hat{\mu}_{K_1^*} = 0.730$ and $\hat{\mu}_{K_1^*-1} = 0.691$. In this case, $\hat{\mu}_{K_1^*-1}$ has the smallest variance. Over all, the reduced set size $K_O^* = K_1^*$ appears to be working reasonably well for moderately large sample sizes($n \geq 30$). For small sample sizes ($n \leq 30$) we suggest using $K_O^* = K_1^* - 1$ for the reduced set sizes.

Tables 1 and 2 also reveal that the rate of convergence of the estimator is slow for large values of K. For example, for standard log normal distribution, when $n = 15$ and $K = 1$ the simulated variance estimate $V(\hat{\mu}_1) = 4.774$ is very close to the

Table 2 Mean, variance and efficiency of the median estimator $\hat{\mu}_K$ from a standard log normal distribution, $H = 5$, simulation and bootstrap simulation sizes are 5000 and 500, respectively

n	ρ	K	$P_0(K)$	$P_1(K)$	$\hat{\mu}_K$	$C_{\hat{\mu}}(K)$	$V(\hat{\mu}_K)$	$\text{RE}_K(\hat{\mu})$	$B(\hat{\mu}_K)$	$A(\hat{\mu}_K)$
15	1.00	1	0.000	0.000	1.650	0.830	4.774	1.518	4.436	4.671
15	1.00	2	0.000	0.006	1.625	0.800	3.429	1.097	2.854	3.566
15	1.00	3	0.005	0.145	1.642	0.792	3.416	1.093	2.723	3.160
15	1.00	4	0.167	0.530	1.623	0.776	3.135	1.000	2.475	2.988
15	1.00	5	0.829	0.320	1.651	0.804	3.832	1.225	3.304	2.939
15	0.75	1	0.000	0.000	1.650	0.838	4.600	1.123	4.312	
15	0.75	2	0.000	0.006	1.632	0.833	3.943	0.963	3.343	
15	0.75	3	0.005	0.145	1.627	0.824	3.827	0.935	3.121	
15	0.75	4	0.167	0.530	1.633	0.818	4.095	1.000	3.122	
15	0.75	5	0.829	0.320	1.647	0.809	4.793	1.170	3.775	
30	1.00	1	0.000	0.000	1.649	0.876	4.602	1.111	4.425	4.671
30	1.00	2	0.000	0.000	1.641	0.849	3.635	0.878	3.342	3.566
30	1.00	3	0.000	0.000	1.639	0.836	3.184	0.769	2.865	3.160
30	1.00	4	0.006	0.052	1.638	0.833	3.273	0.790	2.801	2.988
30	1.00	5	0.994	0.948	1.651	0.853	4.141	1.000	3.531	2.939
30	0.75	1	0.000	0.000	1.647	0.886	4.446	0.935	4.319	
30	0.75	2	0.000	0.000	1.642	0.870	3.974	0.836	3.707	
30	0.75	3	0.000	0.000	1.632	0.873	3.574	0.752	3.316	
30	0.75	4	0.006	0.052	1.653	0.868	4.250	0.894	3.767	
30	0.75	5	0.994	0.948	1.654	0.882	4.753	1.000	4.547	
100	1.00	1	0.000	0.000	1.649	0.914	4.720	1.506	4.684	4.671
100	1.00	2	0.000	0.000	1.648	0.902	3.514	1.121	3.455	3.566
100	1.00	3	0.000	0.000	1.646	0.901	3.191	1.018	3.066	3.160
100	1.00	4	0.000	0.000	1.641	0.899	2.936	0.937	2.869	2.988
100	1.00	5	1.000	1.000	1.653	0.908	3.135	1.000	3.127	2.939
100	0.75	1	0.000	0.000	1.649	0.917	4.451	1.155	4.486	
100	0.75	2	0.000	0.000	1.647	0.907	4.167	1.081	4.037	
100	0.75	3	0.000	0.000	1.645	0.908	3.961	1.028	3.826	
100	0.75	4	0.000	0.000	1.645	0.909	3.936	1.021	3.782	
100	0.75	5	1.000	1.000	1.647	0.923	3.854	1.000	3.926	

asymptotic variance $A(\hat{\mu}_1) = 4.671$. On the other hand, when $n = 15$ and $K = 5$ there is a big difference between the simulated variance estimate and the asymptotic variance, $V(\hat{\mu}_5) = 3.832$ and $A(\hat{\mu}_1) = 2.939$. For large sample size n, $n = 100$, the simulated variance estimates and the asymptotic variances are practically equal for all $K = 1, \ldots, H$. Similar results also hold for standard normal distribution.

The simulation study also investigated the properties of the bootstrap variance estimates and coverage probabilities ($C_{\hat{\mu}}(K_O^*)$ of the percentile confidence interval of population mean. For small sample sizes n, $n = 15$, the bootstrap variance estimate, $B(\hat{\mu}_K)$, slightly under estimates the simulated variance estimates $V(\hat{\mu}_K)$. For mod-

Table 3 Mean, variance and efficiency of the mean estimator $\tilde{\mu}_K$ from a standard normal distribution, $H = 5$, simulation and bootstrap simulation sizes are 5000 and 500, respectively

n	ρ	K	$P_0(K)$	$P_1(K)$	$\tilde{\mu}_K$	$C_{\tilde{\mu}}(K)$	$V(\tilde{\mu}_K)$	$\mathrm{RE}_K(\tilde{\mu})$	$B(\tilde{\mu}_K)$	$A(\tilde{\mu}_K)$
15	1.00	1	0.000	0.000	−0.004	0.920	1.520	1.761	1.873	1.571
15	1.00	2	0.000	0.006	0.005	0.923	0.985	1.141	1.196	0.982
15	1.00	3	0.005	0.145	0.000	0.910	0.894	1.036	1.086	0.835
15	1.00	4	0.167	0.530	−0.005	0.919	0.863	1.000	0.983	0.785
15	1.00	5	0.829	0.320	−0.003	0.947	1.041	1.206	1.124	0.773
15	0.75	1	0.000	0.000	0.002	0.926	1.527	1.141	1.894	
15	0.75	2	0.000	0.006	−0.004	0.922	1.306	0.976	1.548	
15	0.75	3	0.005	0.145	0.005	0.920	1.238	0.925	1.488	
15	0.75	4	0.167	0.530	0.004	0.913	1.338	1.000	1.538	
15	0.75	5	0.829	0.320	−0.006	0.926	1.578	1.179	1.532	
30	1.00	1	0.000	0.000	0.000	0.956	1.494	1.591	1.715	1.571
30	1.00	2	0.000	0.000	−0.003	0.928	0.999	1.064	1.151	0.982
30	1.00	3	0.000	0.000	−0.004	0.937	0.845	0.900	1.030	0.835
30	1.00	4	0.006	0.052	0.004	0.936	0.803	0.855	0.974	0.785
30	1.00	5	0.994	0.948	0.001	0.937	0.939	1.000	1.142	0.773
30	0.75	1	0.000	0.000	−0.001	0.963	1.460	1.021	1.706	
30	0.75	2	0.000	0.000	0.001	0.937	1.265	0.885	1.485	
30	0.75	3	0.000	0.000	0.004	0.939	1.190	0.832	1.417	
30	0.75	4	0.006	0.052	0.001	0.933	1.259	0.878	1.436	
30	0.75	5	0.994	0.948	−0.003	0.936	1.430	1.000	1.674	
100	1.00	1	0.000	0.000	0.001	0.953	1.558	1.907	1.687	1.571
100	1.00	2	0.000	0.000	−0.002	0.941	0.972	1.190	1.098	0.982
100	1.00	3	0.000	0.000	−0.004	0.945	0.841	1.029	0.950	0.835
100	1.00	4	0.000	0.000	−0.000	0.940	0.785	0.961	0.905	0.785
100	1.00	5	1.000	1.000	−0.001	0.941	0.817	1.000	0.936	0.773
100	0.75	1	0.000	0.000	−0.003	0.951	1.587	1.303	1.693	
100	0.75	2	0.000	0.000	0.002	0.947	1.247	1.024	1.409	
100	0.75	3	0.000	0.000	−0.002	0.938	1.224	1.005	1.340	
100	0.75	4	0.000	0.000	0.001	0.946	1.182	0.970	1.330	
100	0.75	5	1.000	1.000	−0.003	0.941	1.218	1.000	1.370	

erate ($n = 30$) and large sample sizes ($n = 100$), both the simulated and bootstrap variance estimates are close to each other. Similar pattern also appears in the coverage probabilities. For smaller sample sizes ($n = 15$), the coverage probabilities are smaller than the nominal coverage probability 0.95. The bootstrap confidence intervals provide reasonable coverage probabilities for the moderate and large sample sizes.

Tables 3 and 4 present the simulation results for the median estimator, $\tilde{\mu}_K = \hat{\eta}_{0.5}$, from standard normal and log normal distributions. Results similar to the ones we observed in Tables 1 and 2 hold for the median estimators as well. We note that the median estimators $\tilde{\mu}_K$ appear to have a negligible amount of bias for the population

Table 4 Mean, variance and efficiency of the median estimator $\hat{\mu}_K$ from a standard log normal distribution, $H = 5$, simulation and bootstrap simulation sizes are 5000 and 500, respectively

n	ρ	K	$P_0(K)$	$P_1(K)$	$\hat{\mu}_K$	$C_{\hat{\mu}}(K)$	$V(\hat{\mu}_K)$	$\mathrm{RE}_K(\hat{\mu})$	$B(\hat{\mu}_K)$	$A(\hat{\mu}_K)$
15	1.00	1	0.000	0.000	1.052	0.918	1.811	1.882	2.818	1.571
15	1.00	2	0.000	0.006	1.030	0.923	1.084	1.126	1.509	0.982
15	1.00	3	0.005	0.145	1.033	0.914	0.973	1.011	1.419	0.835
15	1.00	4	0.167	0.530	1.026	0.891	0.963	1.000	1.321	0.785
15	1.00	5	0.829	0.320	1.044	0.921	1.223	1.271	1.553	0.773
15	0.75	1	0.000	0.000	1.050	0.920	1.732	1.009	2.777	
15	0.75	2	0.000	0.006	1.049	0.918	1.620	0.943	2.173	
15	0.75	3	0.005	0.145	1.046	0.921	1.645	0.958	2.190	
15	0.75	4	0.167	0.530	1.041	0.907	1.718	1.000	2.279	
15	0.75	5	0.829	0.320	1.054	0.901	2.119	1.234	2.336	
30	1.00	1	0.000	0.000	1.037	0.940	1.685	1.722	2.095	1.571
30	1.00	2	0.000	0.000	1.019	0.933	1.097	1.120	1.306	0.982
30	1.00	3	0.000	0.000	1.018	0.938	0.908	0.927	1.164	0.835
30	1.00	4	0.006	0.052	1.016	0.921	0.893	0.912	1.097	0.785
30	1.00	5	0.994	0.948	1.021	0.938	0.979	1.000	1.310	0.773
30	0.75	1	0.000	0.000	1.031	0.940	1.702	0.968	2.090	
30	0.75	2	0.000	0.000	1.021	0.940	1.481	0.842	1.879	
30	0.75	3	0.000	0.000	1.020	0.937	1.515	0.862	1.870	
30	0.75	4	0.006	0.052	1.015	0.931	1.506	0.856	1.897	
30	0.75	5	0.994	0.948	1.028	0.931	1.758	1.000	2.183	
100	1.00	1	0.000	0.000	1.008	0.947	1.570	1.960	1.778	1.571
100	1.00	2	0.000	0.000	1.007	0.946	1.022	1.276	1.140	0.982
100	1.00	3	0.000	0.000	1.004	0.938	0.885	1.105	0.981	0.835
100	1.00	4	0.000	0.000	1.004	0.942	0.789	0.985	0.921	0.785
100	1.00	5	1.000	1.000	1.004	0.946	0.801	1.000	0.945	0.773
100	0.75	1	0.000	0.000	1.011	0.945	1.607	1.101	1.816	
100	0.75	2	0.000	0.000	1.009	0.945	1.437	0.984	1.667	
100	0.75	3	0.000	0.000	1.004	0.939	1.431	0.980	1.597	
100	0.75	4	0.000	0.000	1.004	0.943	1.409	0.965	1.586	
100	0.75	5	1.000	1.000	1.006	0.943	1.460	1.000	1.681	

medians of normal ($\eta_{0.5} = 0$) and log-normal ($\eta_{0.5} = 1$) distributions for all K, n and ρ. Again inspection of the efficiency values reveals that the selection of K_1^* and $K_1^* - 1$ for $\tilde{\mu}_K$ yields a reasonable estimator for moderately large ($n \geq 30$) and small ($n < 30$) sample sizes, respectively. The rate of convergence, the bootstrap variance estimates and the bootstrap coverage probabilities of the median confidence intervals follow patterns similar to the ones we observed in Tables 1 and 2.

We conduct another simulation study to compare our estimator with its competing isotonized JPS estimator in Wang et al. (2008) and JPS estimator in MacEachern et al. (2004). The isotonized ($\hat{\mu}_I$) and JPS ($\hat{\mu}_J$) estimators of population mean are given by

$$\hat{\mu}_I = \frac{1}{H}\sum_{h=1}^{H} \bar{X}^*_{[h]}, \quad \bar{X}^*_{[h]} = \max_{r \le h} \min_{s \ge h} \sum_{g=r}^{s} \frac{n_g \bar{X}_{[g]}}{n_{rs}},$$

$$n_{rs} = \sum_{g=r}^{s} n_g \quad \text{and} \quad \hat{\mu}_J = \frac{1}{H}\sum_{h=1}^{H} \bar{X}_{[h]},$$

where $\bar{X}_{[g]}$ and n_g are the sample mean and size for the observations falling in the judgment class g, $g = 1, \ldots, H$. The simulation study is conducted with set sizes; $H = 2, 3, 4, 5, 10$, average judgment class sizes; $\bar{n} = 1, 2, 3, 4, 5$, so that the actual sample sizes are $n = H\bar{n}$, and perfect ($\rho = 1$) and random ($\rho = 0$) ranking. The JPS samples are generated from standard normal ($N(0, 1)$), uniform ($U(0, 1)$) and standard lognormal ($LN(0, 1)$) distributions. The number of replications in the simulation is taken to be 20,000.

Table 5 provides relative efficiencies of the isotonized and JPS mean estimators with respect to $\hat{\mu}_{K^*_O}$,

$$\mathrm{RE}_1 = \mathrm{MSE}(\hat{\mu}_I)/\mathrm{MSE}(\hat{\mu}_{K^*_O}), \quad \mathrm{RE}_2 = \mathrm{MSE}(\hat{\mu}_J)/\mathrm{MSE}(\hat{\mu}_{K^*_O}),$$

where K^*_O is defined in Eq. (10). It is clear from Table 5 that the proposed estimator ($\hat{\mu}_{K^*_O}$) is as good as or better than both $\hat{\mu}_I$ and $\hat{\mu}_J$ for most $\bar{n}$ ($\bar{n} > 2$) and H ($H > 2$) with the exception that $\hat{\mu}_I$ and μ_J are slightly better than $\hat{\mu}_{K^*_O}$ when $H = 2, \bar{n} \le 2$ and $\rho = 1$ for normal and uniform distributions. Another important observation in Table 5 is that the proposed estimator $\hat{\mu}_{K^*_O}$ outperforms both $\hat{\mu}_I$ and $\hat{\mu}_J$ even if there is no viable ranking information ($\rho = 0$).

7 Example

We apply the proposed estimator to 2012 United States Department of Agriculture (USDA) census data to estimate the mean corn production in bushels at county levels. We downloaded the USDA census data from http://quickstats.nass.usda.gov for the variable corn production and the number of agricultural land operations in each county in seven states. These states include California, Arizona, Iowa, Ohio, Florida, Kansas, Oklahoma and Pennsylvania. We removed the counties where either the corn production or the number of operations was not reported due to the identifiability concern of certain operations. We considered 467 counties in these states as our population. Our interest in this population is to estimate the mean corn production (X, in 100,000 bushels) by using the number of agricultural land operations (Y) as an auxiliary variable. The corn production in each county is divided by 100,000 to simplify the computation. The scatter plot of X and Y variables indicate that the corn production (X) is highly nonlinear and exponentially decaying with Y. Since regression and ratio estimators require that X must be linear in Y, we used $Y^* = 1/\sqrt{Y}$ transformation. After the transformation, relationship between X and Y^* was approximately linear. The population means of Y^* and X are 0.217 and 58.98, respectively. Pearson corre-

Table 5 Relative efficiency of $\hat{\mu}_{K_O^*}$ with respect to isotonized ($\hat{\mu}_I$) and JPS ($\hat{\mu}_J$) mean estimator, $\mathrm{RE}_I = \mathrm{MSE}(\hat{\mu}_I)/\mathrm{MSE}(\hat{\mu}_{K_O^*})$, $\mathrm{RE}_J = \mathrm{MSE}(\hat{\mu}_J)/\mathrm{MSE}(\hat{\mu}_{K_O^*})$

D	H	ρ	$\bar{n}=1$		$\bar{n}=2$		$\bar{n}=3$		$\bar{n}=4$		$\bar{n}=5$	
			RE_I	RE_J	RE_I	RE_J	RE_I	RE_J	RE_I	RE_J	RE_I	RE_J
$N(0,1)$	2	1.00	1.000	1.000	0.935	0.955	0.866	0.883	0.819	0.833	0.782	0.788
	3	1.00	0.890	0.964	0.758	0.836	1.093	1.179	1.061	1.108	1.015	1.041
	4	1.00	0.771	0.908	1.067	1.243	0.998	1.112	1.112	1.189	1.075	1.120
	5	1.00	0.698	0.865	0.963	1.213	1.105	1.301	1.114	1.230	1.088	1.158
	10	1.00	0.878	1.402	1.197	1.976	1.166	1.669	1.088	1.349	0.869	1.000
	2	0.00	1.000	1.000	1.085	1.170	1.109	1.207	1.091	1.184	1.070	1.137
	3	0.00	1.087	1.084	1.119	1.226	1.043	1.187	1.054	1.199	1.054	1.166
	4	0.00	1.117	1.119	1.072	1.206	1.081	1.256	1.018	1.177	1.021	1.154
	5	0.00	1.140	1.135	1.102	1.236	1.049	1.244	1.001	1.162	1.004	1.158
	10	0.00	1.144	1.169	1.048	1.240	0.971	1.214	0.942	1.160	0.824	1.000
$U(0,1)$	2	1.00	1.000	1.000	0.921	0.947	0.843	0.869	0.800	0.814	0.759	0.767
	3	1.00	0.917	0.954	0.745	0.813	1.061	1.149	1.020	1.080	0.999	1.035
	4	1.00	0.818	0.903	1.021	1.218	0.959	1.098	1.072	1.173	1.058	1.120
	5	1.00	0.717	0.859	0.924	1.177	1.062	1.283	1.080	1.239	1.076	1.174
	10	1.00	0.822	1.376	1.148	2.104	1.142	1.758	1.063	1.427	0.829	1.000
	2	0.00	1.000	1.000	1.085	1.177	1.104	1.213	1.097	1.190	1.070	1.140
	3	0.00	1.088	1.083	1.111	1.219	1.042	1.188	1.053	1.176	1.042	1.159
	4	0.00	1.118	1.125	1.075	1.196	1.077	1.246	1.021	1.173	1.024	1.151
	5	0.00	1.148	1.145	1.094	1.238	1.048	1.234	0.998	1.175	1.007	1.164
	10	0.00	1.133	1.164	1.047	1.234	0.977	1.209	0.947	1.161	0.835	1.000
$LN(0,1)$	2	1.00	1.000	1.000	1.071	1.078	1.028	1.035	0.956	0.964	0.942	0.945
	3	1.00	1.146	1.035	0.928	0.925	1.193	1.213	1.145	1.163	1.103	1.108
	4	1.00	1.134	1.011	1.255	1.171	1.233	1.237	1.154	1.167	1.150	1.164
	5	1.00	1.202	1.022	1.232	1.196	1.217	1.225	1.180	1.210	1.194	1.211
	10	1.00	1.305	1.187	1.379	1.452	1.297	1.348	1.157	1.210	0.977	1.000
	2	0.00	1.000	1.000	1.094	1.182	1.125	1.246	1.053	1.174	1.146	1.267
	3	0.00	1.112	1.073	1.141	1.334	1.072	1.212	1.037	1.243	1.038	1.247
	4	0.00	1.205	1.156	1.087	1.163	1.108	1.293	1.038	1.211	1.032	1.190
	5	0.00	1.300	1.140	1.148	1.221	1.108	1.287	1.015	1.179	0.996	1.127
	10	0.00	1.209	1.160	1.067	1.266	1.004	1.194	0.961	1.162	0.814	1.000

lation coefficient between X and Y^* is 0.289. The histogram of X is strongly skewed to right. The population variances of Y^* and X are 0.020 and 6817.667, respectively.

We performed another simulation study by constructing JPS samples from this population. Simulation parameters are taken to be $\bar{n} = 2, 3, 4, 5, 6$, $H = 3, 4, 5, 10$. The sample sizes are constructed from $n = \bar{n}H$. For each simulation parameter combination, we generated 20,000 JPS samples and computed proposed ($\hat{\mu}_{K_O^*}$), isotonized ($\hat{\mu}_I$), JPS ($\hat{\mu}_J$), regression ($\hat{\mu}_R$) and ratio ($\hat{\mu}_{\mathrm{Ra}}$) estimators. The ratio and regression estimators are given by

Table 6 Biases and relative efficiencies of the estimators of the mean corn production (in 100,000 bushels) of 467 counties in seven states in the USA, $\mathrm{RE}_I = \mathrm{MSE}(\hat{\mu}_I)/\mathrm{MSE}(\hat{\mu}_{K_O^*})$, $\mathrm{RE}_J = \mathrm{MSE}(\hat{\mu}_J)/\mathrm{MSE}(\hat{\mu}_{K_O^*})$, $\mathrm{RE}_R = \mathrm{MSE}(\hat{\mu}_R)/\mathrm{MSE}(\hat{\mu}_{K_O^*})$, $\mathrm{RE}_{\mathrm{Ra}} = \mathrm{MSE}(\hat{\mu}_{\mathrm{Ra}})/\mathrm{MSE}(\hat{\mu}_{K_O^*})$, $\hat{\mu}_R$ and $\hat{\mu}_{\mathrm{Ra}}$ are regression and ratio estimators, respectively

$\bar{n}$	H	K_O^*	Bias					Efficiency			
			$\hat{\mu}_{K_O^*}$	$\hat{\mu}_I$	$\hat{\mu}_J$	$\hat{\mu}_R$	$\hat{\mu}_{\mathrm{Ra}}$	RE_I	RE_J	RE_R	$\mathrm{RE}_{\mathrm{Ra}}$
2	3	1	−0.236	0.593	−0.123	1.241	2.135	1.136	1.191	1.003	1.429
3	3	2	−0.232	0.190	−0.209	0.876	2.059	1.082	1.182	0.983	1.226
4	3	2	0.267	0.506	0.323	0.892	1.982	1.069	1.164	0.976	1.139
5	3	2	0.030	0.216	0.105	0.728	1.674	1.060	1.146	0.987	1.117
6	3	2	0.126	0.356	0.296	0.705	1.414	1.062	1.140	0.985	1.091
2	4	2	−0.161	0.991	0.068	1.034	2.284	1.135	1.211	1.001	1.283
3	4	2	0.062	0.651	0.209	0.799	1.958	1.125	1.242	0.996	1.178
4	4	3	0.128	0.368	0.206	0.664	1.684	1.063	1.191	0.969	1.096
5	4	3	0.011	0.209	0.106	0.443	1.172	1.046	1.165	0.972	1.050
6	4	3	0.088	0.166	0.041	0.548	1.108	1.031	1.115	0.972	1.042
2	5	2	−0.099	1.015	−0.093	0.824	1.947	1.156	1.244	1.007	1.217
3	5	3	0.086	0.638	0.181	0.729	1.730	1.098	1.224	0.992	1.123
4	5	4	−0.104	0.194	−0.056	0.451	1.201	1.034	1.184	0.952	1.042
5	5	4	0.011	0.182	0.061	0.360	1.008	1.031	1.154	0.962	1.029
6	5	5	0.032	0.107	0.032	0.286	0.770	0.906	1.000	0.858	0.905
2	10	5	−0.213	0.921	−0.096	0.309	1.007	1.135	1.259	0.992	1.086
3	10	8	0.035	0.496	0.053	0.391	0.925	1.017	1.216	0.950	1.003
4	10	9	0.002	0.225	−0.002	0.283	0.681	0.978	1.169	0.942	0.961
5	10	10	0.045	0.164	0.045	0.168	0.460	0.850	1.000	0.817	0.829
6	10	10	−0.001	0.098	−0.001	0.176	0.419	0.876	1.000	0.856	0.861

The integer K_O^* is defined in Eq. (10)

$$\hat{\mu}_{\mathrm{Ra}} = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} Y_i^*}\bar{y}^*, \quad \hat{\mu}_R = \bar{X} + \hat{\beta}(\bar{Y}^* - \bar{y}^*),$$

where $\bar{X}$, $\bar{Y}^*$ are the sample averages of X- and Y^*-variables, and $\bar{y}^*$ is the population average of Y^*-variable. The slope parameter $\hat{\beta}$ is obtained from the regression fit of Y^* and X.

Table 6 presents the biases and relative efficiencies of the mean estimators. The relative efficiencies are given as the ratio of MSEs

$$\mathrm{RE}_I = \frac{\mathrm{MSE}(\hat{\mu}_I)}{\mathrm{MSE}(\hat{\mu}_{K_O^*})}, \quad \mathrm{RE}_J = \frac{\mathrm{MSE}(\hat{\mu}_J)}{\mathrm{MSE}(\hat{\mu}_{K_O^*})},$$
$$\mathrm{RE}_R = \frac{\mathrm{MSE}(\hat{\mu}_R)}{\mathrm{MSE}(\hat{\mu}_{K_O^*})}, \quad \mathrm{RE}_{\mathrm{Ra}} = \frac{\mathrm{MSE}(\hat{\mu}_{\mathrm{Ra}})}{\mathrm{MSE}(\hat{\mu}_{K_O^*})},$$

where K_O^* is defined in Eq. (10). The value of RE_R greater than 1 indicates that the proposed estimator has higher efficiency than the regression estimator. It is clear from Table 6 that the proposed estimator has substantially lower biases and higher efficiencies than the ratio estimator. On the other hand, it has smaller bias but comparable efficiency with respect to regression estimators. We note that regression and ratio estimators require a transformation for the linearity assumption. The proposed estimator is not affected by nonlinear relationship between response and auxiliary variables and it does not require a transformation. As long as there exists monotonicity between X and Y to rank the within set units, the proposed estimator performs fairly well. The biases of $\hat{\mu}_{K_O^*}$, $\hat{\mu}_I$ and $\hat{\mu}_J$ are all comparable, but the proposed estimator $\hat{\mu}_{K_O^*}$ has higher relative efficiency than the other two estimators $\hat{\mu}_I$ and $\hat{\mu}_J$.

8 Concluding remark

The judgment class sample sizes in a JPS sample form a multinomial random vector. Hence, it is highly possible that some of its entries become zero. The zero sample size does not only produce bias, it may also decrease the efficiency of the estimators. To reduce the impact of the empty judgment classes, we reconstructed a JPS sample with a reduced set size K, by conditioning on the ranks of the measured observations in the original JPS sample of the set size H.

We have used the reduced set JPS sample to construct a set of estimators for population characteristics for each reduced set size $K = 1, \ldots, H$. These estimators reduce to SRS estimator for $K = 1$ and JPS estimator for $K = H$. The choice of the reduced set size K in the construction plays an important role in the efficiency of the estimator. As a general rule, we suggest selecting a set size K so that the probability of having at least two observations in each of the K set is maximized.

We developed the asymptotic distribution of the reduced set JPS sample estimators. A simulation study suggested that the convergence rate is slow for large reduced set size K. We constructed a bootstrap inference for moderate sample sizes. The simulation study indicates that the bootstrap inference works reasonably well for moderate sample sizes.

Acknowledgements The author thanks Professor Jesse Frey who provided the sketch of the proof of the unbiasedness of the estimator in Theorem 1, and the Editor, Associate Editor and two anonymous referees for their helpful comments to improve the presentation of the paper.

Appendix

Proof of Theorem 1 For the proof of (i), from Eqs. (2) and (3) and from the law of large numbers, we write

$$\lim_{n\to\infty}\left\{\frac{1}{n}\sum_{i=1}^{n} a_{k:K|R_i}\, g(X_i - b)\right\} = \frac{1}{K} E\{g(X_{[k:K]} - b)\} + o_p(1)$$

and

$$\lim_{n\to\infty}\left\{\frac{1}{n}\sum_{i=1}^{n}a_{k:K|R_i}\right\}=\frac{1}{K}+o_p(1).$$

It is now easy to observe that

$$\begin{aligned}\lim_{n\to\infty}\hat{\theta}_{g,K}&=\sum_{k=1}^{K}\lim_{n\to\infty}\left(\frac{I_k}{d_n}\right)E\{g(X_{[k:K]}-b)\}+o_p(1)\\&=\sum_{k=1}^{K}\frac{E\{g(X_{[k:K]}-b)\}}{K}+o_p(1)=\theta_g+o_p(1).\end{aligned}$$

For the proof of (ii), without loss of generality assume that H is an even integer and $b=0$. We consider

$$\begin{aligned}E(\hat{\theta}_{g,K})&=E\left\{E(\hat{\theta}_{g,K}|\boldsymbol{R})\right\}=\sum_{h=1}^{H}E\left\{\sum_{k=1}^{K}\frac{I_k a_{k:K|h}N_h}{d_n\sum_{h=1}^{H}a_{k:K|h}N_h}\right\}E(g(X_{[h:H]}))\\&=\sum_{h=1}^{H}E(w_h)E(g(X_{[h:H]}))=\sum_{h=1}^{H/2}\{E(w_h)E(g(X_{[h:H]}))\\&\quad+E(w_{H+1-h})E(g(X_{[H+1-h:H]}))\}.\end{aligned}$$

In a JPS sample, since $E(N_h)=E(N_{H+1-h})$, the expected weights also preserve the equality $E(w_h)=E(w_{H+1-h})$. By using the assumption of the theorem and the symmetry of the expected weights, we write

$$\begin{aligned}E(\hat{\theta}_{g,K})&=\sum_{h=1}^{H/2}E(w_h)\left\{E(g(X_{[h:H]}-b))+E(g(X_{[H+1-h:H]}-b))\right\}\\&=\sum_{h=1}^{H/2}E(w_h)E(g(X-b))/2=E(g(X-b))\sum_{h=1}^{H}E(w_h)=E(g(X-b)).\end{aligned}$$

This completes the proof. □

Proof of Lemma (1) $\bar{\boldsymbol{T}}_g(b)$ is the mean of independent random vectors. Hence, central limit theorem gives the asymptotic normality. The expression $E(\bar{\boldsymbol{T}}_g)$ follows from Eqs. (2) and (3). We now compute the variance and covariances. For $\sigma_{r,s}$, $1\le r,s\le K$, we use the conditional covariance given R_i

$$\begin{aligned}\sigma_{r,s} &= \frac{1}{n}\sum_{h=1}^{H} E\left\{\sum_{i=1}^{n} I(R_i = h)\text{cov}(W_{r,R_i,g}(b), W_{s,R_i,g}(b))|R_i\right\} \\ &\quad + \frac{1}{n}\text{cov}\left\{\sum_{i=1}^{n} I(R_i = h)E(W_{r,R_i,g}(b)|R_i), \sum_{i=1}^{n} I(R_i = h)E(W_{s,R_i,g}(b)|R_i)\right\} \\ &= A + B.\end{aligned}$$

In the above equation the first term reduces to

$$\begin{aligned}A &= \sum_{h=1}^{H} E\left(\frac{N_h}{n}\right) a_{r:K|h} a_{s:K|h}\text{var}(g(X_{[h:H]} - b)) \\ &= \frac{1}{H}\sum_{h=1} a_{r:K|h} a_{s:K|h}\text{var}(g(X_{[h:H]} - b)).\end{aligned}$$

In a similar approach, we have

$$\begin{aligned}B &= \frac{1}{n}\text{cov}\left\{\sum_{h=1}^{H} N_h a_{r:K|h} E(g(X_{[h:H]} - b)), \sum_{h=1}^{H} N_h a_{s:K|h} E(g(X_{[h:H]} - b))\right\} \\ &= \sum_{h=1}^{H}\sum_{h'=1}^{H} \frac{\text{cov}(N_h, N_{h'})}{n} a_{r:K|h} E(g(X_{[h:H]} - b)) a_{r:K|h'} E(g(X_{[h':H]} - b)) \\ &= \sum_{h=1}^{H} \frac{\text{Var}(N_h)}{n} a_{r:K|h} a_{s:K|h} (E(g(X_{[h:H]} - b)))^2 \\ &\quad + \sum_{h=1}^{H}\sum_{h'\neq h}^{H} \frac{\text{cov}(N_h, N_{h'})}{n} E(g(X_{[h:H]} - b)) a_{r:K|h'} E(g(X_{[h':H]} - b)) \\ &= \sum_{h=1}^{H} \frac{H-1}{H^2} a_{r:K|h} a_{s:K|h} (E(g(X_{[h:H]} - b)))^2 \\ &\quad + \sum_{h=1}^{H}\sum_{h'\neq h}^{H} \frac{-1}{H^2} E(g(X_{[h:H]} - b)) a_{r:K|h'} E(g(X_{[h':H]} - b)) \\ &= \frac{1}{H}\sum_{h=1}^{H} a_{r:K|h} a_{s:K|h} (E(g(X_{[h:H]} - b)))^2 - \frac{1}{K^2} E(g(X_{[r:K]} - b)) E(g(X_{[s:K]} - b)).\end{aligned}$$

The second term in the right side of the above expression follows from Eq. (2). Combining expressions A and B, we obtain

$$\sigma_{r,s} = \frac{1}{H}\sum_{h=1}^{H} a_{r:K|h} a_{s:K|h} E g^2(X_{[h:H]} - b) - \frac{1}{K^2} E(g(X_{[r:K]} - b)) E(g(X_{[s:K]} - b)).$$

For $\tau_{r,s}$, $1 \le r \le k$ and $1 \le s \le K-1$ we compute

$$\begin{aligned}
\tau_{r,s} &= \frac{1}{n}\operatorname{cov}\left\{\sum_{i=1}^{n} W_{r,R_i,g}(b), \sum_{i=1}^{n} a_{s:K|R_i}\right\}\\
&= \frac{1}{n}E\left\{\operatorname{cov}\left(\sum_{i=1}^{n} W_{r,R_i,g}(b), \sum_{i=1}^{n} a_{s:K|R_i}\right)|\boldsymbol{R}\right\}\\
&\quad +\frac{1}{n}\operatorname{cov}\left\{E\left(\sum_{i=1}^{n} W_{r,R_i,g}(b)|\boldsymbol{R}\right), E(\sum_{i=1}^{n} a_{s:K|R_i})|\boldsymbol{R}\right\}\\
&= 0+\frac{1}{n}\operatorname{cov}\left\{\sum_{h=1}^{H} N_h W_{r,h,g}(b), \sum_{h=1}^{H} N_h a_{s:K|h})\right\}\\
&= \frac{1}{n}\sum_{h=1}^{H}\sum_{h'=1}^{H}\operatorname{cov}(N_h, N_{h'})a_{r:K|h}a_{s:K|h'}Eg(X_{[h:H]}-b)\\
&= \frac{1}{H}\sum_{h=1}^{H} a_{r:K|h}a_{s:K|h}Eg(X_{[h:H]}-b) - \frac{1}{K^2}Eg(X_{[r:K]}-b).
\end{aligned}$$

Finally, the $\gamma_{r,s}$, $1 \le r, s \le K-1$, follows from

$$\begin{aligned}
\gamma_{r,s} &= \frac{1}{n}\operatorname{cov}\left\{\sum_{i=1}^{n} a_{r:K|R_i}, \sum_{i=1}^{n} a_{s:K|R_i}\right\}\\
&= \frac{1}{n}E\left\{\operatorname{cov}\left(\sum_{i=1}^{n} a_{r:K|R_i}, \sum_{i=1}^{n} a_{s:K|R_i}\right)|\boldsymbol{R}\right\}\\
&\quad +\frac{1}{n}\operatorname{cov}\left\{E\left(\sum_{i=1}^{n} a_{r:K|R_i}\right)|\boldsymbol{R}, E\left(\sum_{i=1}^{n} a_{s:K|R_i}\right)|\boldsymbol{R}\right\}\\
&= 0+\frac{1}{H}\sum_{h=1}^{H} a_{r:K|h}a_{s:K|h} - \frac{1}{K^2}.
\end{aligned}$$

The proof is completed. □

Proof of Theorem 2 Let Q be a transformation from $\bar{\boldsymbol{T}}_g(b)$ to $\hat{\theta}_g(b)$

$$\hat{\theta}_g(b) = \frac{1}{d_n}\sum_{k=1}^{K}\frac{I_k \bar{T}_{g,k}(b)}{\bar{T}_{g,K+k}(b)} \stackrel{D}{=} \frac{1}{K}\sum_{k=1}^{K}\frac{\bar{T}_{g,k}(b)}{\bar{T}_{g,K+k}(b)} = Q(\bar{\boldsymbol{T}}_g(b)),$$

where $\bar{T}_{g,k}(b)$ is the k th component of vector $\bar{\boldsymbol{T}}_g(b)$ and $\bar{T}_{g,2K}(b) = 1 - \sum_{k=1}^{K-1}\bar{T}_{g,K+k}(b)$. The second equality in the above equation follows from the fact that I_k/d_n converges in probability to $1/K$ for $k = 1, \ldots, K$. It is clear that

$Q(E(\bar{\boldsymbol{T}}_g(b))) = E(g(X-b))$. For notational convenience, let $\boldsymbol{\mu}_T = E(\bar{\boldsymbol{T}}_g(b))$. By using a Taylor expansion of $Q(\bar{\boldsymbol{T}}_g(b))$ around $\boldsymbol{\mu}_T$ we write

$$\sqrt{n}\left(Q(\bar{\boldsymbol{T}}_g(b)) - E(g(X-b))\right) = \frac{1}{K}\boldsymbol{L}_T^{\top}\sqrt{n}\left(\bar{\boldsymbol{T}}_g(b) - \boldsymbol{\mu}_T\right) + o_p(1),$$

where $\boldsymbol{L}_T$ is a $2K-1$ dimensional partial derivative vector of $Q(\boldsymbol{\mu}_T)$,

$$\begin{aligned} L_r &= \frac{\mathrm{d}}{\mathrm{d}\mu_T(r)}Q(\boldsymbol{\mu}_T) \\ &= \begin{cases} K & r = 1, \ldots, K \\ K\left(Eg(X_{[K:K]} - b) - Eg(X_{[r-K:K]} - b)\right) & r = K+1, \ldots, 2K-1, \end{cases} \end{aligned}$$

and $\mu_T(r)$ is the r th component of vector μ_T. Let $\boldsymbol{L}^{\top} = (\boldsymbol{L}_1^{\top}, \boldsymbol{L}_2^{\top})$ with $\boldsymbol{L}_1^{\top} = (K, \ldots, K)$ and $\boldsymbol{L}_2 = K(Eg(X_{[K:K]} - b) - Eg(X_{[1:K]} - b), \ldots, Eg(X_{[K:K]} - b) - Eg(X_{[K-1:K]} - b))$. It is then easy to see that $\sqrt{n}\left(Q(\bar{\boldsymbol{T}}_g(b)) - E(g(X-b))\right)$ converges to a normal distribution with mean zero and variance

$$\sigma^2_{\hat{\theta}_g} = \frac{1}{K^2}\boldsymbol{L}_1^{\top}\boldsymbol{\Sigma}_{1,1}\boldsymbol{L}_1 + \frac{1}{K^2}2\boldsymbol{L}_1^{\top}\boldsymbol{\Sigma}_{1,2}\boldsymbol{L}_2 + \frac{1}{K^2}\boldsymbol{L}_2^{\top}\boldsymbol{\Sigma}_{2,2}\boldsymbol{L}_2.$$

By using Lemma 1, we write

$$\begin{aligned} \frac{1}{K^2}\boldsymbol{L}_1^{\top}\boldsymbol{\Sigma}_{1,1}\boldsymbol{L}_1 &= \sum_{r=1}^{K}\sum_{s=1}^{K}\left\{\frac{1}{H}\sum_{h=1}^{H} a_{r:K|h}a_{s:K|h}E\left\{g^2(X_{[h:H]} - b)\right\}\right. \\ &\quad \left. - \frac{E\left\{g(X_{[r:K]} - b)\right\}E\left\{g(X_{[s:K]} - b)\right\}}{K^2}\right\} \\ &= \frac{1}{H}\sum_{h=1}^{H} E\left\{g^2(X_{[h:H]} - b)\right\} - \{Eg(X-b)\}^2 = \sigma_g^2. \end{aligned}$$

Let $d_s = Eg(X_{[K:K]} - b) - Eg(X_{[s:K]} - b)$, $s = 1, \ldots, K-1$.

$$\begin{aligned} \frac{1}{K^2}\boldsymbol{L}_1^{\top}\boldsymbol{\Sigma}_{1,2}\boldsymbol{L}_2 &= \sum_{r=1}^{K}\sum_{s=1}^{K-1} d_s\tau_{r,s} \\ &= \sum_{r=1}^{K}\sum_{s=1}^{K-1} d_s\left[\left\{\frac{1}{H}\sum_{h=1}^{H}\alpha_{r:K|h}\alpha_{s:K|h}Eg(X_{[h:H]} - b)\right\} - \frac{Eg(X_{[r:K]} - b)}{K^2}\right] \\ &= \sum_{s=1}^{K-1} d_s\left\{\frac{Eg(X_{[s:K]} - b)}{K} - \frac{Eg(X-b)}{K}\right\} \\ &= \sum_{s=1}^{K-1}\left\{Eg(X_{[K:K]} - b) - Eg(X_{[s:K]} - b)\right\}\left\{\frac{Eg(X_{[s:K]} - b)}{K} - \frac{Eg(X-b)}{K}\right\} \\ &= -\sum_{s=1}^{K}\left\{Eg(X_{[s:K]} - b) - Eg(X-b)\right\}^2/K. \end{aligned}$$

For the last expression, we consider

$$\begin{aligned}
\frac{1}{K^2}\boldsymbol{L}_2^\top \boldsymbol{\Sigma}_{2,2}\boldsymbol{L}_2 &= \sum_{r=1}^{K-1}\sum_{s=1}^{K-1} d_r d_s \gamma_{r,s} \\
&= \sum_{r=1}^{K-1}\sum_{s=1}^{K-1} \left\{Eg(X_{[K:K]}-b) - Eg(X_{[r:K]}-b)\right\} \\
&\quad \times \left\{Eg(X_{[K:K]}-b) - Eg(X_{[s:K]}-b)\right\}\gamma_{r,s} \\
&= \sum_{r=1}^{K}\sum_{s=1}^{K} \left\{Eg(X_{[K:K]}-b) - Eg(X_{[r:K]}-b)\right\} \\
&\quad \times \left\{Eg(X_{[K:K]}-b) - Eg(X_{[s:K]}-b)\right\}\gamma_{r,s} \\
&= \sum_{r=1}^{K}\sum_{s=1}^{K} \frac{1}{H}\sum_{h=1}^{H} \alpha_{r:K|h}\alpha_{s:K|h} Eg(X_{[r:K]}-b)Eg(X_{[s:K]}-b) \\
&\quad - \{Eg(x-b)\}^2 .
\end{aligned}$$

This completes the proof. □

Proof of Corollary 1 We first rewrite $\sigma^2_{\theta:K}$ as

$$\begin{aligned}
\sigma^2_{\hat{\theta}_{g:K}} &= \sigma_g^2 - \sum_{r=1}^{K}(Eg(X_{[r:K]}-b) - Eg(X-b))^2/K - \left\{\sum_{r=1}^{K}(Eg(X_{[r:K]}-b))^2/K\right. \\
&\quad \left. - \sum_{r=1}^{K}\sum_{s=1}^{K}\sum_{h=1}^{H} \frac{a_{r:K|h}a_{s:K|h}}{H} Eg(X_{[r:K]}-b)Eg(X_{[s:K]}-b)\right\} \\
&= \sigma^2_{\hat{\theta}_{g:RSS(K)}} - \left\{\sum_{r=1}^{K}\frac{(Eg(X_{[r:K]}-b))^2}{K}\right. \\
&\quad \left. - \sum_{r=1}^{K}\sum_{s=1}^{K}\sum_{h=1}^{H} \frac{a_{r:K|h}a_{s:K|h}}{H} Eg(X_{[r:K]}-b)Eg(X_{[s:K]}-b)\right\} \\
&= \sigma^2_{\hat{\theta}_{g:RSS(K)}} - A_d .
\end{aligned}$$

We need to show A_d is non-negative. We first consider

$$B_d = \sum_{r=1}^{K}\sum_{s=1}^{K}\frac{1}{H}\sum_{h=1}^{H} a_{r:K|h}a_{s:K|h}\left\{E\left(g(X_{[r:K]}-b)\right) - E\left(g(X_{[r:K]}-b)\right)\right\}^2 \geq 0.$$

We re-write expression B_d as

$$B_d = 2\sum_{r=1}^{K}\sum_{s=1}^{K}\frac{1}{H}\sum_{h=1}^{H} a_{r:K|h}a_{s:K|h}\left\{E\left(g(X_{[r:K]}-b)\right)\right\}^2 \\ -2\sum_{r=1}^{K}\sum_{s=1}^{K}\frac{1}{H}\sum_{h=1}^{H} a_{r:K|h}a_{s:K|h}\left\{E\left(g(X_{[r:K]}-b)\right)\right\}\left\{E\left(g(X_{[r:K]}-b)\right)\right\}.$$

Use of equalities $\sum_{s=1}^{K} a_{s:K|h} = 1$ and $\sum_{h=1}^{H} a_{s:K|h} = H/K$ in the first term of the above equation reduces B_d to $2A_d$

$$B_d = \frac{2}{K}\sum_{r=1}^{K}\left\{E\left(g(X_{[r:K]}-b)\right)\right\}^2 \\ -2\sum_{r=1}^{K}\sum_{s=1}^{K}\frac{1}{H}\sum_{h=1}^{H} a_{r:K|h}a_{s:K|h}\left\{E\left(g(X_{[r:K]}-b)\right)\right\}\left\{E\left(g(X_{[r:K]}-b)\right)\right\} \\ = 2A_d \geq 0.$$

This completes the proof. □

Proof of Theorem 3 Let $U_n(b/\sqrt{n}) = \left\{S_n(\eta_p + b/\sqrt{n}) - S_n(\eta_p))/(b/\sqrt{n}\right\}$. By using conditional expectation given rank vector $\boldsymbol{R}$, we obtain

$$E(U_n(b/\sqrt{n}) = \sum_{h=1}^{H} E\left\{\sum_{k=1}^{K}\frac{I_k N_h a_{k:K|h}}{d_n \sum_{t=1}^{H} N_t a_{k:K|t}}\right\}\frac{F_{[h]}(\eta_p) - F_{[h]}(\eta_p + b/\sqrt{n})}{b/\sqrt{n}}.$$

For large n, we can write

$$\lim_{n\to\infty} E\left\{\frac{I_k N_h a_{k:K|h}}{d_n \sum_{t=1}^{H} N_t a_{k:K|t}}\right\} = E\left\{\lim_{n\to\infty}\frac{I_k}{d_n}\lim_{n\to\infty}\frac{N_h a_{k:K|h}}{d_n \sum_{t=1}^{H} N_t a_{k:K|t}}\right\} \\ = \frac{1}{K}\frac{a_{k:K|h}/H}{\sum_{t=1}^{H} a_{k:K|t}/H} = a_{k:K|h}/H.$$

It is now easy to observe that $E(U_n(b/\sqrt{n})$ has a limit at $-bf(\eta_p)$

$$\lim_{n\to\infty} E(U_n(b/\sqrt{n}) = \sum_{h=1}^{H}\sum_{k=1}^{K}\frac{a_{k:K|h}}{H}\lim_{n\to\infty}\frac{F_{[h]}(\eta_p) - F_{[h]}(\eta_p + b/\sqrt{n})}{b/\sqrt{n}} \\ = -b\sum_{k=1}^{K}\sum_{h=1}^{H}\frac{a_{k:K|h}}{H} f_{[h]}(\eta_p) = -b\sum_{k=1}^{K}\frac{H}{KH} f(\eta_p) = -bf(\eta_p).$$

We now show that the variance of $U_n(b/\sqrt{n})$ converges to zero as n goes to infinity. Without loss of generality, assume that $b > 0$ and $\eta_p = 0$. In this case, $U_n(b/\sqrt{n})$ can be written as

$$U_n(b/\sqrt{n}) = \sum_{i=1}^{n}\sum_{k=1}^{K} \frac{I_k a_{k:K|R_i}}{d_n \sum_{j=1}^{n} a_{k:K|R_j}} I(0 \le X_i \le b/\sqrt{n}).$$

The variance of $U_n(b/\sqrt{n})$ can be written as

$$\mathrm{Var}(U_n(b/\sqrt{n})) = \mathrm{Var}\left(E\left\{U_n(b/\sqrt{n})|\boldsymbol{R}\right\}\right) + E\left(\mathrm{Var}\left\{U_n(b/\sqrt{n})|\boldsymbol{R}\right\}\right) = D_n + G_n.$$

The expression D_n is given by

$$\begin{aligned} D_n = \frac{1}{b^2}\sum_{h=1}^{H}\sum_{h'=1}^{H} &\left(F_{[h]}(b/\sqrt{n}) - F_{[h]}(0)\right)\left(F_{[h']}(b/\sqrt{n}) - F_{[h']}(0)\right) \\ &\times \sum_{k=1}^{K}\sum_{k'=1}^{K} \mathrm{Cov}\left(\frac{\sqrt{n} I_k N_h a_{k:K|h}}{d_n \sum_{t=1}^{H} N_t a_{k:K|t}}, \frac{\sqrt{n} I_{k'} N_{h'} a_{k':K|h'}}{d_n \sum_{t=1}^{H} N_t a_{k':K|t}}\right). \end{aligned}$$

In the above expression, the covariances are all finite and the double sum in the first term converges to zero as n gets large. Hence, D_n has a limit at 0. In a similar fashion, G_n can be written as

$$\begin{aligned} G_n = \sum_{k=1}^{K}\sum_{h=1}^{H} E &\left\{\frac{I_k^2 N_h a_{k:K|h}^2}{d_n^2 (\sum_{t=1}^{H} N_t a_{k:K|t})^2}\right\}\left(F_{[h']}(b/\sqrt{n}) - F_{[h']}(0)\right) \\ &\times \left\{1 - \left(F_{[h']}(b/\sqrt{n}) + F_{[h']}(0)\right)\right\}. \end{aligned}$$

Since the expectation in the above expression has a finite limit, G_n converges to zero as n goes to infinity. We then conclude that variance $U_n(b/\sqrt{n})$ goes to zero as n approaches to infinity. This establishes the point-wise convergence

$$\sqrt{n} S_n(\eta_p + b/\sqrt{n}) = \sqrt{n} S_n(\eta_p) - b f(\eta_p) + o_p(1).$$

The uniform convergence follows from the monotonicity of $S_n(b)$. □

Proof of Theorem 4 We first observe that sample size vector $\boldsymbol{N}^\top = (N_1, \ldots, N_H)$ has a multinomial distribution with parameters n and $(1/H, \ldots, 1/H)$. The probability that $W_n(q)$ equals K is

$$P(W_n(q) = K) = \binom{H}{K} P(N_1 > q, \ldots, N_K > q, N_{K+1} \le q, \ldots, N_H \le q).$$

Let

$$A_{K,q} = \{N_1 > q, \ldots, N_K > q\} \text{ and } B_{H-K,q} = \{N_{K+1} \le q, \ldots, N_H \le q\}.$$

For a fixed j, we also define $B_{j,H-K,q}$ to be the event that each one of the j judgment classes in set $B_{H-K,q}$ has exactly q observations. Note that by the definition of the set $B_{H-K,q}$, $B_{j,H-K,q} = B_{H-K,H-K,q}$ when $q = 0$. With this new notation we write

$$P((W_n(q) = K) = \binom{H}{K} \sum_{j=0}^{q(H-K)} \binom{q(H-K)}{j} P(A_{K,q}|B_{j,H-K,q})P(B_{j,H-K,q}).$$

It is clear that

$$P(B_{j,H-K,q}) = \frac{n!}{(n-qj)!} \left(\frac{K}{H}\right)^{n-qj} \left(\frac{1}{H}\right)^{qj}$$

and

$$P(A_{K,q}|B_{j,H-K,q}) = \sum_{n_1>q,\ldots,n_K>q} \binom{n-qj}{n_1,\ldots,n_K} (1/K)^{n-qj}.$$

Let

$$A_{K,q} = \{A_{1:K,q} \cap \cdots \cap A_{K:K,q}\},$$

where $A_{h:K,q}$ is the event that $N_h > q$ for $h \le K$. By using DeMorgan's law, we write

$$\begin{aligned} P(A_{K,q}|B_{j,H-K,q}) &= 1 - P\left(A^C_{K,q}|B_{j,H-K,q}\right) = 1 - P\left(\cup_{h=1}^K A^C_{h:K,q}|B_{j,H-K,q}\right) \\ &= 1 - \sum_{i=1}^K \binom{K}{i} (-1)^{i-1} P\left(A^C_{1:K,q} \cap \cdots \cap A^C_{i:K,q}|B_{j,H-K,q}\right). \end{aligned} \tag{11}$$

We now evaluate the following conditional probability

$$\begin{aligned} P\left(A^C_{1:K,q} \cap \cdots \cap A^C_{i:K,q}|B_{j,H-K,q}\right) = \sum_{r_1=0}^q \cdots \sum_{r_i=0}^q &\binom{n-qj}{r_1,\ldots,r_i,n-qj-T_i} \\ &\times (1/K)^{T_i} \{1-i/K\}^{(n-qj-T_i)}, \end{aligned}$$

where $T_i = \sum_{y=1}^i r_y$. With some algebra, the above equation simplifies to

$$P\left(A^C_{1:K,q} \cap \cdots \cap A^C_{i:K,q}|B_{j,H-K,q}\right) = \sum_{s=0}^{iq} \binom{qi}{s} \binom{n-qj}{s} \frac{s!(K-i)^{n-qj-s}}{K^{n-qj}}.$$

We complete the proof by putting this expression in (11). □

References

Chen, M., Ahn, S., Wang, X., Lim, J. (2014). Generalized isotonized mean estimators for judgment post-stratification with multiple rankers. *Journal of Agricultural, Biological, and Environmental Statistics*, *19*, 405–418.

Dastbaravarde, A., Arghami, N. R., Sarmad, M. (2016). Some theoretical results concerning non-prametric estimation by using a judsgment post-stratification sample with perfect ranking. *Communications in Statistics, Theory and Methods*, *45*, 2182–2203.

Dell, T. R., Clutter, J. L. (1972). Ranked set sampling theory with order statistics backround. *Biometrics*, *28*, 545–555.

Du, J., MacEachern, S. N. (2008). Judgement post-stratification for designed experiments. *Biometrics*, *64*, 345–354.

Frey, J. (2012). Constrained nonparametric estimation of the mean and the CDF using ranked-set sampling with a covariate. *Annals of the Institute of Statistical Mathematics*, *64*, 439–456.

Frey, J., Feeman, T. G. (2012). An improved mean estimator for judgment post-stratification. *Computational Statistics and Data Analysis*, *56*, 418–426.

Frey, J., Feeman, T. G. (2013). Variance estimation using judgment post-stratification. *Annals of the Institute of Statistical Mathematics*, *65*, 551–569.

Frey, J., Ozturk, O. (2011). Constrained estimation using judgment post-stratification. *Annals of the Institute of Statistical Mathematics*, *63*, 769–789.

Hollander, M., Wolfe, D. A., Chicken, E. (2014). *Nonparametric Statistical Methods* (3rd ed.). New York: Wiley.

MacEachern, S. N., Stasny, E. A., Wolfe, D. A. (2004). Judgment post-stratification with imprecise rankings. *Biometrics*, *60*, 207–215.

Ozturk, O. (2012). Quantile inference based on partially rank-ordered set samples. *Journal of Statistical Planning and Inference*, *142*, 2116–2127.

Ozturk, O. (2013). Combining multi-observer information in partially rank-ordered judgment post-stratified and ranked set samples. *The Canadian Journal of Statistics*, *41*, 304–324.

Ozturk, O. (2014). Statistical inference for population quantiles and variance in judgment post-stratified samples. *Computational Statistics and Data Analysis*, *77*, 188–205.

Ozturk, O. (2015). Distribution free two-sample methods for judgment post-stratified data. *Statistica Sinica*, *25*, 1691–1712.

Stokes, S. L., Wang, X., Chen, M. (2007). Judgment post-stratification with multiple rankers. *Journal of Statistical Theory and Applications*, *6*, 344–359.

Wang, X., Stokes, S. L., Lim, J., Chen, M. (2006). Concomitants of multivariate order statistics with application to judgment post-stratification. *Journal of the American Statistical Association*, *101*, 1693–1704.

Wang, X., Lim, J., Stokes, S. L. (2008). A nonparametric mean estimator for judgment post-stratified data. *Biometrics*, *64*, 355–363.

Wang, X., Wang, K., Lim, J. (2012). Isotonized CDF estimation from judgment post-stratification data with empty strata. *Biometrics*, *68*, 194–202.

Wolfe, D. A. (2012). Ranked set sampling: its relevance and impact on statistical inference. *ISRN Probability and Statistics*. doi:10.5402/2012/568385.