Generalized Isotonized Mean Estimators for Judgment Post-stratification with Multiple Rankers

Abstract

In this paper, we propose a new set of mean estimators for judgment post-stratified data with multiple rankers. The new estimators take into account matrix partial ordering in cumulative distribution functions of rank strata, and they are derived by improving existing estimators through employing the order constraints and solving a generalized isotonic regression problem. Numerical studies show that the proposed isotonized mean estimators outperform the existing estimators. Finally, the proposed estimators are applied to estimating the average tree height using the tree data in Chen et al. (2006).

Key Words: best linear unbiased estimator; generalized isotonic regression; matrix partial order; simple stochastic order; ranked set sampling; Raking; relative efficiency.

1 Introduction

Both ranked set sampling (RSS) and judgment post-stratification (JPS) are established cost-effective methods of data collection, and are useful in situations when the characteristic of interest Y is expensive to measure but sampling units can be easily gathered and ranked by some means without exact measurements. Both utilize the assigned ranks to provide auxiliary information on the measured units. They can provide improved estimators of the population mean, variance and distribution functions over simple random samples (SRS) of the same size. However, unlike RSS, JPS is based on a simple random sample (SRS), in which the actually measured units are post-stratified on ranks, and so the number of measured units n_h in each rank stratum h (in the one-ranker case), $h = 1, \dots, H$, is a random variable, where H is a pre-determined set size, and $(n_1, \dots, n_H)$ jointly follows the multinomial distribution with parameters $\big(n, 1/H, ..., 1/H\big)$. JPS has several advantages over RSS as explained in Wang et al. (2012). For example, JPS can easily allow more than one ranker to provide ranking information about the same measured unit, which can further improve the estimation efficiency when rankers have some ranking skill and are not identical.

JPS experiments with multiple rankers are first introduced by MacEachern et al. (2004). For example, suppose we are interested in estimating the mean volume of trees in a forest so that Y denotes the tree volume. The JPS data with m rankers are constructed as follows. First, a simple random sample of H trees is selected from the forest. Among the H trees, one tree is randomly selected and marked for later actual measurement; and each ranker assesses the rank of this tree in comparison to the $(H-1)$ trees based on visual inspection of their volumes, and assigns a rank value among $\{1, \cdots, H\}$ to the tree. Here, ranking errors could occur so that rankers do not necessarily agree on the assigned ranks. This step is repeated for an additional $(n-1)$ times until n trees in total are marked for later measurements, along with their ranks assigned by the m rankers among their own comparison groups. Next, each of the n trees is cut down, dragged to a mill and cut into logs for measuring its volume. Suppose Y is absolutely continuous with population mean μ and finite variance σ^2. Let $O_{i,k}$ denote the judgment order assigned by ranker k for the ith measured tree, where $O_{i,k} \in \{1, \cdots, H\}$ for $i = 1, \dots, n$ and $k = 1, \dots, m$. The JPS sample with m rankers consists of data in the form of $\big\{Y_i, \mathbf{O}_i\big\}_{i=1}^{n}$ where $\mathbf{O}_i = \big(O_{i,1}, \dots, O_{i,m}\big)$. The n measured units form a simple random sample from the population; and they fall into H^m post-strata formed by the orders. Let $Y_{[\mathbf{r}]}$ denote $Y|\mathbf{O} = \mathbf{r}$, any observation falling in the $\mathbf{r}$-th post-stratum, where $\mathbf{r} = (r_1, \dots, r_m)$ and each $r_k \in \{1, \cdots, H\}$. Let $n_{\mathbf{r}}$, $\overline{Y}_{[\mathbf{r}]}$, $\mu_{[\mathbf{r}]}$, and $\sigma^2_{[\mathbf{r}]}$ denote the sample size, sample mean, population mean and variance in the $\mathbf{r}$-th stratum, respectively.

In the literature several nonparametric mean estimators have been proposed for JPS with multiple rankers. MacEachern et al. (2004), who first introduced JPS, proposed a mean estimator that prorates each measured value Y_i to the original H judgment strata according to the proportion of the rankers that assign the value to stratum h, $h = 1, \dots, H$. Stokes et al. (2007) considered two different estimators to utilize the information from multiple rankers. The first estimator is a weighted average of sample means of the H^m post-strata jointly formed by the m rankers, where each sample mean $\overline{Y}_{[\mathbf{r}]}$ is weighted by the estimated probability from the raking method that a randomly selected observation falls in the corresponding $\mathbf{r}$-th post-stratum. The second is the Best Linear Unbiased Estimator (BLUE) in the form of a linear combination of JPS mean estimators based on each single ranker, whose weights are set to minimize the variance of the linear estimator.

In both RSS and JPS (with one ranker only), researchers often exploit distributional properties of $Y_{[r]}$s to improve the estimation efficiency (Ozturk, 2007; Frey and Ozturk, 2011; Wang et al., 2008). An example is stochastic ordering among cumulative distribution functions (CDF) of different rank strata. For every

$y \in \mathcal{R}$,

$$F_{(1)}(y) \geq F_{(2)}(y) \geq \cdots \geq F_{(H)}(y),$$

where $F_{(h)}(y)$ is the CDF of the h-th order statistic. Further, in the presence of imperfect ranking, it is often reasonable to assume

$$F_{[1]}(y) \geq F_{[2]}(y) \geq \cdots \geq F_{[H]}(y),$$

where $F_{[h]}(y)$ is the CDF of the h-th judgment order statistic (Wang et al. 2008). By considering the above constraints, Ozturk (2007) proposed to modify the empirical CDF estimator by minimizing the Cramer-von Mises distance. The new estimator has been shown to have smaller integrated Mean Squared Error (MSE) than the empirical one. Wang et al. (2008) also introduced an isotonized estimator of the population mean for JPS data, which improves estimation efficiency of the conventional JPS mean estimator.

In this paper, we consider the monotonicity in CDFs of rank strata formed by multiple rankers, and propose a new set of estimators by monotonizing $\overline{Y}_{[\mathbf{r}]}$s using generalized isotonic regression. In the cases of multiple rankers, it is reasonable to assume that $F_{[\mathbf{r}]}$, the CDF of the $\mathbf{r}$-th judgment stratum, follows a matrix partial order; that is, if $\mathbf{r}_1 = \big(r_{11}, \ldots, r_{1m}\big)$, $\mathbf{r}_2 = \big(r_{21}, \ldots, r_{2m}\big)$, and $\mathbf{r}_1 \leq \mathbf{r}_2$ in a component-wise manner (i.e., $r_{1k} \leq r_{2k}$ for all $k = 1, \ldots m$), then

$$F_{[\mathbf{r}_1]}(y) \geq F_{[\mathbf{r}_2]}(y) \tag{1}$$

for every y. The order (1) further induces an order in the corresponding stratum means

$$\mu_{[\mathbf{r}_1]} \leq \mu_{[\mathbf{r}_2]}. \tag{2}$$

Here we propose to modify the sample means $\overline{Y}_{[\mathbf{r}]}$ by solving the generalized isotonic regression problem under the matrix partial order (2). Finally, we develop a set of new estimators based on existing estimators for JPS with multiple rankers, where we use isotonized mean estimators $\widehat{\mu}^*_{[\mathbf{r}]}$ to replace the sample means $\overline{Y}_{[\mathbf{r}]}$.

This paper is organized as follows. In Section 2, we review existing mean estimators for JPS data with multiple rankers. In Section 3, we propose a new set of estimators that incorporate the matrix partial ordering among stratum means into estimation. In Section 4, we numerically compare the performance of the new set of isotonized estimators to their non-isotonic counterparts. In Section 5, we apply the new estimators to estimating the mean of tree height data reported in Chen et al. (2006). In Section 6, we conclude the paper with a brief discussion.

2 Existing JPS Mean Estimators with Multiple Rankers

Recall that JPS data form a post-stratified sample using the auxiliary information from multiple rankers. Thus, the population mean has the following form:

$$\mu = \sum_{\mathbf{r}} \pi_{\mathbf{r}} \mu_{[\mathbf{r}]}, \tag{3}$$

where $\pi_{\mathbf{r}} \equiv P(\mathbf{O} = \mathbf{r})$ denotes the probability that an observation falls in the $\mathbf{r}$-th post-stratum. If we have a single ranker (i.e., $m = 1$), the probability $\pi_{\mathbf{r}}$ is known to be $1/H$. However, in the case of multiple rankers, unless assumptions are made about both the distribution of Y and the ranking process, $\pi_{\mathbf{r}}$s are known only up to their one-dimensional marginal probabilities, which are all equal to $1/H$. They should be estimated or pre-decided, which makes the problem more complicated. To estimate the population mean from JPS data with multiple rankers, we may first consider the nonparametric estimator proposed by MacEachern et al. (2004),

$$\widehat{\mu}^{\text{msw}} = \frac{1}{H}\sum_{h=1}^{H} \frac{\sum_{k=1}^{m}\sum_{i=1}^{n} \mathrm{I}\big(O_{i,k}=h\big)Y_i}{\sum_{k=1}^{m}\sum_{i=1}^{n} \mathrm{I}\big(O_{i,k}=h\big)}, \tag{4}$$

where $\mathrm{I}(\,\cdot\,)$ is the indicator function. The estimator can be rewritten in the form of (3),

$$\widehat{\mu}^{\text{msw}} = \sum_{\mathbf{r}} \widehat{\pi}_{\mathbf{r}}^{\text{msw}} \overline{Y}_{[\mathbf{r}]}, \tag{5}$$

where

$$\widehat{\pi}_{\mathbf{r}}^{\text{msw}} = \frac{1}{H}\sum_{h=1}^{H} \frac{S_h(\mathbf{r})n_{\mathbf{r}}}{\sum_{\mathbf{r}'} S_h(\mathbf{r}')n_{\mathbf{r}'}}$$

and $S_h(\mathbf{r})$ is the count of rank h in the row vector $\mathbf{r}$. If there are only two rankers (i.e., $m=2$), then for every $s,t \in \{1,\ldots,H\}$,

$$\widehat{\pi}_{s,t}^{\text{msw}} = \frac{1}{H}\left(\frac{n_{s,t}}{n_{s.}+n_{.s}} + \frac{n_{s,t}}{n_{t.}+n_{.t}}\right),$$

where $n_{s.}$(or $n_{.t}$) denotes the number of sample units for which $O_{i,1}=s$ (or $O_{i,2}=t$). Note that $\widehat{\pi}_{\mathbf{r}}^{\text{msw}}=0$ for empty cells $\mathbf{r}$ whenever $n_{\mathbf{r}}=0$ ($0/0$ is defined as 0).

The estimator by MacEachern et al. (2004) inherently assumes that all rankers are equally effective by assigning equal weight to each of them. However, this is often restrictive in practice. More recently, Stokes et al. (2007) considered two estimators, the Raking estimator and the Best Linear Unbiased Estimator (BLUE). The Raking estimator, denoted by $\widehat{\mu}^{\text{rake}}$, estimates $\mu_{[\mathbf{r}]}$ with $\bar{Y}_{[\mathbf{r}]}$, and estimates $\pi_{\mathbf{r}}$ with the maximum likelihood estimator (MLE) under marginal probability constraints $\sum_{\mathbf{r}:r_k=h}\pi_{\mathbf{r}}=1/H$ for every $k=1,\ldots,m$ and $h=1,\ldots,H$. The constrained MLE, $\widehat{\pi}_{\mathbf{r}}^{\text{rake}}$, is computed using an Iterative Proportional Fitting procedure, called "raking" (Pelz and Good, 1986). Thus, $\widehat{\mu}^{\text{rake}}$ is given by

$$\widehat{\mu}^{\text{rake}} = \sum_{\mathbf{r}} \widehat{\pi}_{\mathbf{r}}^{\text{rake}} \bar{Y}_{[\mathbf{r}]}. \tag{6}$$

Note that for empty cells, $\widehat{\pi}_{\mathbf{r}}^{\text{rake}}$ is set to zero unless $0/0$ occurs when the raking procedure cannot proceed and returns "NA". When there exists a large portion of empty cells, $\widehat{\pi}_{\mathbf{r}}^{\text{rake}}$ often performs poorly (Thompson, 1981).

The BLUE linearly combines the JPS mean estimators from individual rankers by considering the class of linear estimators in the form of $\sum_{k=1}^{m} w_k\widehat{\mu}_k$, where the weights $w_k \geq 0$ for every k, satisfying $\sum_{k=1}^{m} w_k = 1$; and $\widehat{\mu}_k$ is the JPS mean estimator using only the rank information from the $k-$th ranker,

$$\widehat{\mu}_k = \frac{1}{H}\sum_{h=1}^{H} \frac{\sum_{i=1}^{n} \mathrm{I}\big(O_{i,k}=h\big)Y_i}{\sum_{i=1}^{n} \mathrm{I}\big(O_{i,k}=h\big)}.$$

Then the BLUE finds the weights w_k^*s that minimize the variance of $\sum_{k=1}^{m} w_k\widehat{\mu}_k$, which requires the covariance matrix of $(\hat{\mu}_1,\ldots,\hat{\mu}_m)$. Stokes et al. (2007) estimated the covariance matrix using a bootstrapping method. The BLUE estimator, $\widehat{\mu}^{\text{blue}}$, can also be written in the form of (3):

$$\widehat{\mu}^{\text{blue}} = \sum_{\mathbf{r}} \widehat{\pi}_{\mathbf{r}}^{\text{blue}} \overline{Y}_{[\mathbf{r}]}, \tag{7}$$

with

$$\widehat{\pi}_{\mathbf{r}}^{\text{blue}} = \frac{1}{H} \sum_{h=1}^{H} \sum_{k=1}^{m} w_k^* \frac{\mathrm{I}(r_k = h) n_{\mathbf{r}}}{\sum_{\mathbf{r}'} \mathrm{I}(r_k' = h) n_{\mathbf{r}'}},$$

where r_k is the $k-$th component of $\mathbf{r}$. When there are empty cells, $\widehat{\pi}_{\mathbf{r}}^{\text{blue}} = 0$ where $n_{\mathbf{r}} = 0$ (0/0 is defined as 0). Similarly, for the case with two rankers,

$$\widehat{\mu}^{\text{blue}} = \sum_{s=1}^{H} \sum_{t=1}^{H} \widehat{\pi}_{s,t} \overline{Y}_{[s,t]},$$

with

$$\widehat{\pi}_{s,t} = \frac{1}{H} \left(w_1^* \frac{n_{s,t}}{n_{s.}} + w_2^* \frac{n_{s,t}}{n_{.t}} \right).$$

3 Generalized Isotonized Mean Estimators

In this section, we propose a new set of mean estimators for JPS data with multiple rankers. Wang et al. (2008) considered the monotonicity among means of judgment strata,

$$\mu_{[1]} \leq \cdots \leq \mu_{[H]},$$

and proposed an isotonized mean estimator

$$\tilde{\mu} = \frac{1}{H} \sum_{h=1}^{H} \tilde{\mu}_{[h]},$$

where $\tilde{\mu}_{[h]}$ is the solution to the isotonic regression problem:

$$\begin{array}{ll} \text{minimize} & \sum_{h=1}^{H} n_h \left(\overline{Y}_{[h]} - \mu_{[h]} \right)^2 \\ \text{subject to} & \mu_{[1]} \leq \cdots \leq \mu_{[H]}. \end{array} \tag{8}$$

They numerically showed that the isotonized mean estimator performed better than the original JPS estimator (5).

For JPS with multiple rankers, we consider the matrix partial orders among $\mu_{[\mathbf{r}]}$s. To be specific, we assume that $F_{[\mathbf{r}]}$s are stochastically ordered in the sense that, if $\mathbf{r}_1 \leq \mathbf{r}_2$ in a component-wise manner, then for any $y \in \mathcal{R}$, $F_{[\mathbf{r}_1]}(y) \geq F_{[\mathbf{r}_2]}(y)$. The stochastic order in stratum CDFs induces the built-in ordering in the means of the strata; that is, if $\mathbf{r}_1 \leq \mathbf{r}_2$, then $\mu_{[\mathbf{r}_1]} \leq \mu_{[\mathbf{r}_2]}$. However, the sample means $\overline{Y}_{[\mathbf{r}]}$s, the most common empirical estimator of $\mu_{[\mathbf{r}]}$, often do not satisfy the order constraints (2). We propose to modify the sample means $\overline{Y}_{[\mathbf{r}]}$ by solving the following generalized isotonic regression problem:

$$\begin{array}{ll} \text{minimize} & \sum_{\mathbf{r}} n_{\mathbf{r}} \left(\overline{Y}_{[\mathbf{r}]} - \mu_{[\mathbf{r}]} \right)^2 \\ \text{subject to} & \mu_{[\mathbf{r}_1]} \leq \mu_{[\mathbf{r}_2]}, \quad \text{if} \quad \mathbf{r}_1 \leq \mathbf{r}_2. \end{array} \tag{9}$$

We finally suggest a set of new estimators using the isotonized mean estimators $\widehat{\mu}^*_{[\mathbf{r}]}$, which are the solution to the above optimization problem, for $\overline{Y}_{[\mathbf{r}]}$ in (5), (6), and (7).

As in JPS with a single ranker, we have that $\widehat{\mu}^*_{[\mathbf{r}]}$ is a strongly consistent estimator of $\mu_{[\mathbf{r}]}$, if $\overline{Y}_{[\mathbf{r}]}$ is strongly consistent for $\mu_{[\mathbf{r}]}$ (Barlow et al., 1972). Let

$$\tilde{\mu} = \sum_{\mathbf{r}} \widehat{\pi}_{\mathbf{r}} \widehat{\mu}^*_{[\mathbf{r}]}$$

be an isotonized mean estimator. We note that

$$\begin{aligned} |\tilde{\mu} - \mu| &= \left| \sum_{\mathbf{r}} \widehat{\pi}_{[\mathbf{r}]} \left(\widehat{\mu}^*_{[\mathbf{r}]} - \mu_{[\mathbf{r}]} \right) + \sum_{\mathbf{r}} \left(\widehat{\pi}_{[\mathbf{r}]} - \pi_{[\mathbf{r}]} \right) \mu_{[\mathbf{r}]} \right| \\ &\leq \sum_{\mathbf{r}} \widehat{\pi}_{[\mathbf{r}]} \left| \widehat{\mu}^*_{[\mathbf{r}]} - \mu_{[\mathbf{r}]} \right| + \sum_{\mathbf{r}} \left| \widehat{\pi}_{[\mathbf{r}]} - \pi_{[\mathbf{r}]} \right| \mu_{[\mathbf{r}]}. \end{aligned} \tag{10}$$

Based on Slutsky's Theorem, we can conclude from (10) that $\tilde{\mu}$ is also consistent when $\widehat{\pi}_{[\mathbf{r}]}$ is a consistent estimator of $\pi_{\mathbf{r}}$.

We now explain the algorithm to solve (9). There are several efficient algorithms in the literature (Block et al., 1994; Qian and Eddy, 1996; Sysoev et al., 2011). We introduce the iterative procedure by Block et al. (1994). Below we assume two rankers for notation simplicity. The extension to more than two rankers is straightforward. For the case of two rankers, the problem becomes

$$\begin{array}{ll} \text{minimize} & \sum_{s,t} n_{[s,t]} \left(\overline{Y}_{[s,t]} - \mu_{[s,t]} \right)^2 \\ \text{subject to} & \mu_{[1,t]} \leq \mu_{[2,t]} \leq \cdots \leq \mu_{[H,t]} \\ & \mu_{[s,1]} \leq \mu_{[s,2]} \leq \cdots \leq \mu_{[s,H]}, \end{array}$$

for $s, t \in \{1, \dots, H\}$. The algorithm is outlined by the following steps.

- **(S1)** Let $\widehat{\mathbf{Y}}^{(1)} = \left(\widehat{Y}^{(1)}_{[s,t]},\ s,t = 1, \dots, H \right)$ be the solution to

$$\begin{array}{ll} \text{minimize} & \sum_{s,t} n_{[s,t]} \left(\overline{Y}_{[s,t]} - \mu_{[s,t]} \right)^2 \\ \text{subject to} & \mu_{[1,t]} \leq \mu_{[2,t]} \leq \cdots \leq \mu_{[H,t]}, \quad t = 1, \dots, H \end{array}$$

  We define the row increments $\mathbf{R}^{(1)} = \left(R^{(1)}_{[s,t]},\ s,t = 1, \dots, H \right)$ where

$$R^{(1)}_{[s,t]} = \widehat{Y}^{(1)}_{[s,t]} - \overline{Y}_{[s,t]}.$$

- **(S2)** Let $\widetilde{\mathbf{Y}}^{(1)} = \left(\widetilde{Y}^{(1)}_{[s,t]},\ s,t = 1, \dots, H \right)$ be the solution to

$$\begin{array}{ll} \text{minimize} & \sum_{s,t} n_{[s,t]} \left(\bar{Y}_{[s,t]} + R^{(1)}_{[s,t]} - \mu_{[s,t]} \right)^2 \\ \text{subject to} & \mu_{[s,1]} \leq \mu_{[s,2]} \leq \cdots \leq \mu_{[s,H]}, \quad s = 1, \dots, H. \end{array}$$

  We define the column increments $\mathbf{C}^{(1)} = \left(C^{(1)}_{[s,t]},\ s,t = 1, \dots, H \right)$ where

$$C^{(1)}_{[s,t]} = \widetilde{Y}^{(1)}_{[s,t]} - \left[\overline{Y}_{[s,t]} + R^{(1)}_{[s,t]} \right].$$

- **(S3)** For $k \geq 2$, we obtain $\widehat{\mathbf{Y}}^{(k)}$ by solving the isotonic regression

$$\begin{array}{ll} \text{minimize} & \sum_{s,t} n_{[s,t]} \left[\overline{Y}_{[s,t]} + C_{[s,t]}^{(k-1)} - \mu_{[s,t]} \right]^2 \\ \text{subject to} & \mu_{[1,t]} \leq \mu_{[2,t]} \leq \cdots \leq \mu_{[H,t]}, \quad t = 1, \ldots, H, \end{array}$$

and define the $k-$th row increments $\mathbf{R}^{(k)} = \left(R_{[s,t]}^{(k)},\ s, t = 1, \ldots, H \right)$ where

$$R_{[s,t]}^{(k)} = \widehat{Y}_{[s,t]}^{(k)} - \left[\overline{Y}_{[s,t]} + C_{[s,t]}^{(k-1)} \right].$$

Then, we solve the generalized isotonic regression on $\widehat{\mathbf{Y}}^{(k)}$ over columns, that is

$$\begin{array}{ll} \text{minimize} & \sum_{s,t} n_{[s,t]} \left(\bar{Y}_{[s,t]} + R_{[s,t]}^{(k)} - \mu_{(s,t)} \right)^2 \\ \text{subject to} & \mu_{[s,1]} \leq \mu_{[s,2]} \leq \cdots \leq \mu_{[s,H]}, \quad s = 1, \ldots, H. \end{array}$$

Let $\widetilde{\mathbf{Y}}^{(k)} = \left(\widetilde{Y}_{[s,t]}^{(k)},\ s, t = 1, \ldots, H \right)$ be the solution to the above problem. We define the $k-$th column increments $\mathbf{C}^{(k)} = \left(C_{[s,t]}^{(k)},\ s, t = 1, \ldots, H \right)$ where

$$C_{[s,t]}^{(k)} = \widetilde{Y}_{[s,t]}^{(k)} - \left[\overline{Y}_{[s,t]} + R_{[s,t]}^{(k)} \right].$$

- **(S4)** We iterate **(S3)** until the row and the column increments are smaller than a given error bound.

4 Simulation Studies

In this section, we study the behavior of the isotonized mean estimators discussed in section 3. We use the same imperfect ranking model as in Stokes et al. (2007). In the study, we consider the case of two rankers ($m = 2$) with six scenarios. The first three scenarios assume that rankers are equally effective, and the last three ones assume rankers are not equally effective. We set H to be $\{2, 4, 10\}$ and n to be $\{10, 30, 60, 150\}$. Let Y be the characteristic of interest, which is generated from four types of distributions: normal, uniform, lognormal, and exponential (see Table 1 for the parameter settings of the simulation).

For each simulated JPS data set, we compute six estimators:

- $\widehat{\mu}^{\text{msw}}$ (MSW): the estimator by MacEachern et al. (2004).
- $\widehat{\mu}^{*\text{msw}}$ (Iso-MSW): the isotonized version of $\widehat{\mu}^{\text{msw}}$.
- $\widehat{\mu}^{\text{rake}}$ (Raking): the Raking estimator by Stokes et al. (2007).
- $\widehat{\mu}^{*\text{rake}}$ (Iso-Raking): the isotonized version of $\widehat{\mu}^{\text{rake}}$.
- $\widehat{\mu}^{\text{blue}}$ (BLUE): the BLUE by Stokes et al. (2007).
- $\widehat{\mu}^{*\text{blue}}$ (Iso-BLUE): the isotonized version of $\widehat{\mu}^{\text{blue}}$.

To calculate $\widehat{\mu}^{\text{blue}}$, as in Stokes et al. (2007), we use 200 bootstrapped data sets with size of $[n/2]$.

In our study, the MSE is estimated using $10,000$ JPS data sets. Empty cells frequently arise in JPS samples with multiple rankers. In all methods, when $n_{\mathbf{r}} = 0$, $\widehat{\pi}_{\mathbf{r}}$ is estimated by zero except Raking when

Scenario	Normal	Log-normal	σ_y^2	σ_1^2	σ_2^2	ρ	$\rho(X_1, Y)$	$\rho(X_2, Y)$	$\rho(X_1, X_2)$
1. EE	$N(0, 1.34)$	$LogN(0, 0.81)$	1.8	1	1	-0.5	0.8	0.8	0.47
2. EE	$N(0, 1.34)$	$LogN(0, 0.81)$	1.8	1	1	0.0	0.8	0.8	0.64
3. EE	$N(0, 1.34)$	$LogN(0, 0.81)$	1.8	1	1	0.5	0.8	0.8	0.82
4. NE	$N(0, 1)$	$LogN(0, 0.69)$	1	0.56	3	-0.5	0.8	0.5	0.14
5. NE	$N(0, 1)$	$LogN(0, 0.69)$	1	0.56	3	0.0	0.8	0.5	0.40
6. NE	$N(0, 1)$	$LogN(0, 0.69)$	1	0.56	3	0.5	0.8	0.5	0.66

Table 1: Parameters of six simulation scenarios. EE: equally effective; NE: non-equally effective.

$0/0$ occurs, in which case $\widehat{\pi}_{\mathbf{r}}^{\text{rake}}$ cannot be estimated. Equivalently, the JPS estimator is in the form of

$$\widehat{\mu} = \sum_{\mathbf{r}\in\mathcal{E}} \widehat{\pi}_{\mathbf{r}} \widehat{\mu}_{[\mathbf{r}]},$$

where $\mathcal{E}$ denotes the set of non-empty cells.

We assume that ranking is imperfect. The linear ranking error model is employed, where the two rankers determine the order of the measurement Y based on ranking variables $X_1 = Y + \epsilon_1$ and $X_2 = Y + \epsilon_2$, respectively. The ranking errors $(\epsilon_1, \epsilon_2)^{\mathrm{T}}$ are simulated from a multivariate normal distribution with mean 0 and variance-covariance Σ as

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Note that we can choose σ_i^2 to control $\rho(X_i, Y), i = 1, 2$, the correlation between X_i and Y, which measures the ranking quality. The correlation between X_1 and X_2, denoted by $\rho(X_1, X_2)$, is jointly affected by all parameters and it represents the similarity in the rankings of the two rankers. We consider six scenarios that are combinations of two choices of effectiveness of the rankers (i.e., equally versus non-equally effective) and three correlation values for ranking errors (i.e., $\rho \in \{-0.5, 0, 0.5\}$). Table 1 shows the six scenarios with the choices of distribution parameters, σ_1^2, σ_2^2 and ρ, along with the resulting correlation coefficients between any two of X_1, X_2, and Y.

The performance of the three isotonized estimators are measured by the relative efficiency (RE) with respect to their un-isotonized peers, i.e., the ratio of the mean squared errors (MSE) of the original JPS estimators and the MSE of their isotonized versions. Figure 1 shows the RE of the three isotonized estimators for the normal distribution. From the figure we can make the following observations.

1. In all cases, the isotonized estimators have smaller MSE than their un-isotonized counterparts. This suggests that no matter which estimator is used, when data show violation of matrix partial orderings, it is beneficial to use the isotonized version of that estimator.

2. By comparing the six different panels in Figure 1, we find that the RE curves follow similar patterns and their values are roughly the same, no matter whether the two rankers are equally effective or not, or whether their ranking errors are correlated or not (although they may affect the choice of estimators).

3. The figure further shows that there is not necessarily a monotone trend of RE with respect to the sample size n. But as n gets large, RE decreases. This is not surprising because as $n \to \infty$, the probability of violating the ordering constraints goes to zero, and so RE $\to 1$.

4. For a fixed sample size, RE tends to be large for large values of H for all three estimators.

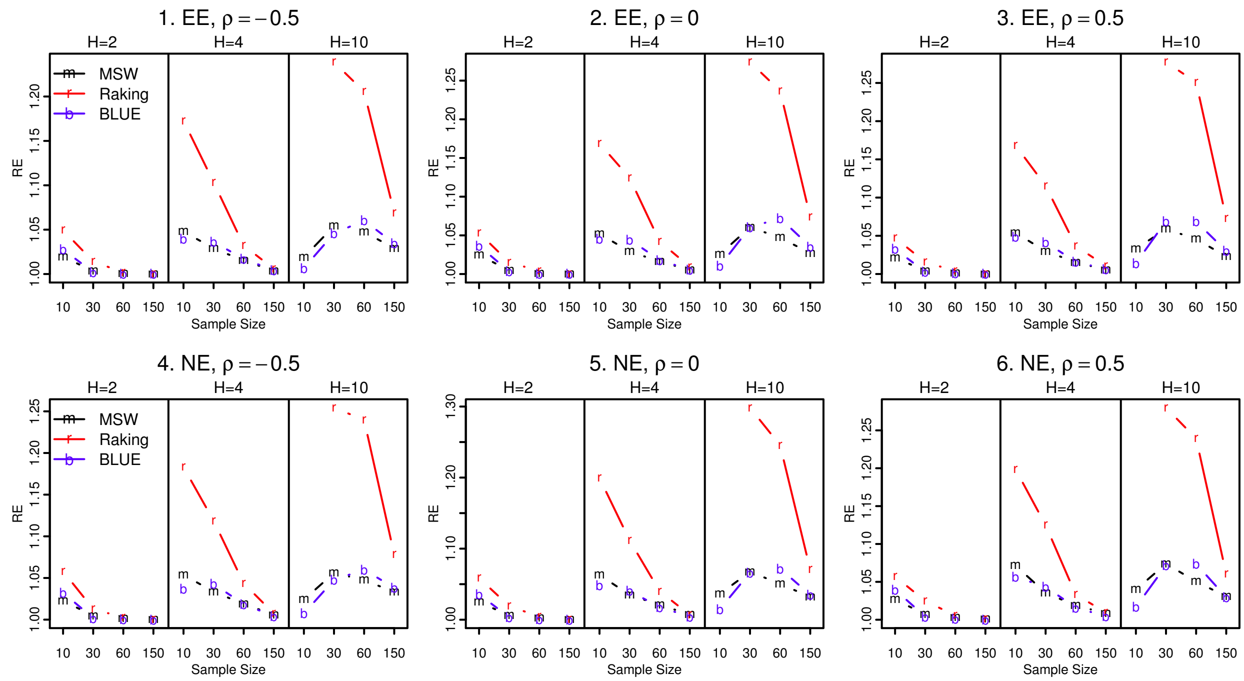

Figure 1: Relative efficiency of the isotonized estimators versus un-isotonized ones for the normal distribution in all six scenarios. EE: equally effective; NE: non-equally effective.

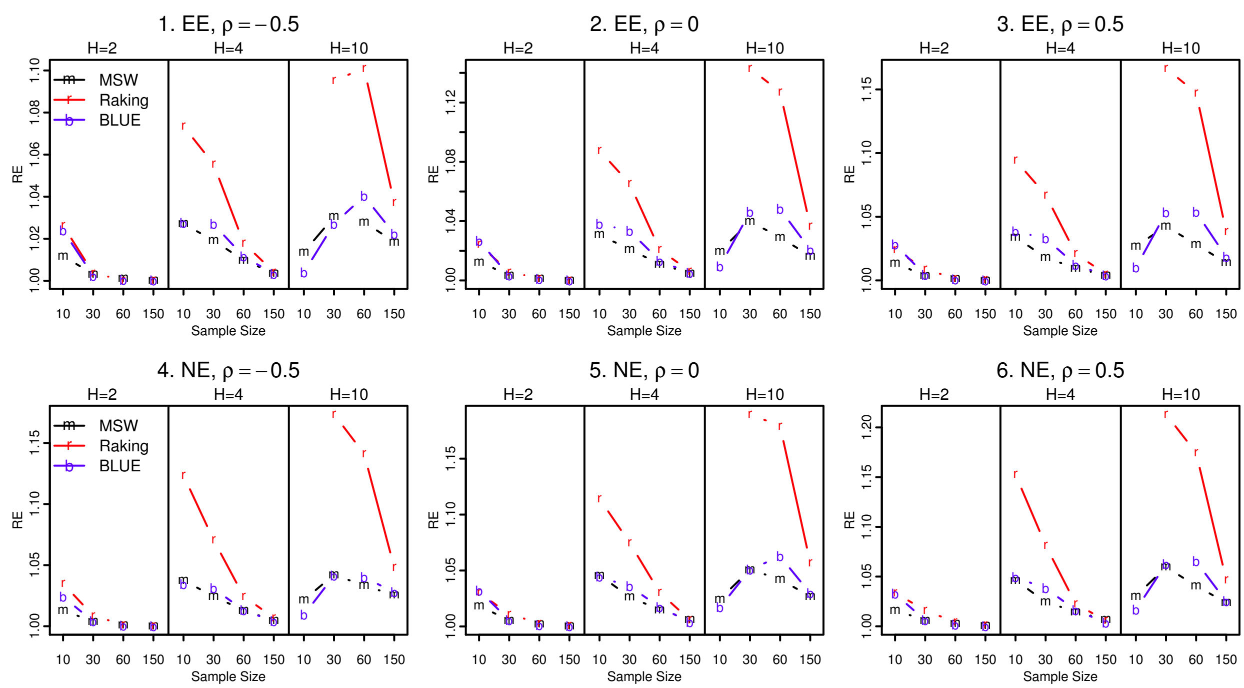

Figure 2: Relative efficiency of the isotonized estimators versus un-isotonized ones for the lognormal distribution in all six scenarios. EE: equally effective; NE: non-equally effective.

5. Among the three estimators, Raking benefits the most, especially when H is large and the sample size is small.

Figure 2 shows the RE of the lognormal distribution for the four scenarios. We also plot the RE for the uniform and exponential distributions (data not reported here). Their patterns are similar to the normal distribution with the exception that the uniform distribution has larger RE than others while the RE values of the exponential distribution are between the normal and the lognormal distributions.

Next we study the efficiency of the three estimation approaches, after employing the monotonicity constraints, for different choices of ranking efficiencies, correlations of ranking errors, the sample size n and the set size H, and the distributions of Y. In below, to measure the size of the improvements in the three isotonized estimators, we use the relative efficiency with respect to the SRS, a common reference estimator for the purpose of fair comparisons. The results are listed in Table 2.

From Table 2 we make the following observations.

1. It is easy to see that, in most cases, RE increases with H when the sample size n is fixed, and RE also increases with n when H is fixed.

2. In all three isotonized estimators, in most situations the ones based on negatively correlated and uncorrelated rankers (i.e., negatively correlated or uncorrelated ranking errors) have higher RE than those based on correlated rankers, regardless of whether the effectiveness of the two rankers are equal. Similar observations can be made in the lognormal (the right panel of Table 2) and other distributions (data not reported). This can be explained by the fact that two negatively correlated or uncorrelated rankers provide complementary or independent information on the ranks of Y, unlike the correlated rankers that offer overlapping information. If two rankers are extremely correlated and their rankings are almost the same, then the marginal benefit from the additional information provided by the second ranker is virtually zero.

3. When the two rankers are equally effective in ranking, Iso-MSW works quite well when compared with the other two methods. However, when the rankers are unequally effective, its performance is not as good as the other two in a large portion of the simulation situations. This is not surprising since MSW assumes equal effectiveness of multiple rankers and assigns equal weights to all rankers. Iso-Raking works poorly for cases with small n, but does well for cases with large n. When the sample size n is small or the set size H is large, there are many empty strata (i.e., those with $n_{\mathbf{r}} = 0$) and $\widehat{\pi}_{\mathbf{r}}^{\text{rake}}$ is known to perform poorly for these cases. However, if n is large enough, $\widehat{\pi}_{\mathbf{r}}^{\text{rake}}$ is equal to the non-parametric maximum likelihood estimator of $\pi_{\mathbf{r}}$ and is statistically efficient.

4. Iso-BLUE performs well in all cases when the sample size is not very small, especially in the case where rankers are unequally effective.

5 Example

In this section, we analyze the tree height data from Platt et al. (1988) and Chen et al. (2006). This data set contains the heights of 396 trees whose mean is 52.7 and variance is 3253.4. We treat the reported tree height data as the true population and generate JPS samples with $n = 6$, 12, 18, and 24. We assume there are two rankers, each of whom ranks a set of $H = 3$ units using noisy copies of the original observations. For the first ranker, the noisy copy is made by adding an independent Gaussian random variable with mean

ρ	H	n	Normal						Lognormal					
			Equally Eff.			Non-equally Eff.			Equally Eff.			Non-equally Eff.		
			I-M	I-R	I-B	I-M	I-R	I-B	I-M	I-R	I-B	I-M	I-R	I-B
-0.5	2	10	1.30	1.32	1.20	1.21	1.23	1.14	1.20	1.02	1.21	1.15	0.98	1.16
		30	1.34	1.34	1.30	1.23	1.26	1.22	1.20	1.17	1.24	1.16	1.13	1.18
		60	1.38	1.43	1.35	1.23	1.31	1.24	1.19	1.21	1.21	1.17	1.21	1.19
		150	1.37	1.42	1.35	1.24	1.34	1.27	1.21	1.26	1.20	1.17	1.23	1.17
	4	10	1.65	1.49	1.30	1.39	1.29	1.17	1.42	0.94	1.27	1.34	0.99	1.20
		30	1.91	1.89	1.80	1.53	1.61	1.52	1.50	1.30	1.49	1.37	1.28	1.38
		60	1.92	1.98	1.85	1.55	1.75	1.58	1.43	1.38	1.44	1.38	1.46	1.39
		150	1.93	2.07	1.89	1.54	1.84	1.63	1.49	1.58	1.48	1.38	1.53	1.41
	10	10	1.59	–	1.17	1.28	–	1.05	1.62	–	1.21	1.37	–	1.15
		30	2.49	2.31	1.91	1.75	1.76	1.55	1.86	1.26	1.56	1.60	1.22	1.42
		60	2.83	2.82	2.56	1.87	2.13	1.88	1.95	1.55	1.84	1.69	1.64	1.65
		150	2.90	3.25	2.81	1.95	2.59	2.17	2.02	2.11	1.98	1.72	2.04	1.80
0.0	2	10	1.27	1.29	1.18	1.18	1.20	1.13	1.20	1.07	1.24	1.12	0.95	1.15
		30	1.32	1.31	1.28	1.21	1.22	1.20	1.18	1.16	1.22	1.13	1.11	1.17
		60	1.33	1.34	1.31	1.22	1.26	1.23	1.18	1.17	1.20	1.14	1.14	1.15
		150	1.36	1.40	1.34	1.23	1.30	1.26	1.19	1.22	1.19	1.15	1.18	1.16
	4	10	1.55	1.38	1.30	1.33	1.20	1.17	1.37	0.92	1.28	1.28	0.99	1.18
		30	1.79	1.63	1.71	1.47	1.44	1.48	1.42	1.20	1.43	1.33	1.16	1.36
		60	1.84	1.81	1.79	1.47	1.57	1.53	1.42	1.34	1.43	1.34	1.32	1.37
		150	1.85	1.89	1.82	1.44	1.56	1.53	1.44	1.45	1.43	1.35	1.42	1.38
	10	10	1.51	–	1.14	1.24	–	1.04	1.59	–	1.24	1.37	–	1.15
		30	2.31	1.97	1.87	1.63	1.44	1.53	1.70	1.20	1.51	1.53	1.18	1.40
		60	2.48	2.20	2.29	1.78	1.82	1.90	1.82	1.43	1.78	1.61	1.42	1.60
		150	2.63	2.67	2.57	1.82	2.08	2.05	1.85	1.75	1.85	1.65	1.75	1.75
0.5	2	10	1.25	1.27	1.19	1.16	1.20	1.13	1.16	1.05	1.22	1.11	0.96	1.17
		30	1.32	1.31	1.28	1.20	1.20	1.19	1.18	1.15	1.23	1.14	1.11	1.19
		60	1.33	1.33	1.29	1.21	1.25	1.23	1.15	1.14	1.17	1.14	1.16	1.16
		150	1.33	1.34	1.32	1.22	1.29	1.26	1.18	1.19	1.18	1.14	1.16	1.15
	4	10	1.51	1.35	1.28	1.32	1.23	1.19	1.32	1.02	1.27	1.24	0.96	1.19
		30	1.72	1.58	1.66	1.44	1.39	1.46	1.39	1.16	1.42	1.32	1.14	1.35
		60	1.78	1.73	1.73	1.44	1.49	1.50	1.40	1.33	1.42	1.31	1.29	1.36
		150	1.79	1.81	1.77	1.47	1.60	1.58	1.42	1.41	1.41	1.33	1.39	1.38
	10	10	1.48	–	1.13	1.23	–	1.07	1.61	–	1.25	1.36	–	1.15
		30	2.20	1.81	1.82	1.60	1.41	1.53	1.67	1.14	1.50	1.49	1.15	1.42
		60	2.36	2.11	2.23	1.69	1.69	1.82	1.77	1.40	1.74	1.58	1.39	1.62
		150	2.48	2.44	2.44	1.76	1.96	2.00	1.81	1.71	1.82	1.59	1.65	1.72

Table 2: RE (w.r.t. SRS) of the isotonized estimators. The distributions of Y are normal and lognormal. $\rho = -0.5$: negatively correlated ranking errors; $\rho = 0$: uncorrelated ranking errors; $\rho = 0.5$: positively correlated ranking errors; Equally Eff: equally effective rankers; Non-equally Eff: non-equally effective rankers; I-M: Iso-MSW; I-R: Iso-Raking; I-B: Iso-BLUE.

0 and variance $1821.9 (= 0.56 \times 3253.4)$ to each observation; for the second ranker, we add an independent Gaussian noise with mean 0 and variance $9760.3 (= 3 \times 3253.4)$ to each observation. This is similar to the 5th senario in Table 1. For each setting, we generate $10,000$ replicate data sets, in each of which sampling is conducted without replacement; and we apply the same six estimators as in the simulation study to estimate the mean tree height. The empirical MSEs and REs (w.r.t. SRS) of the six estimators are reported in Table 3.

Table 3 shows that, again, the isotonized estimators perform better than their non-isotonized peers. Also,

n	Method	MSE	RE	Mean	s.d.	n	Method	MSE	RE	Mean	s.d.
6	SRS	521.4	1.00	52.5	22.8	18	SRS	171.5	1.00	52.6	13.1
	MSW	460.2	1.13	51.3	21.4		MSW	134.4	1.28	52.3	11.6
	Iso-MSW	438.0	1.19	51.3	20.9		Iso-MSW	130.8	1.31	52.3	11.4
	Raking	562.2	0.93	53.3	23.7		Raking	164.9	1.04	53.1	12.8
	Iso-Raking	506.5	1.03	53.5	22.5		Iso-Raking	150.2	1.14	53.1	12.2
	BLUE	520.4	1.00	49.4	22.6		BLUE	145.4	1.18	50.8	11.9
	Iso-BLUE	500.6	1.04	49.6	22.2		Iso-BLUE	139.4	1.23	50.9	11.7
12	SRS	251.7	1.00	52.2	15.9	24	SRS	128.9	1.00	52.7	11.4
	MSW	207.6	1.21	51.7	14.4		MSW	99.0	1.30	52.4	9.9
	Iso-MSW	199.9	1.26	51.8	14.1		Iso-MSW	96.6	1.33	52.4	9.8
	Raking	284.0	0.89	52.7	16.9		Raking	110.8	1.16	52.8	10.5
	Iso- Raking	242.9	1.04	52.8	15.6		Iso- Raking	102.7	1.25	52.9	10.1
	BLUE	236.2	1.07	49.6	15.1		BLUE	104.7	1.23	51.1	10.1
	Iso-BLUE	223.3	1.13	49.8	14.7		Iso-BLUE	101.2	1.27	51.2	9.9

Table 3: Results of the tree height data. In the table, "Mean" represents the average of 10,000 mean estimates, and "s.d." represents their standard deviation.

we find that MSW and Iso-MSW perform better than the other two pairs. BLUE and Iso-BLUE do well when the sample size is large. This is consistent with the results in the numerical study for non-equally effective rankers. The distribution of tree height data is skewed to the right and has a heavy tail. Thus, its distributional properties are close to that of exponential or lognormal distributions. In the simulation study, if Y is generated from exponential or lognormal distribution, MSW and Iso-MSW often perform better than the other estimators if the sample size is not large.

6 Discussions and conclusions

We have proposed a new set of estimators, which monotonize existing estimators using the matrix partial order. The new estimators are numerically shown to perform better than their un-isotonized counterparts. It is worth mentioning that the performance depends on the validity of the monotonicity assumption. When the assumption barely holds or fails, the proposed isotonized estimation would have little or no help at all. It is rare that the assumption is violated in theory (i.e., $\mu_{[\mathbf{r}_1]} \geq \mu_{[\mathbf{r}_2]}$ holds when $\mathbf{r}_1 \leq \mathbf{r}_2$ in a component-wise manner). After all, the rankers are supposed to **order** the units, and no matter how bad they could be, the judgment order statistics are expected to resemble the true order statistics. As long as the rankers are rational, the matrix partial ordering we impose is a reasonable and mild assumption.

We notice that small sample sizes often induce many empty strata; the number of empty strata becomes large as the number of rankers increases. The Raking estimators frequently disregard empty strata and the performance is relatively poor for small samples in comparison to the other two types of estimators; the MSW and BLUE estimators are, in principle, based on marginal ranks and so are less affected by the presence of empty strata. We have also considered the case of four rankers ($m = 4$). We observe that there are some improvements by considering the monotonicity constraints, but they are not as significant as the cases of two rankers. This is because the rank matrix has four dimensions and it is so sparse that the matrix partial order is seldom violated, making the isotonized estimation almost identical to the corresponding un-isotonized one in most simulated data. For future work, one may improve the performance of the Raking estimators by estimating both $\pi_{\mathbf{r}}$ and $\mu_{[\mathbf{r}]}$ through borrowing information from its neighboring strata.

Acknowledgement

The work of Johan Lim was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2011-0029104).

References

Barlow, R.E., Bartholomew, D.J., Bremner, J.M., and Brunk, H.D. (1972). *Statistical Inference under Order Restrictions*. Wiley, New York.

Block, H., Qian, S. and Sampson, A. (1994). Structure Algorithms for Partially Ordered Isotonic Regression. *Journal of the Computational and Graphical Statistics*, **3(3)**, 285-300.

Chen, Z., Bai, Z., and Sinha, B. K. (2006). *Ranked Set Sampling: Theory and Applications*. Springer, New York.

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

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

Mode, N. A., Conquest, L. L., and Marker, D. A. (1999). Ranked set sampling for ecological research: Accounting for the total costs of sampling. *Environmetrics*, 10:179–194.

Ozturk, O. (2007). Statistical inference under a stochastic ordering constraint in ranked set sampling. *Journal of Nonparametric Statistics*, **16(3)**, 131-144.

Pelz, W. and Good, I. J. (1986). Estimating probabilities from contingency tables when the marginal probabilities are known, by using additive objective functions. *The Statistician*, **35(1)**, 45–50.

Platt, W.J., Evans, G.M., and Rathbun, S.L. (1988). The population dynamics of long-lived conifer (Pinus palustris). *American Naturalist*, **131,** 491-525.

Qian, S. and Eddy, W. (1996). An algorithm for isotonic regression on ordered rectangular grids. *Journal of Computational and Graphical Statistics*, **5(3)**, 225-235.

Sysoev, O., Burdakov, O. and Grimvall, A. (2011). A segmentation-based algorithm for large-scale partially ordered monotonic regression. *Computational Statistics and Data Analysis*, **55**, 2463-2476.

Stokes, S. L., Wang, X., and Chen, M. (2007). Judgment post-stratification with multiple rankers. *Journal of Statistical Theory and Applications,* **6(4)**, 344-359.

Thompson, J. (1981). Convergence properties of the iterative 1980 census estimator. In *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 182–185.

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

Wang, X., Wang, K., and Lim, J. (2012). Isotonized CDF estimation from judgment post-stratification data with empty strata. *Biometrics*. **68(1)**, 194-202.