International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 11(Special Issue), 2011, pp 37-58

Ranking Procedures for Several Normal Populations: An Empirical Investigation

Martin Klein and Tommy Wright
Center for Statistical Research and Methodology
U.S. Census Bureau
Washington, DC, USA

[Received December 16, 2010; Revised April 5, 2011; Accepted April 9, 2011]

Abstract

Statistical agencies commonly release ranking tables in which a collection of entities such as states are ranked based on an estimate of some quantity of interest. We have a particular interest in assessing the U.S. Census Bureau's current methodology for producing ranking tables based on data from the American Community Survey. Taking such ranking tables as our motivation, this paper presents several methods for obtaining a complete ranking of several normal populations with respect to the unknown means. Most of the methods are derived from a Bayesian perspective. We explore the computational aspects of each ranking method, and we use simulation to compare the various methods under several scenarios. The procedures are then applied to a dataset from the American Community Survey of the Census Bureau. We demonstrate a use of the parametric bootstrap to quantify uncertainty in the empirical ranking. In our empirical studies, we find that the rankings produced by the various methods, including the Census Bureau's current method, are largely in agreement.

Keywords and Phrases: American Community Survey (ACS), Empirical Bayes, Hierarchical Bayes, Monte Carlo methods, Normal populations, Ranking, Simulation, Parametric bootstrap.

AMS Classification: Primary 62F07; Secondary 62D05, 62F15.

1 Introduction

The American Community Survey (ACS) of the U.S. Census Bureau seeks to collect data from a sample of approximately 250,000 randomly selected households and from group quarters (e.g., college and university dormitories, health care facilities, prisons, etc.) distributed across the United States each month. The data from the ACS are used to produce estimates of various parameters over a one-year period, a three-year period, or a five-year period for a number of different demographic characteristics. In addition to national level estimates, estimates are produced for various lower levels of geography, including the fifty states and the District of Columbia. When the true values of parameters θ_i for $i = 1, \ldots, 51$ are known for the states and the District of Columbia (e.g., average annual income or proportion below poverty level), it is possible to rank the states from smallest to largest. When θ_i are unknown, sample data are used to produce sample estimates x_i, and a ranking can be based on the observed estimates x_i which are subject to sampling error.

In this paper we present and compare several procedures for obtaining a complete ranking of k normal populations with respect to the unknown means when the variances are assumed to be known. Our motivation for this problem lies in the ranking tables that are commonly released by the Census Bureau in which the fifty states and the District of Columbia are ranked according to a sample estimate of some quantity of interest. A series of such tables is provided from the ACS as a portion of the released data products. The ranking procedure used in these released tables is the straightforward one in which the 51 estimates for the states and the District of Columbia are ordered to obtain a complete ranking.

The primary objective of this paper is an assessment of the Census Bureau's ranking procedure for the ACS produced estimates through an empirical comparison with alternative Bayesian procedures. The Census Bureau's current method is to obtain a ranking by sorting the frequentist based point estimates of the θ_i. From the Bayesian perspective, one can devise several methods to go from the joint posterior distribution of the θ_i to a complete ranking of the populations. In this paper we develop Bayesian procedures from several different points of view: ranking through posterior probabilities, ranking by sorting a Bayes estimate of the parameter of interest, and ranking by using characteristics of the posterior distribution of the ranks themselves. We then compare all of the procedures. Alternative frequentist based ranking methods can also be developed and included in the comparison, but we do not pursue this task here. We are attracted to the Bayesian approach because, with the Bayesian approach "... once the basic step of describing uncertainty through probability is admitted, we have a formal procedure for solving all inference problems; and it is a procedure that works" (Lindley, 2006). A secondary focus of the present paper is in providing a statement of the uncertainty in a released ranking.

Typically rankings are based on the model

$$x_i \,|\, \boldsymbol{\theta} \overset{ind}{\sim} N(\theta_i, \sigma_i^2), \quad i = 1, \ldots, k, \tag{1}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ is the vector of unknown population means and the information available to us is the estimate vector $\boldsymbol{x} = (x_1, \ldots, x_k)$ along with the known standard deviations $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_k)$. More specifically, x_i is the sample estimate of θ_i, σ_i is the standard error of x_i, and the sampling distribution of x_i is assumed to be normal, for $i = 1, \ldots, 51$. The straightforward or simple (SI) ranking in Table 2, which is obtained by sorting the estimates $x_1, \ldots, x_k$, ignores the known values of $\boldsymbol{\sigma}$. The goal then is to use the information ($\boldsymbol{x}$ and $\boldsymbol{\sigma}$), to obtain estimates of the unknown true ranks r_i, of the populations $i = 1, \ldots, k$, which would be obtained by ordering the θ_i. The true ranks can then be defined formally by

$$r_i = \text{rank}(\theta_i) = \sum_{j=1}^{k} I(\theta_j \leq \theta_i) = 1 + \sum_{j\,:\,j \neq i} I(\theta_j \leq \theta_i), \text{ for } i = 1, \ldots, k, \tag{2}$$

where $I(\cdot)$ denotes the indicator function. Thus the rank of the smallest θ_i is 1, the rank of the second smallest θ_i is 2, and so on. An estimated ranking, computed using the observed data $\boldsymbol{x}$ and the known standard deviations $\boldsymbol{\sigma}$, is denoted by $\hat{r}_1, \ldots, \hat{r}_k$. Note that the SI ranking method is defined by $\hat{r}_i = 1 + \sum_{j\,:\,j \neq i} I(x_j \leq x_i)$, for $i = 1, \ldots, k$. As indicated previously, the primary purpose of this paper is to evaluate the performance of various methods for computing $\hat{r}_1, \ldots, \hat{r}_k$.

We choose to judge the estimated rankings based on probabilities of the form $P(|\hat{r}_i - r_i| \leq c_0)$, $i = 1, \ldots, k$, for some fixed c_0. This choice of evaluation criteria is based on the fact that these probabilities are simple and straightforward to interpret. For instance, one can construct the confidence interval $[\hat{r}_i - c_0,\ \hat{r}_i + c_0]$ for r_i, and then the probability $P(|\hat{r}_i - r_i| \leq c_0)$ would represent the individual confidence level of this interval.

There is a large body of literature on ranking and selection (Gibbons, Olkin, and Sobel, 1999; Gupta and Panchapakesan, 2002). Much of this literature considers the probability that the largest population is chosen correctly as the criterion of interest. Other authors such as Goldstein and Spiegelhalter (1996), Hall and Miller (2009), Laird and Louis (1989), and Xie et al. (2009) study intervals and other one at a time criteria for individual ranks, with some extensions for making multiple comparisons. As these authors indicate, there are certainly other possible criteria of interest. For instance, if one is interested in assessing multiple comparisons, then a probability such as $P(|\hat{r}_i - r_i| \leq c_0, \forall\, i = 1, \ldots, k)$ would be a relevant criterion. We do not pursue such a multiple comparison criterion in the present paper, but instead choose to focus on the one at a time criteria $P(|\hat{r}_i - r_i| \leq c_0)$, $i = 1, \ldots, k$.

The outline of the rest of the paper is as follows. In Section 2 we discuss Bayesian approaches to ranking, and we outline details of the empirical and hierarchical Bayes models that we will consider. Section 3 introduces several ranking procedures and addresses the relevant computational issues for each one. The section concludes with a simulation based comparison of all the procedures. Section 4 provides an example to illustrate the application of each of the ranking procedures to a publicly available dataset from the ACS. Section 4 also illustrates the use of the parametric bootstrap to

assess the uncertainty in the empirical rankings. Section 5 concludes the paper with some discussion of our results, and topics for future research.

Ranking or comparing overlapping populations is a challenging task. Here we work under the premise of ordering several normal populations with respect to their unknown means, when the standard deviation of each population is known. To conclude this section, we note that we use the statistical computing software R (R Development Core Team, 2010) to obtain the empirical results presented in this paper.

2 Bayesian Approaches To Ranking

In this paper we consider both frequentist and Bayesian settings. The frequentist setting is described by (1) with $\boldsymbol{\theta}$ fixed and unknown. The Bayesian setting involves placing a prior distribution on $\boldsymbol{\theta}$ and utilizing the posterior distribution to obtain a ranking of the populations conditional on $\boldsymbol{x}$. One can employ various procedures to obtain a ranking from the posterior distribution. For instance, a ranking can be obtained by sorting an appropriate Bayes estimate of $\boldsymbol{\theta}$ such as the posterior mean. Alternatively one could select the ranking based on evaluation of posterior probabilities, for example, choose the ranking that is most probable. By defining the ranks as in (2), one can even look at the posterior distribution of the ranks themselves. More generally, if one specifies a loss function that depends on a ranking and on $\boldsymbol{\theta}$, then an optimal ranking procedure is obtained by minimizing the posterior expected loss. A theoretical development of this approach is provided by Govindarajulu and Harvey (1971). In our Bayesian modeling we specify the prior distribution on $\boldsymbol{\theta}$ by

$$\theta_1, \ldots, \theta_k \,|\, \mu, \tau \overset{iid}{\sim} N(\mu, \tau^2). \tag{3}$$

In this application it is desirable for the analysis to be data driven as much as possible, and hence we do not wish to make any subjective specification of the hyper-parameters μ and τ. Hence empirical and hierarchical Bayes (Berger, 1985) are viable approaches, and we consider and compare both. We now briefly describe each approach.

1. Empirical Bayes. Under the empirical Bayes approach, the prior parameters μ and τ^2 can be estimated using the marginal density of $\boldsymbol{x}$, which is given by

$$p(\boldsymbol{x} \,|\, \mu, \tau) = \left(\prod_{i=1}^{k} [2\pi(\sigma_i^2 + \tau^2)]^{-1/2} \right) \exp\left\{ -\frac{1}{2} \sum_{i=1}^{k} \frac{(x_i - \mu)^2}{\sigma_i^2 + \tau^2} \right\}. \tag{4}$$

   Estimates of μ and τ^2 can then be found by maximizing $p(\boldsymbol{x} \,|\, \mu, \tau)$ with respect to μ and τ^2. The estimates can be computed by numerically solving the following system of nonlinear equations:

$$g_1(\mu, \tau^2) \equiv \mu - \frac{\sum_{i=1}^{k} \frac{x_i}{\sigma_i^2 + \tau^2}}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2 + \tau^2}} = 0, \quad g_2(\mu, \tau^2) \equiv \tau^2 - \frac{\sum_{i=1}^{k} \frac{(x_i - \mu)^2 - \sigma_i^2}{(\sigma_i^2 + \tau^2)^2}}{\sum_{i=1}^{k} \frac{1}{(\sigma_i^2 + \tau^2)^2}} = 0. \tag{5}$$

To solve these equations, we use the R library 'nleqslv' (Hasselman, 2010) along with the explicit expressions for the derivatives of $g_1(\mu, \tau^2)$ and $g_2(\mu, \tau^2)$ with respect to μ and τ^2. Letting $\hat{\mu}$ and $\hat{\tau}^2$ denote the solution to (5), we then set $\mu = \hat{\mu}$, $\tau^2 = \hat{\tau}^2$, and proceed with the analysis. The posterior density of $\boldsymbol{\theta}$ is given by

$$p(\boldsymbol{\theta} \,|\, \boldsymbol{x}, \mu, \tau) = \left(\prod_{i=1}^{k} \left[\frac{2\pi}{1/\sigma_i^2 + 1/\tau^2} \right]^{-1/2} \right) \exp \left\{ -\frac{1}{2} \sum_{i=1}^{k} \frac{\left(\theta_i - \frac{x_i \tau^2 + \mu \sigma_i^2}{\tau^2 + \sigma_i^2} \right)^2}{\frac{1}{1/\sigma_i^2 + 1/\tau^2}} \right\},$$

and hence

$$\theta_i \,|\, \boldsymbol{x}, \mu, \tau \overset{ind}{\sim} N\left[\delta_i(x_i, \mu, \tau),\, \omega_i^2(\tau) \right] \tag{6}$$

where

$$\delta_i(x_i, \mu, \tau) = \frac{x_i \tau^2 + \mu \sigma_i^2}{\tau^2 + \sigma_i^2},\ \omega_i^2(\tau) = \frac{1}{1/\sigma_i^2 + 1/\tau^2}.$$

2. Hierarchical Bayes. Under the hierarchical Bayes approach, we place a second stage prior density $p(\mu, \tau)$ on the hyperparameters (μ, τ). In this paper, we assume that

$$p(\mu, \tau) = p(\mu \,|\, \tau) p(\tau) \propto 1, \tag{7}$$

for $-\infty < \mu < \infty$ and $0 < \tau < \infty$. Hence we assume independent and non-informative uniform prior distributions on μ and τ. Next we briefly discuss the posterior distribution of $\boldsymbol{\theta}$, μ, and τ obtained under this hierarchical model. For details, see Gelman et al. (2004).

Under this model it can be shown that

$$\mu \,|\, \tau, \boldsymbol{x} \sim N(\hat{\mu}_\tau, V_\tau) \tag{8}$$

where

$$\hat{\mu}_\tau = \frac{\sum_{i=1}^{k} \frac{x_i}{\sigma_i^2 + \tau^2}}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2 + \tau^2}} \quad \text{and} \quad V_\tau^{-1} = \sum_{i=1}^{k} \frac{1}{\sigma_i^2 + \tau^2}.$$

It is also true that

$$p(\tau \,|\, \boldsymbol{x}) \propto p(\tau)\, V_\tau^{1/2} \left[\prod_{i=1}^{k} (\sigma_i^2 + \tau^2)^{-1/2} \right] \exp \left(-\frac{1}{2} \sum_{i=1}^{k} \frac{(x_i - \hat{\mu}_\tau)^2}{\sigma_i^2 + \tau^2} \right).$$

One can then take random draws from the posterior distribution of $(\boldsymbol{\theta}, \mu, \tau)$ by making use of the factorization

$$p(\boldsymbol{\theta}, \mu, \tau \,|\, \boldsymbol{x}) = p(\tau \,|\, \boldsymbol{x})\, p(\mu \,|\, \tau, \boldsymbol{x})\, p(\boldsymbol{\theta} \,|\, \mu, \tau, \boldsymbol{x}). \tag{9}$$

Hence in order to simulate from the posterior distribution of $\boldsymbol{\theta}$, one can proceed as follows (Gelman et al., 2004).

(i) Draw from the density $p(\tau \,|\, \boldsymbol{x})$ numerically using the inverse cumulative distribution function method.

(ii) Draw $\mu \,|\, \tau, \boldsymbol{x} \sim N(\hat{\mu}_\tau, V_\tau)$.

(iii) Draw $\theta_i \,|\, \mu, \tau, \boldsymbol{x} \sim N\left(\frac{x_i\tau^2+\mu\sigma_i^2}{\tau^2+\sigma_i^2}, \frac{1}{1/\sigma_i^2+1/\tau^2}\right)$ independently for $i = 1, \ldots, k$.

In the sequel, we will make use of the following fact which follows from (6) and (8):

$$\boldsymbol{\theta} \,|\, \tau, \boldsymbol{x} \sim N_k\left[\boldsymbol{a}(\tau, \boldsymbol{x}) + \hat{\mu}_\tau \boldsymbol{b}(\tau),\, \boldsymbol{\Sigma}(\tau) + V_\tau \boldsymbol{b}(\tau)\boldsymbol{b}'(\tau)\right], \tag{10}$$

where

$$\boldsymbol{a}(\tau, \boldsymbol{x}) = \begin{pmatrix} \frac{x_1\tau^2}{\tau^2+\sigma_1^2} \\ \vdots \\ \frac{x_k\tau^2}{\tau^2+\sigma_k^2} \end{pmatrix}, \quad \boldsymbol{b}(\tau) = \begin{pmatrix} \frac{\sigma_1^2}{\tau^2+\sigma_1^2} \\ \vdots \\ \frac{\sigma_k^2}{\tau^2+\sigma_k^2} \end{pmatrix},$$

$$\boldsymbol{\Sigma}(\tau) = \operatorname{diag}\left(\frac{1}{1/\sigma_1^2 + 1/\tau^2}, \ldots, \frac{1}{1/\sigma_k^2 + 1/\tau^2}\right).$$

3 Ranking Procedures

In this section, we introduce several different procedures for obtaining a complete ranking of the populations based upon the observed data $\boldsymbol{x}$. In subsection 3.1 we introduce two ranking procedures that make explicit use of posterior probabilities, in subsection 3.2 we discuss some ranking procedures that result by sorting certain point estimates of $\theta_1, \ldots, \theta_k$, and in subsection 3.3 we discuss ranking procedures based on the posterior distribution of the ranks. Throughout, we address the relevant computational issues for each ranking procedure. A summary and short-hand notation for each method is provided in subsection 3.4. In subsection 3.5 we present a comparison of the procedures.

3.1 Ranking Based on Posterior Probabilities

Below we define a ranking procedure that sequentially determines the population with rank 1, population with rank 2, and so on, by maximizing certain posterior probabilities. The posterior probabilities appearing in each step are computed with respect to the distribution (6) if we adopt the empirical Bayes model, or (9) if we adopt the hierarchical Bayes model.

Step 1. For each $s \in \mathcal{I}_1 = \{1, \ldots, k\}$, compute the joint posterior probability

$$p_{1,s}(\boldsymbol{x}) = P[\theta_s \le \theta_j,\, j \in \mathcal{I}_1 \setminus \{s\} \,|\, \boldsymbol{x}]. \tag{11}$$

Determine $\hat{q}_1$ such that $p_{1,\hat{q}_1}(\boldsymbol{x}) = \max\{p_{1,s}(\boldsymbol{x}) : s \in \mathcal{I}_1\}$ and associate population $\hat{q}_1$ with the smallest rank. That is, $\hat{r}_{\hat{q}_1} = 1$.

Step 2. We next look for the population with the second smallest rank. For each $s \in \mathcal{I}_2 = \{1, \ldots, k\} \setminus \{\hat{q}_1\}$, compute the joint posterior probability

$$p_{2,s}(\boldsymbol{x}) = P[\theta_s \leq \theta_j,\, j \in \mathcal{I}_2 \setminus \{s\} \,|\, \boldsymbol{x}]. \tag{12}$$

Determine $\hat{q}_2$ such that $p_{2,\hat{q}_2}(\boldsymbol{x}) = \max\{p_{2,s}(\boldsymbol{x}) : s \in \mathcal{I}_2\}$ and associate population $\hat{q}_2$ with the second smallest rank. That is, $\hat{r}_{\hat{q}_2} = 2$.

Step 3. We next look for the population with the third smallest rank. For each $s \in \mathcal{I}_3 = \{1, \ldots, k\} \setminus \{\hat{q}_1, \hat{q}_2\}$, compute the joint posterior probability

$$p_{3,s}(\boldsymbol{x}) = P[\theta_s \leq \theta_j,\, j \in \mathcal{I}_3 \setminus \{s\} \,|\, \boldsymbol{x}]. \tag{13}$$

Determine $\hat{q}_3$ such that $p_{3,\hat{q}_3}(\boldsymbol{x}) = \max\{p_{3,s}(\boldsymbol{x}) : s \in \mathcal{I}_3\}$ and associate population $\hat{q}_3$ with the third smallest rank. That is, $\hat{r}_{\hat{q}_3} = 3$.

Continue in this fashion to get a complete estimated ranking of the populations.

One can easily modify this procedure so that it instead sequentially computes the population with rank k, then the population with rank $k-1$, and so on. In the sequel, we refer to these two procedures as Procedure 1 and Procedure 2, respectively.

We now discuss some computational methods for the evaluation of the probabilities that appear in the Procedures 1 and 2. For ease of exposition, we focus on the form of the probabilities as they appear in Step 1. It is straightforward to generalize these computational methods for the probabilities in an arbitrary Step ℓ of each procedure.

Implementation of the above procedures requires one to compute probabilities of the form

$$P(\theta_s \leq \theta_1, \ldots, \theta_s \leq \theta_{s-1}, \theta_s \leq \theta_{s+1}, \ldots, \theta_s \leq \theta_k \,|\, \boldsymbol{x}) \quad \text{[Procedure 1]} \tag{14}$$

and

$$P(\theta_s \geq \theta_1, \ldots, \theta_s \geq \theta_{s-1}, \theta_s \geq \theta_{s+1}, \ldots, \theta_s \geq \theta_k \,|\, \boldsymbol{x}) \quad \text{[Procedure 2]}. \tag{15}$$

A simple Monte Carlo estimator of these probabilities is obtained by independently drawing several vectors $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(m)}$ from the posterior distribution of $\boldsymbol{\theta}$, and approximating the probabilities (14) and (15) by

$$\frac{1}{m} \sum_{j=1}^{m} I(\theta_s^{(j)} \leq \theta_1^{(j)}, \ldots, \theta_s^{(j)} \leq \theta_{s-1}^{(j)}, \theta_s^{(j)} \leq \theta_{s+1}^{(j)}, \ldots, \theta_s^{(j)} \leq \theta_k^{(j)}) \tag{16}$$

and

$$\frac{1}{m} \sum_{j=1}^{m} I(\theta_s^{(j)} \geq \theta_1^{(j)}, \ldots, \theta_s^{(j)} \geq \theta_{s-1}^{(j)}, \theta_s^{(j)} \geq \theta_{s+1}^{(j)}, \ldots, \theta_s^{(j)} \geq \theta_k^{(j)}), \tag{17}$$

respectively. Under both the empirical and hierarchical Bayes models, we may use Rao-Blackwellization (Givens and Hoeting, 2005) to obtain more efficient Monte Carlo estimators of (14) and (15).

Consider the empirical Bayes model in which the posterior distribution of $\boldsymbol{\theta}$ is given by (6), with $\mu = \hat{\mu}$ and $\tau = \hat{\tau}$, the solutions to (5). To obtain a Rao-Blackwellized version of (14), we note the following:

$$\begin{aligned}
&E[I(\theta_s \le \theta_1, \ldots, \theta_s \le \theta_{s-1}, \theta_s \le \theta_{s+1}, \ldots, \theta_s \le \theta_k) \,|\, \theta_s = \nu, \boldsymbol{x}] \\
&= P[\theta_s \le \theta_1, \ldots, \theta_s \le \theta_{s-1}, \theta_s \le \theta_{s+1}, \ldots, \theta_s \le \theta_k \,|\, \theta_s = \nu, \boldsymbol{x}] \\
&= P\left(\frac{\theta_s - \delta_i(x_i, \mu, \tau)}{\omega_i(\tau)} \le \frac{\theta_i - \delta_i(x_i, \mu, \tau)}{\omega_i(\tau)}, i \ne s \,\Big|\, \theta_s = \nu, \boldsymbol{x}\right) \\
&= \prod_{i\,:\,i \ne s} P\left[\frac{\nu - \delta_i(x_i, \mu, \tau)}{\omega_i(\tau)} \le \frac{\theta_i - \delta_i(x_i, \mu, \tau)}{\omega_i(\tau)} \,\Big|\, \boldsymbol{x}\right] \\
&= \prod_{i\,:\,i \ne s} \left[1 - \Phi\left(\frac{\nu - \delta_i(x_i, \mu, \tau)}{\omega_i(\tau)}\right)\right], \qquad (18)
\end{aligned}$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function, and the third equality above followed from independence of the θ_i in the posterior distribution. Thus a Rao-Blackwellized version of the Monte Carlo estimator (16), obtained by conditioning on θ_s, is given by

$$\frac{1}{m}\sum_{j=1}^{m}\{\prod_{i\,:\,i \ne s}[1 - \Phi\left(\frac{\theta_s^{(j)} - \delta_i(x_i, \mu, \tau)}{\omega_i(\tau)}\right)]\}, \qquad (19)$$

where $\theta_s^{(1)}, \ldots, \theta_s^{(m)}$ are iid as $N[\delta_s(x_s, \mu, \tau), \omega_s^2(\tau)]$. An analogous argument leads to the following Rao-Blackwellized version of the Monte Carlo estimator (17):

$$\frac{1}{m}\sum_{j=1}^{m}\{\prod_{i\,:\,i \ne s}[\Phi\left(\frac{\theta_s^{(j)} - \delta_i(x_i, \mu, \tau)}{\omega_i(\tau)}\right)]\}, \qquad (20)$$

where again, $\theta_s^{(1)}, \ldots, \theta_s^{(m)}$ are iid as $N[\delta_s(x_s, \mu, \tau), \omega_s^2(\tau)]$.

Next consider the hierarchical Bayes model in which the joint posterior density of $(\boldsymbol{\theta}, \mu, \tau)$ can be expressed as in (9). To obtain a Rao-Blackwellized version of (14), we can apply a similar logic as in the case of empirical Bayes. However, in order to get a result similar to (18), we must also condition on μ and τ. Thus, we obtain the expression:

$$\begin{aligned}
&E[I(\theta_s \le \theta_1, \ldots, \theta_s \le \theta_{s-1}, \theta_s \le \theta_{s+1}, \ldots, \theta_s \le \theta_k) \,|\, \theta_s = \nu, \mu, \tau, \boldsymbol{x}] \\
&= \prod_{i\,:\,i \ne s} \left[1 - \Phi\left(\frac{\nu - \delta_i(x_i, \mu, \tau)}{\omega_i(\tau)}\right)\right],
\end{aligned}$$

and therefore a Rao-Blackwellized version of (16) is now given by

$$\frac{1}{m}\sum_{j=1}^{m}\{\prod_{i\,:\,i \ne s}[1 - \Phi\left(\frac{\theta_s^{(j)} - \delta_i(x_i, \mu^{(j)}, \tau^{(j)})}{\omega_i(\tau^{(j)})}\right)]\}, \qquad (21)$$

where $\{(\theta_s^{(j)}, \mu^{(j)}, \tau^{(j)}),\ j = 1, \ldots, m\}$ are drawn as *iid* from the posterior density of (θ_s, μ, τ) which is obtained from (9). The Rao-Blackwellized version of (17) is obtained in a similar fashion as

$$\frac{1}{m}\sum_{j=1}^{m}\{\prod_{i\,:\,i\neq s}[\Phi\left(\frac{\theta_s^{(j)} - \delta_i(x_i, \mu^{(j)}, \tau^{(j)})}{\omega_i(\tau^{(j)})}\right)]\}, \tag{22}$$

where again, $\{(\theta_s^{(j)}, \mu^{(j)}, \tau^{(j)}),\ j = 1, \ldots, m\}$ are drawn as *iid* according to the posterior density of (θ_s, μ, τ) which is obtained from (9).

The Rao-Blackwellized estimators above are more efficient than the respective simple Monte Carlo estimators in (16) and (17) in terms of possessing smaller mean squared error. They are also more efficient in the sense that a total of m (empirical Bayes) or $3m$ (hierarchical Bayes) random draws are needed as opposed to $m \times k$.

3.2 Ranking Based on Estimates of $\boldsymbol{\theta}$

Once we have estimates $x_1, \ldots, x_k$ of $\theta_1, \ldots, \theta_k$, a ranking can be easily obtained by sorting the estimated values. We consider the ranking obtained by sorting the components of the posterior mean of $\boldsymbol{\theta}$. Under empirical Bayes, the posterior mean is given by (6) with μ and τ^2 replaced by the appropriate estimates. Under hierarchical Bayes, an expression for the posterior mean can be obtained as follows:

$$\begin{aligned}
E(\theta_i \,|\, \boldsymbol{x}) &= \int_0^\infty \int_{-\infty}^\infty \int_{\mathbb{R}^k} \theta_i\, p(\boldsymbol{\theta}, \mu, \tau \,|\, \boldsymbol{x})\, d\boldsymbol{\theta}\, d\mu\, d\tau \\
&= \int_0^\infty \int_{-\infty}^\infty \int_{\mathbb{R}^k} \theta_i\, p(\tau \,|\, \boldsymbol{x})\, p(\mu \,|\, \tau, \boldsymbol{x})\, p(\boldsymbol{\theta} \,|\, \mu, \tau, \boldsymbol{x})\, d\boldsymbol{\theta}\, d\mu\, d\tau \\
&= \int_0^\infty \int_{-\infty}^\infty p(\tau \,|\, \boldsymbol{x})\, p(\mu \,|\, \tau, \boldsymbol{x}) \left(\frac{x_i\tau^2 + \mu\sigma_i^2}{\tau^2 + \sigma_i^2}\right) d\mu\, d\tau \\
&= \int_0^\infty \frac{x_i\tau^2 + \hat{\mu}_\tau\sigma_i^2}{\tau^2 + \sigma_i^2}\, p(\tau \,|\, \boldsymbol{x})\, d\tau.
\end{aligned} \tag{23}$$

Thus, for the hierarchical Bayes model, (23) provides an expression for the posterior mean of each component of $\boldsymbol{\theta}$ that can be evaluated using one dimensional Monte Carlo or numerical integration.

In the empirical Bayes setting, we also consider a ranking obtained by sorting the Bayes estimate of $\boldsymbol{\theta}$ presented by Louis (1984). The estimator has the form

$$\hat{\theta}_i^L = \zeta + A_i(x_i - \zeta) \tag{24}$$

where

$$A_i = \frac{D_i}{1 + D_i\sigma_i^2\lambda},\ D_i = \frac{\tau^2}{\sigma_i^2 + \tau^2},$$

and the values of ζ and λ are obtained by solving the system of equations:

$$f_1(\zeta,\lambda) \equiv \zeta + \frac{1}{k}\sum_{i=1}^{k} A_i(x_i-\zeta) - \mu - \frac{1}{k}\sum_{i=1}^{k} Y_i(\mu) = 0, \tag{25}$$

$$f_2(\zeta,\lambda) \equiv \sum_{i=1}^{k}[A_i(x_i-\zeta) - \frac{1}{k}\sum_{j=1}^{k} A_j(x_j-\zeta)]^2 - \frac{k-1}{k}\sum_{i=1}^{k} D_i\sigma_i^2 - \sum_{i=1}^{k}[Y_i(\mu) - \frac{1}{k}\sum_{j=1}^{k} Y_j(\mu)]^2 = 0, \tag{26}$$

where $Y_i(\mu) = D_i(x_i - \mu)$. The estimator (24) is designed so that the empirical distribution of the estimates $\hat{\theta}_1^L, \ldots, \hat{\theta}_k^L$ closely resembles the empirical distribution of the unobserved $\theta_1, \ldots, \theta_k$. More precisely, the equations (25) and (26) ensure that the sample mean and sample variance of the estimators $\hat{\theta}_1^L, \ldots, \hat{\theta}_k^L$ match the posterior expectation of the sample mean and sample variance of $\theta_1, \ldots, \theta_k$, respectively. That is, (25) and (26) ensure that (and are equivalent to):

$$\frac{1}{k}\sum_{i=1}^{k}\hat{\theta}_i^L = E\left[\frac{1}{k}\sum_{i=1}^{k}\theta_i \,\middle|\, \boldsymbol{x}\right],$$

$$\frac{1}{k-1}\sum_{i=1}^{k}(\hat{\theta}_i^L - \frac{1}{k}\sum_{j=1}^{k}\hat{\theta}_j^L)^2 = E\left[\frac{1}{k-1}\sum_{i=1}^{k}(\theta_i - \frac{1}{k}\sum_{j=1}^{k}\theta_j)^2 \,\middle|\, \boldsymbol{x}\right],$$

under the independent normal posterior given by (6). Louis' (1984) estimator (24) may be more appropriate for ranking than the posterior mean since it leads to estimates that are more representative of the parameter ensemble, i.e., the empirical distribution of $\theta_1, \ldots, \theta_k$. Just as we did to solve (5), to solve the nonlinear system of equations defined by (25) and (26), we use the R library 'nleqslv' (Hasselman, 2010) along with the explicit expressions for the derivatives of $f_1(\zeta,\lambda)$ and $f_2(\zeta,\lambda)$ with respect to ζ and λ.

3.3 Ranking Based on the Posterior Distribution of the Ranks

Laird and Louis (1989) look directly at the posterior distribution of the ranks. Because the ranks can be formally defined by (2), it becomes clear that the ranks themselves are random variables in the Bayesian setting. Thus we may use the posterior distribution of the r_i to obtain a ranking, for instance, by sorting the posterior mean ranks:

$$\hat{r}_i = E(r_i \,|\, \boldsymbol{x}) = 1 + \sum_{j\,:\,j\neq i} P(\theta_j \leq \theta_i \,|\, \boldsymbol{x}). \tag{27}$$

The expression above clearly shows how a ranking based on sorting the posterior mean ranks can be interpreted as an ordering of sums of certain posterior probabilities. A ranking based on the posterior distribution of the ranks has the advantage of providing a straightforward way to quantify uncertainty in the ranking from a Bayesian point of view. For instance, by computing a 95% posterior interval on each r_i.

In the case of empirical Bayes, the posterior mean (27) of each rank easily simplifies to

$$\hat{r}_i = 1 + \sum_{j\,:\,j\neq i} \Phi\left(\frac{\delta_i(x_i,\mu,\tau) - \delta_j(x_j,\mu,\tau)}{[\omega_i^2(\tau)+\omega_j^2(\tau)]^{1/2}}\right).$$

In the case of hierarchical Bayes, it follows from (10) (after some simplification) that for $i \neq j$,

$$(\theta_j - \theta_i) \,|\, \tau, \boldsymbol{x} \sim N\left[\mu_{[\theta_j-\theta_i]}(\tau, x_j, x_i),\ \sigma^2_{[\theta_j-\theta_i]}(\tau)\right]$$

where

$$\mu_{[\theta_j-\theta_i]}(\tau,x_j,x_i) = \frac{x_j\tau^2 + \hat{\mu}_\tau\sigma_j^2}{\tau^2+\sigma_j^2} - \frac{x_i\tau^2+\hat{\mu}_\tau\sigma_i^2}{\tau^2+\sigma_i^2},$$

$$\sigma^2_{[\theta_j-\theta_i]}(\tau) = \frac{1}{1/\sigma_j^2+1/\tau^2} + \frac{1}{1/\sigma_i^2+1/\tau^2} + V_\tau \frac{\tau^4(\sigma_j^2-\sigma_i^2)^2}{[(\tau^2+\sigma_j^2)(\tau^2+\sigma_i^2)]^2}.$$

Therefore, in this case we have:

$$\begin{aligned} P[\theta_j \leq \theta_i \,|\, \boldsymbol{x}] &= E[P(\theta_j - \theta_i \leq 0 \,|\, \tau, \boldsymbol{x}) \,|\, \boldsymbol{x}] \\ &= E\left[\Phi\left(\frac{-\mu_{[\theta_j-\theta_i]}(\tau,x_j,x_i)}{\sigma_{[\theta_j-\theta_i]}(\tau)}\right) \,|\, \boldsymbol{x}\right] \\ &= \int_0^\infty \Phi\left(\frac{-\mu_{[\theta_j-\theta_i]}(\tau,x_j,x_i)}{\sigma_{[\theta_j-\theta_i]}(\tau)}\right) p(\tau \,|\, \boldsymbol{x})\, d\tau \end{aligned}$$

which provides an expression that can be evaluated using one dimensional Monte Carlo or numerical integration. The resulting Monte Carlo estimator of $P(\theta_j \leq \theta_i \,|\, \boldsymbol{x})$ is

$$\frac{1}{m}\sum_{\ell=1}^m \Phi\left(\frac{-\mu_{[\theta_j-\theta_i]}(\tau^{(\ell)},x_j,x_i)}{\sigma_{[\theta_j-\theta_i]}(\tau^{(\ell)})}\right)$$

where $\tau^{(1)},\ldots,\tau^{(m)} \overset{iid}{\sim} p(\tau \,|\, \boldsymbol{x})$, which is a Rao-Blackwellization of the simple Monte Carlo estimator:

$$\frac{1}{m}\sum_{\ell=1}^m I(\theta_j^{(\ell)} \leq \theta_i^{(\ell)}),$$

with $(\theta_j^{(1)},\theta_i^{(1)}),\ldots,(\theta_j^{(m)},\theta_i^{(m)})$ drawn as *iid* from the posterior distribution of (θ_j,θ_i).

Thus we have convenient and efficient methods of computing the posterior mean rank (27) in both the empirical and hierarchical Bayes models.

3.4 Summary of the Procedures

At this point we have introduced a number of ranking procedures. For convenience, we provide a list of the procedures below, along with a short-hand name that we use in the sequel.

SI: The ranking obtained by replacing the unknown θ_i in (2) with the estimates x_i.

P1EB: Procedure 1 defined in subsection 3.1 under empirical Bayes.

P1HB: Procedure 1 defined in subsection 3.1 under hierarchical Bayes.

P2EB: Procedure 2 defined in subsection 3.1 under empirical Bayes.

P2HB: Procedure 2 defined in subsection 3.1 under hierarchical Bayes.

PMEB: Ranking by sorting the posterior means of the θ_i under empirical Bayes.

PMHB: Ranking by sorting the posterior means of the θ_i under hierarchical Bayes.

LEB: Ranking by sorting Louis' (1984) estimator (24) under empirical Bayes.

PREB: Ranking by sorting the posterior mean ranks (27) under empirical Bayes.

PRHB: Ranking by sorting the posterior mean ranks (27) under hierarchical Bayes.

3.5 Comparison of the Procedures

In this section, we turn to a simulation based comparison of the ten ranking procedures from the frequentist perspective. In each case we simulate from the normal model (1), and we consider five different settings, each time taking $k = 5$:

(i) $\boldsymbol{\theta}$ =(10.0, 10.2, 10.4, 10.6, 10.8), $\boldsymbol{\sigma}$ =(0.07, 0.07, 0.07, 0.07, 0.07);

(ii) $\boldsymbol{\theta}$ =(10.0, 10.2, 10.4, 10.6, 10.8), $\boldsymbol{\sigma}$ =(0.05, 0.05, 0.2, 0.2, 0.2);

(iii) $\boldsymbol{\theta}$ =(10.0, 10.2, 10.7, 11.2, 11.4), $\boldsymbol{\sigma}$ =(0.15, 0.15, 0.25, 0.15, 0.15);

(iv) $\boldsymbol{\theta}$ =(10.0, 10.5, 10.7, 11.0, 11.2), $\boldsymbol{\sigma}$ =(0.1, 0.3, 0.3, 0.1, 0.5);

(v) $\boldsymbol{\theta}$ =(9.8, 10.5, 10.7, 10.9, 11.6), $\boldsymbol{\sigma}$ =(0.5, 0.1, 0.1, 0.1, 0.5).

In each case the true value of each rank is obviously $r_i = i$. Plots of the normal density curves for each of these simulation settings are provided in Figure 1. Our goal is to examine various cases such as when the sampling distributions of $x_1, \ldots, x_5$ are equally spaced with equal dispersion as in setting (i); ranks r_1 and r_2 relatively easy to estimate while r_3, r_4, and r_5 are difficult to distinguish as in setting (ii); upper two and lower two ranks difficult to distinguish while center rank is easier to distinguish

as in setting (iii); the case where the populations overlap very much making nearly all ranks difficult to distinguish as in setting (iv); and finally, the case where r_1 and r_5 are fairly easy to determine while the center ranks are more difficult to distinguish as in setting (v).

Recall that our implementation of Procedures 1 and 2 requires us to determine the maximum of several probabilities that are computed through a Monte Carlo approach. Therefore, in order to ensure the quality of our results, it is important that m, the number of replications appearing in the Monte Carlo estimator, be sufficiently large. To determine an appropriate m, we use the following approach which provides an m that is independent of $\boldsymbol{x}$. Obviously, the data $\boldsymbol{x}$ will be replicated many times throughout the simulation, therefore it is very convenient to have an m that does not depend on $\boldsymbol{x}$. Consider Step ℓ of Procedure 1, and suppose that we are using a simple Monte Carlo estimator along the lines of (16) to compute the probabilities $\{p_{\ell,s}(\boldsymbol{x}) : s \in \mathcal{I}_\ell\}$. Notice that under the Bayesian model, $\sum_{s\in\mathcal{I}_\ell} p_{\ell,s}(\boldsymbol{x}) = 1$, which implies that $\max\{p_{\ell,s}(\boldsymbol{x}) : s \in \mathcal{I}_\ell\} \geq \frac{1}{k-\ell+1}$. Note that the simple Monte Carlo estimator has standard deviation $\frac{\sqrt{p_{\ell,s}(\boldsymbol{x})(1-p_{\ell,s}(\boldsymbol{x}))}}{\sqrt{m}} \leq \frac{1}{2\sqrt{m}}$. Let us take a small $\epsilon > 0$. Then by choosing m to satisfy

$$\frac{1}{2\sqrt{m}} \leq \frac{\epsilon}{k-\ell+1}, \tag{28}$$

we ensure that the standard deviation of the Monte Carlo estimator of $\max\{p_{\ell,s}(\boldsymbol{x}) : s \in \mathcal{I}_\ell\}$ is less than or equal to $\epsilon \cdot \max\{p_{\ell,s}(\boldsymbol{x}) : s \in \mathcal{I}_\ell\}$. It is essential that $\max\{p_{\ell,s}(\boldsymbol{x}) : s \in \mathcal{I}_\ell\}$ be computed precisely since it is this probability that we must locate at Step ℓ. If there are multiple probabilities that are fairly large, we need each of their Monte Carlo estimators to be precise so that we can correctly determine which of these probabilities is largest. Choosing m according to (28) should suffice since it will provide an upper bound of $\epsilon \cdot p_{\ell,s}(\boldsymbol{x})$ (actually, $\frac{\epsilon}{k-\ell+1}$) for the standard deviation of the Monte Carlo estimator for any $p_{\ell,s}(\boldsymbol{x}) \geq \frac{1}{k-\ell+1}$. If a Rao-Blackwellized version of (16) is used instead, the bound (28) can still be applied, although it is likely to give an m that is much larger than needed. This method is designed to work well in our simulation settings where k is not too large, and we require a general method for selecting m that does not require additional tuning. In data analysis situations where k is fairly large, the inequality (28) can naturally lead to a huge m which may be infeasible. Note that the above reasoning is applicable to Procedure 2 as well.

The simulation results are provided in Table 1 which displays the marginal probability $P(\hat{r}_i = r_i)$, $i = 1, \ldots, k$, for each method. In these results, the m at each step of Procedures 1 and 2 is chosen to satisfy (28) with $\epsilon = 0.05$. Throughout subsections 3.1 - 3.3, we presented several Rao-Blackwellized/improved estimators of the various probabilities/expectations needed in the ranking procedures. The simulation results shown here make use of each of these improved estimators. In particular, we use Rao-Blackwellized estimators of the probabilities of Procedures 1 and 2, and therefore the

Monte Carlo standard deviation is likely to be substantially less than the guaranteed bound provided by (28). We consider the one at a time correct ranking probabilities $P(\hat{r}_i = r_i)$, $i = 1, \ldots, k$, as the criteria of comparison. We choose to focus on these individual correct ranking probabilities because they are simple and straightforward to interpret. We now provide a summary of the results.

1. In setting (i) where $\sigma_1, \ldots, \sigma_k$ are equal, all of the ranking procedures give equivalent results. Under each procedure, each estimated rank has a high probability of being correct. The probability that the center rank is correct is 0.96 which is slightly less than 0.98, the probability that the largest (or smallest) estimated rank is correct.

2. The empirical Bayes procedures P1EB, P2EB, PMEB, and PREB provide comparable results. The hierarchical Bayes procedures P1HB, P2HB, and PMHB, provide comparable results, however, in setting (v), the correct ranking probabilities under PRHB are noticeably lower than those from the other hierarchical Bayes methods. In setting (v), P1EB, P2EB, and PMEB perform somewhat poorly in comparison to several of the other methods; in this setting LEB seems to outperform the other empirical Bayes methods to some extent.

3. In settings (ii) and (iii) the Bayes procedures generally (with a few exceptions) provide equal or larger correct ranking probabilities than the simple procedure. In setting (v), however, we notice that SI noticeably outperforms the other procedures. In setting (iv) the Bayes procedures provide a slightly higher correct ranking probability for $\hat{r}_1$ and $\hat{r}_2$ in comparison with SI; the probabilities are essentially equal across procedures for $\hat{r}_3$; and for $\hat{r}_4$ and $\hat{r}_5$ the probabilities are largest for SI.

4. Based on the simulation study and the discussion above, we cannot claim any procedure to be uniformly the best in terms of the probabilities $P(\hat{r}_i = r_i)$, $i = 1, \ldots, k$.

4 Example from the American Community Survey

In this section we apply each of the ranking methods to an example dataset. The data considered here provide the estimated percent of people below the poverty level for each state in the United States and the District of Columbia, along with the estimated standard deviation of the percentage. The data are based on 2008 American Community Survey 1-year estimates, and they are available from http://factfinder.census.gov, Table R1701. For more information about these data, we refer to the U.S. Census Bureau document titled *2008 ACS Accuracy of the Data (US)*, available from `http://www.census.gov/acs/www/data_documentation/documentation_main/`. In the context of this example, x_i denotes the estimate for the percentage of people below the poverty

level in state i. The states are indexed alphabetically, $i = 1$ corresponds to Alabama, $i = 2$ to Alaska, etc. We note that because of large sample sizes within each state, it is reasonable to assume that each x_i is approximately normally distributed with a known standard deviation (2008 ACS Accuracy of the Data (US)). Table 2 presents the rankings obtained by applying each of the ranking procedures. To obtain these data analysis results, we make use of each of the Rao-Blackwellized/improved estimators discussed in subsections 3.1 - 3.3. Note that in this example, the values of x_i and σ_i have been rounded before applying the ranking procedures. As a result, there are two pairs of states, (Missouri and Ohio), (Alabama and South Carolina), with the same estimated percentage of people below the poverty level, and the same standard error for the estimate. For these states the estimated ranks produced by SI, PMEB, PMHB, LEB, PREB, and PRHB, will be tied. Of course, the procedure SI will produce tied rankings for any states with equal values of x_i.

Under our frequentist model (1), a parametric bootstrap can be used to estimate the frequentist properties of the empirical ranks from each procedure. To describe the parametric bootstrap in this setting, suppose that from the data we have the estimated ranks $\hat{r}_1, \ldots, \hat{r}_k$. The parametric bootstrap can then be applied as follows.

1. Generate $x_i^* \sim N(x_i, \sigma_i^2)$, independently for $i = 1, \ldots, k$.

2. Use $x_1^*, \ldots, x_k^*$ to arrive at a ranking $\hat{r}_1^*, \ldots, \hat{r}_k^*$. To obtain $\hat{r}_1^*, \ldots, \hat{r}_k^*$ from $x_1^*, \ldots, x_k^*$, one should use the same procedure as was used originally to obtain $\hat{r}_1, \ldots, \hat{r}_k$ from $x_1, \ldots, x_k$.

3. Repeat Steps (1) and (2) B times to get $(\hat{r}_{1,1}^*, \ldots, \hat{r}_{k,1}^*), \ldots, (\hat{r}_{1,B}^*, \ldots, \hat{r}_{k,B}^*)$, a collection of bootstrap replications of the ranks. We can then use these replications to estimate the distribution of $\hat{r}_1, \ldots, \hat{r}_k$ and its properties.

The results of applying the bootstrap procedure to the example dataset are displayed in Tables 3 and 4, which respectively provide the bootstrap estimates of $P(|r_i - \hat{r}_i| \leq 1)$ and $P(|r_i - \hat{r}_i| \leq 2)$, for each procedure. For a given value c_0, a bootstrap estimate of $P(|r_i - \hat{r}_i| \leq c_0)$ can be computed as $\frac{1}{B}\sum_{b=1}^{B} I(|\hat{r}_i - \hat{r}_{i,b}^*| \leq c_0)$.

It turns out that except for a few cases of minor discrepancies, all the ranking procedures are mostly in agreement with each other.

5 Concluding Remarks

Recall that the primary objective of the paper is to assess the U.S. Census Bureau's ranking methodology, while a secondary interest is in providing a statement of uncertainty for a released ranking. We are reluctant to make strong general statements beyond the numerical information that we have generated. We do believe that conclusion (1) of subsection 3.5 is a generalizable result, and will pursue it in our future

research. At this point we do not see substantial differences among the ranking methods. Therefore we do not see a strong reason to deviate from SI, the Census Bureau's current ranking method, in cases where one does not wish to incorporate prior knowledge into the analysis. There are certainly many other Bayesian and non-Bayesian ranking methods which could be introduced into the comparison. Apart from the SI method, in this paper we have chosen to focus on procedures that are motivated from a Bayesian perspective. It is certainly possible to devise additional non-Bayesian ranking methods, including classical ranking and selection procedures based on other criterion such as the indifference zone and subset selection approaches (Gibbons, Olkin, and Sobel, 1999). Within the Bayesian perspective, we have tried to incorporate procedures from several different points of view: ranking through posterior probabilities, ranking by sorting a Bayes estimate of the parameter of interest, and ranking by using characteristics of the posterior distribution of the ranks themselves. It is possible that a hybrid of two or more methods could outperform those that we have studied. We have also focused on situations in which prior knowledge about the populations is not brought into the analysis. We have used empirical and hierarchical Bayes methods in order to achieve a strongly data driven analysis. It is easy to imagine a situation where prior knowledge is available so that a priori, we would not want to assume that $\theta_1, \ldots, \theta_k$ are identically distributed. The Bayesian perspective provides a straightforward way to incorporate this prior knowledge into the ranking procedure, which could lead to improved results.

We have identified several areas which we intend to pursue in future research. These include (i) a careful study of how to provide a simple and accurate statement of the uncertainty in a released ranking; (ii) generalizations of the conclusions of our empirical studies; and (iii) comparisons of additional ranking methods, including those Bayesian procedures which incorporate informative prior knowledge, additional frequentist based ranking methods, and the comparison of various methods using multiple comparison style criteria such as $P(|\hat{r}_i - r_i| \leq c_0, \forall i = 1, \ldots, k)$. In a future communication, we also intend to present a thorough review and evaluation of ranking methods used by other statistical agencies in the United States and throughout the world.

Acknowledgments

A great deal of appreciation goes to the referee and editor-in-chief for stimulating and provocative questions that have greatly enhanced the content of this paper. This paper is released to inform interested parties of ongoing research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

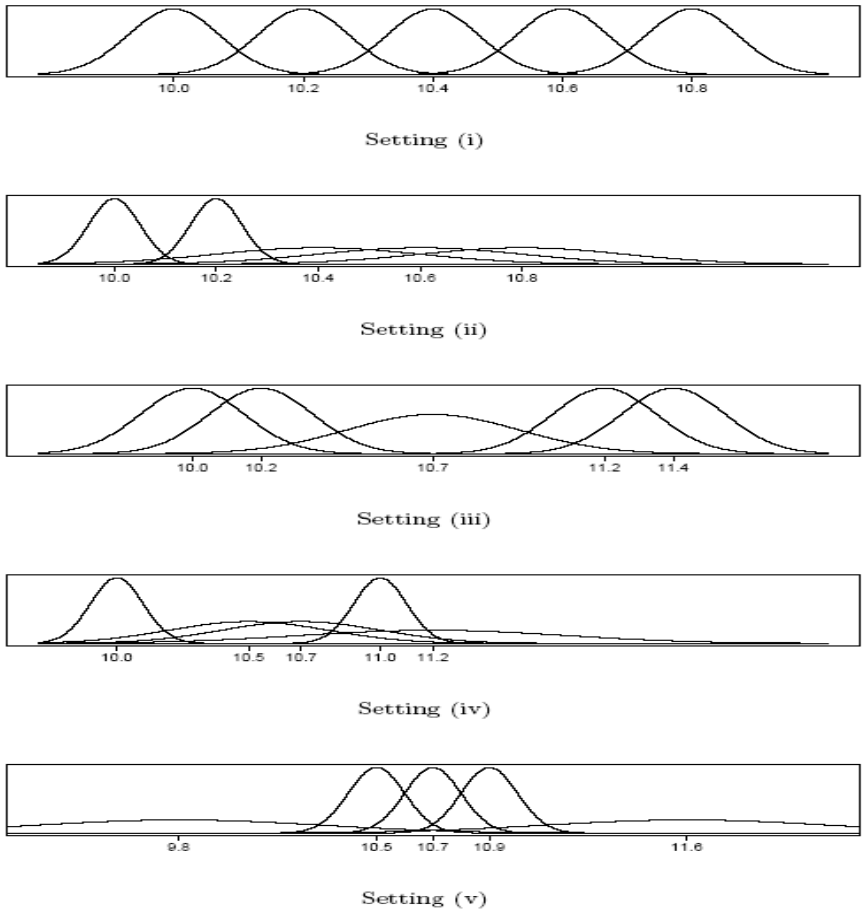

Figure 1: Normal populations to be ranked in the five simulation settings

Table 1: Simulated $P(\hat{r}_i = r_i)$ for estimated rank $\hat{r}_i$ for each procedure

		SI	P1 EB	P1 HB	P2 EB	P2 HB	PM EB	PM HB	L EB	PR EB	PR HB
Setting (i)	$\hat{r}_1$	.98	.98	.98	.98	.98	.98	.98	.98	.98	.98
	$\hat{r}_2$	.96	.96	.96	.96	.96	.96	.96	.96	.96	.96
	$\hat{r}_3$	.96	.96	.96	.96	.96	.96	.96	.96	.96	.96
	$\hat{r}_4$	.97	.97	.97	.97	.97	.97	.97	.97	.97	.97
	$\hat{r}_5$	.98	.98	.98	.98	.98	.98	.98	.98	.98	.98
Setting (ii)	$\hat{r}_1$	.96	.98	.98	.99	.98	.98	.98	.98	.99	.98
	$\hat{r}_2$	.78	.83	.81	.85	.83	.85	.83	.74	.84	.81
	$\hat{r}_3$	.56	.60	.57	.62	.59	.61	.59	.52	.61	.58
	$\hat{r}_4$	.53	.53	.52	.54	.54	.53	.53	.53	.53	.53
	$\hat{r}_5$	.72	.72	.72	.72	.73	.72	.72	.72	.72	.72
Setting (iii)	$\hat{r}_1$	.82	.82	.82	.82	.82	.82	.82	.82	.82	.82
	$\hat{r}_2$	.79	.79	.79	.80	.79	.80	.79	.81	.79	.79
	$\hat{r}_3$	.91	.94	.93	.94	.93	.93	.93	.94	.93	.92
	$\hat{r}_4$	.78	.79	.79	.79	.79	.79	.79	.79	.79	.78
	$\hat{r}_5$	.82	.82	.82	.82	.82	.82	.82	.82	.82	.82
Setting (iv)	$\hat{r}_1$	.93	.98	.97	.98	.97	.98	.97	.96	.98	.97
	$\hat{r}_2$	.56	.59	.58	.60	.59	.59	.59	.57	.60	.59
	$\hat{r}_3$	.43	.44	.43	.43	.44	.43	.44	.45	.43	.43
	$\hat{r}_4$	.56	.34	.46	.36	.48	.32	.46	.49	.29	.37
	$\hat{r}_5$	.63	.31	.46	.35	.50	.31	.48	.47	.27	.37
Setting (v)	$\hat{r}_1$	.91	.37	.78	.34	.74	.35	.80	.51	.30	.62
	$\hat{r}_2$	.84	.36	.72	.34	.68	.34	.73	.48	.29	.57
	$\hat{r}_3$	.80	.63	.75	.63	.74	.64	.76	.76	.61	.73
	$\hat{r}_4$	.85	.34	.69	.35	.72	.34	.74	.48	.30	.56
	$\hat{r}_5$	.92	.34	.74	.37	.78	.35	.80	.51	.30	.61

Table 2: State rankings by estimated percent of population below poverty level in 2008

			Estimated ranks									
State	x_i	σ_i	SI	P1 EB	P1 HB	P2 EB	P2 HB	PM EB	PM HB	L EB	PR EB	PR HB
New Hampshire	7.6	.36	1.0	1	1	1	1	1.0	1.0	1.0	1.0	1.0
Maryland	8.1	.18	2.0	2	2	2	2	2.0	2.0	2.0	2.0	2.0
Alaska	8.4	.49	3.0	3	3	3	3	3.0	3.0	3.0	3.0	4.0
New Jersey	8.7	.18	4.0	4	4	4	4	4.0	4.0	4.0	4.0	3.0
Hawaii	9.1	.43	5.0	5	5	5	5	5.0	5.0	5.0	5.0	5.0
Connecticut	9.3	.24	6.0	6	6	6	6	6.0	6.0	6.0	6.0	6.0
Wyoming	9.4	.55	7.0	7	7	7	7	7.0	7.0	7.0	7.0	9.0
Minnesota	9.6	.18	8.5	8	8	8	8	8.0	8.0	8.0	8.0	7.0
Utah	9.6	.30	8.5	9	9	9	9	9.0	9.0	9.0	9.0	8.0
Delaware	10.0	.49	10.5	11	11	11	11	11.0	11.0	11.0	11.0	11.0
Massachusetts	10.0	.18	10.5	10	10	10	10	10.0	10.0	10.0	10.0	10.0
Virginia	10.2	.18	12.0	12	12	12	12	12.0	12.0	12.0	12.0	12.0
Wisconsin	10.4	.18	13.0	13	13	13	13	13.0	13.0	13.0	13.0	13.0
Vermont	10.6	.55	14.0	14	14	14	14	14.0	14.0	14.0	14.0	14.0
Nebraska	10.8	.30	15.0	15	15	15	15	15.0	15.0	15.0	15.0	15.0
Kansas	11.3	.30	17.0	17	17	17	17	17.0	17.0	17.0	17.0	16.0
Nevada	11.3	.36	17.0	16	16	18	18	18.0	18.0	18.0	18.0	17.0
Washington	11.3	.18	17.0	18	18	16	16	16.0	16.0	16.0	16.0	18.0
Colorado	11.4	.30	19.0	19	19	19	19	19.0	19.0	19.0	19.0	19.0
Iowa	11.5	.30	20.0	20	20	20	20	20.0	20.0	20.0	20.0	20.0
Rhode Island	11.7	.49	21.0	21	21	21	21	21.0	21.0	21.0	21.0	21.0
North Dakota	12.0	.55	22.0	22	22	22	22	22.0	22.0	22.0	22.0	22.0
Pennsylvania	12.1	.12	23.0	23	23	23	23	23.0	23.0	23.0	23.0	23.0
Illinois	12.2	.12	24.0	24	24	24	24	24.0	24.0	24.0	24.0	24.0
Maine	12.3	.36	25.0	25	25	25	25	25.0	25.0	25.0	25.0	25.0
South Dakota	12.5	.55	26.0	26	26	26	26	26.0	26.0	26.0	26.0	26.0
Idaho	12.6	.55	27.0	27	27	27	27	27.0	27.0	27.0	27.0	27.0
Indiana	13.1	.24	28.0	28	28	28	28	28.0	28.0	28.0	28.0	28.0
Florida	13.2	.12	29.0	29	29	29	29	29.0	29.0	29.0	29.0	29.0
California	13.3	.12	30.0	30	30	30	30	30.0	30.0	30.0	30.0	30.0
Missouri	13.4	.18	31.5	31	31	31	31	31.5	31.5	31.5	31.5	31.5
Ohio	13.4	.18	31.5	32	32	32	32	31.5	31.5	31.5	31.5	31.5
New York	13.6	.12	33.5	34	34	33	33	34.0	34.0	34.0	34.0	34.0
Oregon	13.6	.30	33.5	33	33	34	34	33.0	33.0	33.0	33.0	33.0
Michigan	14.4	.18	35.0	35	35	35	35	35.0	35.0	35.0	35.0	35.0
North Carolina	14.6	.24	36.0	36	36	36	36	36.0	36.0	36.0	36.0	36.0
Arizona	14.7	.24	37.5	38	38	38	38	37.0	37.0	37.0	37.0	38.0
Georgia	14.7	.18	37.5	39	39	37	37	38.0	38.0	38.0	38.0	39.0
Montana	14.8	.55	39.0	37	37	39	39	39.0	39.0	39.0	39.0	37.0
Tennessee	15.5	.24	40.0	40	40	40	40	40.0	40.0	40.0	40.0	40.0
Alabama	15.7	.30	41.5	41	42	42	41	41.5	41.5	41.5	41.5	41.5
South Carolina	15.7	.30	41.5	42	41	41	42	41.5	41.5	41.5	41.5	41.5
Texas	15.8	.12	43.0	43	43	43	43	43.0	43.0	43.0	43.0	43.0
Oklahoma	15.9	.30	44.0	44	44	44	44	44.0	44.0	44.0	44.0	44.0
West Virginia	17.0	.43	45.0	46	46	46	45	46.0	46.0	45.0	46.0	46.0
New Mexico	17.1	.43	46.0	47	47	47	46	47.0	47.0	47.0	47.0	47.0
Washington, DC	17.2	.79	47.0	45	45	45	47	45.0	45.0	46.0	45.0	45.0
Arkansas	17.3	.43	49.0	48	48	50	50	48.0	48.0	48.0	48.0	48.0
Kentucky	17.3	.30	49.0	50	50	48	48	50.0	50.0	50.0	50.0	50.0
Louisiana	17.3	.36	49.0	49	49	49	49	49.0	49.0	49.0	49.0	49.0
Mississippi	21.2	.55	51.0	51	51	51	51	51.0	51.0	51.0	51.0	51.0

Table 3: Bootstrap estimate of $P(|\hat{r}_i - r_i| \leq 1)$ for estimated rank $\hat{r}_i$ for each procedure

			Bootstrap estimate of $P(\lvert\hat{r}_i - r_i\rvert \leq 1)$									
State	x_i	σ_i	SI	P1 EB	P1 HB	P2 EB	P2 HB	PM EB	PM HB	L EB	PR EB	PR HB
New Hampshire	7.6	.36	.96	.96	.96	.95	.95	.96	.95	.96	.96	.96
Maryland	8.1	.18	.98	.99	.99	1.0	1.0	.99	.98	.99	1.0	1.0
Alaska	8.4	.49	.73	.74	.74	.72	.72	.73	.72	.73	.73	.80
New Jersey	8.7	.18	.96	.97	.97	.98	.98	.98	.97	.97	.98	.92
Hawaii	9.1	.43	.62	.62	.62	.60	.60	.61	.60	.62	.60	.58
Connecticut	9.3	.24	.77	.77	.77	.80	.80	.78	.77	.78	.78	.79
Wyoming	9.4	.55	.43	.44	.44	.37	.37	.41	.40	.41	.41	.42
Minnesota	9.6	.18	.59	.82	.82	.81	.81	.80	.79	.80	.79	.73
Utah	9.6	.30	.43	.58	.58	.56	.55	.57	.56	.56	.56	.62
Delaware	10.0	.49	.30	.43	.43	.42	.42	.44	.43	.44	.46	.50
Massachusetts	10.0	.18	.61	.73	.73	.79	.79	.75	.74	.74	.74	.77
Virginia	10.2	.18	.78	.79	.79	.74	.74	.77	.77	.79	.77	.76
Wisconsin	10.4	.18	.84	.85	.85	.83	.83	.84	.84	.85	.84	.83
Vermont	10.6	.55	.51	.51	.51	.51	.51	.51	.51	.50	.49	.50
Nebraska	10.8	.30	.73	.73	.73	.73	.73	.73	.72	.72	.73	.71
Kansas	11.3	.30	.50	.50	.50	.51	.51	.50	.50	.50	.50	.43
Nevada	11.3	.36	.44	.44	.44	.39	.39	.39	.39	.39	.40	.45
Washington	11.3	.18	.62	.68	.67	.53	.52	.44	.44	.42	.45	.69
Colorado	11.4	.30	.49	.47	.47	.53	.53	.49	.49	.49	.49	.48
Iowa	11.5	.30	.52	.51	.51	.54	.54	.53	.53	.53	.52	.50
Rhode Island	11.7	.49	.42	.42	.41	.41	.41	.42	.41	.42	.42	.40
North Dakota	12.0	.55	.37	.38	.38	.32	.32	.36	.35	.37	.38	.35
Pennsylvania	12.1	.12	.74	.74	.74	.79	.79	.75	.74	.75	.73	.75
Illinois	12.2	.12	.76	.78	.78	.78	.78	.77	.76	.76	.72	.78
Maine	12.3	.36	.57	.56	.55	.61	.61	.57	.57	.58	.58	.62
South Dakota	12.5	.55	.53	.54	.53	.56	.56	.54	.53	.53	.50	.50
Idaho	12.6	.55	.50	.54	.54	.49	.49	.51	.50	.51	.47	.47
Indiana	13.1	.24	.68	.73	.72	.64	.64	.68	.68	.68	.66	.65
Florida	13.2	.12	.77	.76	.76	.78	.78	.77	.76	.77	.74	.72
California	13.3	.12	.71	.71	.72	.73	.73	.71	.70	.71	.69	.69
Missouri	13.4	.18	.42	.56	.56	.55	.55	.42	.42	.42	.41	.41
Ohio	13.4	.18	.43	.59	.59	.57	.57	.43	.43	.43	.44	.43
New York	13.6	.12	.72	.78	.78	.90	.90	.74	.74	.74	.79	.77
Oregon	13.6	.30	.57	.66	.66	.66	.66	.70	.70	.70	.68	.69
Michigan	14.4	.18	.78	.74	.74	.79	.79	.77	.77	.77	.78	.68
North Carolina	14.6	.24	.67	.65	.65	.67	.67	.66	.66	.67	.66	.70
Arizona	14.7	.24	.51	.73	.73	.71	.71	.69	.68	.69	.70	.69
Georgia	14.7	.18	.60	.58	.57	.76	.76	.81	.80	.81	.81	.67
Montana	14.8	.55	.48	.35	.35	.51	.52	.46	.46	.47	.46	.43
Tennessee	15.5	.24	.76	.73	.74	.76	.76	.74	.74	.75	.75	.70
Alabama	15.7	.30	.41	.61	.58	.58	.58	.42	.42	.42	.43	.41
South Carolina	15.7	.30	.41	.60	.61	.57	.59	.43	.43	.42	.44	.42
Texas	15.8	.12	.80	.87	.87	.80	.80	.82	.81	.82	.81	.79
Oklahoma	15.9	.30	.65	.62	.62	.66	.66	.64	.64	.64	.63	.62
West Virginia	17.0	.43	.51	.66	.67	.67	.51	.66	.66	.50	.66	.64
New Mexico	17.1	.43	.58	.56	.56	.53	.58	.54	.54	.53	.55	.67
Washington, DC	17.2	.79	.32	.46	.46	.42	.29	.45	.44	.48	.45	.53
Arkansas	17.3	.43	.58	.57	.56	.40	.40	.56	.56	.56	.56	.63
Kentucky	17.3	.30	.63	.47	.46	.68	.68	.44	.43	.41	.47	.67
Louisiana	17.3	.36	.61	.67	.66	.62	.62	.64	.64	.63	.65	.74
Mississippi	21.2	.55	1.0	1.0	1.0	1.0	1.0	1.0	.99	1.0	1.0	1.0

Table 4: Bootstrap estimate of $P(|\hat{r}_i - r_i| \leq 2)$ for estimated rank $\hat{r}_i$ for each procedure

			Bootstrap estimate of $P(\lvert\hat{r}_i - r_i\rvert \leq 2)$									
State	x_i	σ_i	SI	P1 EB	P1 HB	P2 EB	P2 HB	PM EB	PM HB	L EB	PR EB	PR HB
New Hampshire	7.6	.36	1.0	1.0	1.0	1.0	1.0	1.0	.99	1.0	1.0	1.0
Maryland	8.1	.18	1.0	1.0	1.0	1.0	1.0	1.0	.99	1.0	1.0	1.0
Alaska	8.4	.49	.93	.93	.93	.92	.92	.92	.92	.92	.92	.92
New Jersey	8.7	.18	1.0	1.0	1.0	1.0	1.0	1.0	.99	1.0	1.0	1.0
Hawaii	9.1	.43	.83	.83	.83	.78	.79	.81	.80	.81	.80	.78
Connecticut	9.3	.24	.92	.92	.92	.94	.94	.93	.93	.93	.94	.94
Wyoming	9.4	.55	.68	.69	.69	.62	.63	.67	.66	.67	.68	.61
Minnesota	9.6	.18	.90	.98	.98	.97	.98	.97	.97	.97	.97	.94
Utah	9.6	.30	.75	.78	.79	.76	.76	.78	.77	.78	.77	.86
Delaware	10.0	.49	.56	.66	.66	.69	.69	.68	.68	.67	.70	.74
Massachusetts	10.0	.18	.90	.93	.93	.96	.96	.95	.94	.94	.94	.94
Virginia	10.2	.18	.96	.96	.96	.94	.94	.95	.94	.95	.94	.95
Wisconsin	10.4	.18	.97	.97	.98	.96	.96	.97	.96	.97	.97	.97
Vermont	10.6	.55	.65	.66	.66	.68	.68	.67	.66	.66	.65	.69
Nebraska	10.8	.30	.88	.87	.87	.88	.88	.88	.88	.88	.88	.87
Kansas	11.3	.30	.73	.71	.71	.74	.74	.72	.72	.72	.72	.63
Nevada	11.3	.36	.66	.61	.60	.68	.68	.67	.67	.68	.67	.66
Washington	11.3	.18	.86	.92	.91	.78	.78	.71	.70	.69	.71	.92
Colorado	11.4	.30	.73	.73	.73	.74	.74	.73	.72	.73	.73	.72
Iowa	11.5	.30	.73	.74	.73	.74	.74	.73	.74	.73	.73	.72
Rhode Island	11.7	.49	.59	.57	.57	.59	.59	.59	.59	.59	.62	.59
North Dakota	12.0	.55	.53	.54	.54	.50	.50	.53	.53	.53	.55	.53
Pennsylvania	12.1	.12	.95	.94	.94	.97	.97	.96	.95	.95	.94	.96
Illinois	12.2	.12	.97	.98	.98	.96	.96	.97	.96	.97	.95	.96
Maine	12.3	.36	.80	.80	.80	.79	.79	.80	.80	.80	.80	.82
South Dakota	12.5	.55	.67	.67	.66	.70	.69	.69	.68	.68	.64	.66
Idaho	12.6	.55	.63	.64	.64	.63	.62	.64	.64	.64	.61	.63
Indiana	13.1	.24	.83	.86	.86	.82	.82	.83	.83	.83	.83	.84
Florida	13.2	.12	.95	.95	.95	.96	.96	.95	.95	.95	.94	.94
California	13.3	.12	.92	.93	.93	.93	.94	.92	.91	.92	.90	.89
Missouri	13.4	.18	.73	.84	.84	.84	.84	.73	.73	.73	.73	.73
Ohio	13.4	.18	.73	.80	.80	.77	.77	.73	.73	.73	.72	.74
New York	13.6	.12	.93	.95	.95	.96	.96	.93	.92	.93	.94	.93
Oregon	13.6	.30	.74	.78	.78	.78	.78	.83	.83	.83	.83	.84
Michigan	14.4	.18	.93	.91	.91	.94	.94	.92	.92	.92	.92	.89
North Carolina	14.6	.24	.87	.86	.86	.87	.87	.86	.86	.86	.86	.88
Arizona	14.7	.24	.90	.92	.92	.92	.91	.98	.97	.98	.98	.88
Georgia	14.7	.18	.95	.83	.83	1.0	1.0	.95	.94	.95	.95	.86
Montana	14.8	.55	.63	.83	.83	.64	.64	.62	.62	.62	.62	.84
Tennessee	15.5	.24	.90	.89	.89	.90	.90	.89	.89	.89	.89	.86
Alabama	15.7	.30	.77	.80	.94	.94	.80	.77	.77	.77	.77	.74
South Carolina	15.7	.30	.77	.92	.81	.80	.94	.78	.77	.78	.78	.73
Texas	15.8	.12	.96	.99	.99	.97	.96	.97	.97	.97	.96	.98
Oklahoma	15.9	.30	.80	.79	.79	.81	.81	.80	.80	.80	.80	.81
West Virginia	17.0	.43	.68	.81	.81	.81	.68	.81	.81	.68	.81	.80
New Mexico	17.1	.43	.73	.85	.86	.86	.73	.86	.86	.86	.86	.86
Washington, DC	17.2	.79	.64	.58	.58	.52	.66	.57	.56	.60	.59	.69
Arkansas	17.3	.43	.74	.91	.91	.58	.58	.91	.90	.91	.91	.96
Kentucky	17.3	.30	.82	.73	.72	.97	.97	.69	.68	.67	.70	.87
Louisiana	17.3	.36	.80	.84	.84	.81	.81	.82	.82	.81	.83	.89
Mississippi	21.2	.55	1.0	1.0	1.0	1.0	1.0	1.0	.99	1.0	1.0	1.0

References

[1] 2008 ACS accuracy of the data (US), (Revised January 25, 2010) *U.S. Census Bureau Document*, available at `http://www.census.gov/acs/www/data_documentation/documentation_main/`.

[2] Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd Ed., Springer, New York.

[3] Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2003). *Bayesian Data Analysis*, 2nd Ed., Chapman & Hall/CRC, New York.

[4] Gibbons, J. D., Olkin, I. and Sobel, M. (1999). *Selecting and Ordering Populations: A New Statistical Methodology*, SIAM, Philadelphia.

[5] Givens, G. H. and Hoeting, J. A. (2005). *Computational Statistics*, Wiley, New York.

[6] Goldstein, H. and Spiegelhalter, D. J. (1996). League tables and their limitations: Statistical issues in comparisons of institutional performance, *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, **159**, 385 - 443.

[7] Govindarajulu, Z. and Harvey, C. (1971). Bayesian procedures for ranking and selection problems, *Annals of the Institute of Statistical Mathematics*, **26**, 35 - 53.

[8] Gupta, S. S. and Panchapakesan, S. (2002). *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations*, SIAM, Philadelphia.

[9] Hall, P. and Miller, H. (2009). Using the bootstrap to quantify the authority of an empirical ranking, *The Annals of Statistics*, **37**, 3929 - 3959.

[10] Hasselman, B. (2010). nleqslv: Solve systems of non linear equations, *R package version 1.8*, URL `http://CRAN.R-project.org/package=nleqslv`.

[11] Laird, N. M. and Louis, T. A. (1989). Empirical Bayes ranking methods, *Journal of Educational Statistics*, **14**, 29 - 46.

[12] Lindley, D. V. (2006). Bayesian inference, *Encyclopedia of Statistical Sciences*, 2nd Ed., (Kotz, S., Balakrishnan, N., Read, C.B. and Vidakovic, B., editors), vol. 1, Wiley, New York, pp. 410 - 418.

[13] Louis, T. A. (1984). Estimating a population of parameter values using Bayes and empirical Bayes methods, *Journal of the American Statistical Association*, **79**, 393 - 398.

[14] R Development Core Team (2010). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0, URL `http://www.R-project.org`.

[15] Xie, M., Singh, K. and Zhang, C. H. (2009). Confidence intervals for population ranks in the presence of ties and near ties, *Journal of the American Statistical Association*, **104**, 775 - 787.