Credible intervals for the Fay-Herriot model

Nadarajasundaram Ganesh
Department of Mathematics, University of Maryland at College Park

Abstract

In this paper, we fill in an important research gap in the small area literature, namely the problem of multiple comparison. For the Fay-Herriot model, we illustrate how the Bayesian approach can be applied to develop different multiple comparison procedures. In the context of multiple comparison, we derive a new class of priors. This class includes the well-known uniform or *superharmonic* prior. Through data analysis we illustrate the use of our class of priors.

KEY WORDS: Bayesian multiple comparison, Fay-Herriot model, hierarchical Bayes, matched priors.

1. Introduction

For effective planning of health, social and other services, and for apportioning government funds, there is a growing demand among various government agencies such as the U.S. Census Bureau, U.K. Central Statistical Office, and Statistics Canada to produce reliable estimates for smaller subpopulations, called small areas. For example, in both developed and developing countries, governmental policies increasingly demand income and poverty estimates for small areas. In fact, in the U.S. more than $130 billion of federal funds per year are allocated based on these estimates.

A sample survey designed for a large population may select a small number of elements or even no element for the small area of interest. Other nonsampling errors such as nonresponse may further reduce sample size for the small area. Thus standard design-based methods that are based solely on the survey data generally fail to provide small area estimates with the desired precisions. Over the last two decades, different model-based approaches that *borrow strength* from related resources have been proposed in the literature. Such methods essentially use explicit models to combine information from the sample survey, various administrative or census records, and even previous surveys. Both Bayesian and non-Bayesian methods were considered to address the point estimation, the associated problem of measuring uncertainty of the point estimator, and the interval estimation. For a good review on small area estimation, readers are referred to the well-known books by Rao (2003) and Longford (2005), and the recent review papers by Pfeffermann (2002), Rao (2005) and Jiang and Lahiri (2006). For Bayesian methods in small area estimation and finite population sampling, see Ghosh and Meeden (1997).

A researcher in public health may report an estimate of the mean body mass index and the associated 95% individual confidence interval for each domain formed by different demographic groups (e.g., for different race×gender× age-group combinations), and then use these individual confidence intervals to find significant difference among pairs of domains. The problem with the above approach, often referred to as *data snooping*, is that even if a table of estimates of the domain mean differences and their associated 95% (individual) confidence intervals are reported for all possible pairs, the confidence level refers to a single comparison and not to a series of comparisons. In fact, the overall confidence level, i.e. the probability that all confidence intervals cover their respective true values, could be much lower than the nominal 95% level. The problem of finding spurious significance results due to data snooping is referred to as the problem of *multiple comparison.*

Exploratory data analysis is a useful part of any scientific investigation, but any claim suggested by such analysis should be validated by an appropriate statistical procedure. Multiple comparison is the most common data snooping problem encountered in small area research. The literature on multiple comparison for linear models is huge, and readers are referred to the books by Hochberg and Tamhane (1987) and Miller (1991). To the best of our knowledge, the multiple comparison problem has not been addressed in the small area literature [at least we have not found any discussions on multiple comparison in the books by Rao (2003) and Longford (2005) or in any of the recent papers except for the discussions and rejoinder section of Jiang and Lahiri (2006)].

In Section 3, using the celebrated Fay-Herriot model, we demonstrate how the Bayesian method can be adapted to address the multiple comparison problem. The Bayesian method is conceptually straightforward. Once the posterior distribution of the parameter(s) of interest is found, this is used in all inferential purposes. The Bayesian approach is attractive since it can incorporate all sources of uncertainties, irrespective of the small area sample size, and can make inferences conditional on the data. However, the choice of prior for the hyperparameters is important and small area inference could be sensitive to the prior choice. See Chen (2001) and Pfeffermann (2006), among others. Thus, one important step in the Bayesian approach is the choice of the prior distribution for the hyperparameter(s).

Morris and Christiansen (1995) used a flat (Lebesgue measure) prior distribution for the regression coefficients, and assumed the prior variance to be independent of the regression coefficients and uniformly distributed over the positive part of the real line. These prior distributions for the hyperparameters are simple to interpret to a nonstatistician and are often recommended. See Berger (1985) and Morris (1983b). The uniform prior for the prior variance, often referred to as Stein's *superharmonic* prior, is noninformative and is known to provide admissible minimax procedures. Unless more information on the hyperparameters is available, these simple prior distributions for the hyperparameters give good frequentist properties to the resulting rules (Morris and Christiansen 1995).

In section 4, we provide an interesting frequentist validation of the well-known superharmonic prior. We show that a weighted average of the posterior variances, under the superharmonic prior, is a second-order unbiased estimator of the corresponding weighted average of the mean squared error (MSE) of the empirical best linear unbiased predictors (EBLUP) or empirical Bayes, the average being taken over all small areas and the weight for a given area being proportional to the inverse of the squared sampling variance. In fact, for a general class of weights, we can obtain a class of priors that satisfies a desirable frequentist property. We call a prior *average moment matching prior* when we choose equal weight for all areas (i.e., simple average). It is interesting to note that all the priors in the class are identical when the sampling variances are all equal; but they could provide considerably different results especially when the sampling variances are very different. The prior suggested by Datta et al. (2005) is another important member of this class. Their prior pays attention to a specific small area by matching the posterior variance of a specific small area mean with the mean squared error (MSE) of the corresponding EBLUP. Thus, their prior is area specific and certainly makes sense when the problem is one of measuring the uncertainty of a specific small area. However, for the multiple comparison problem, we do not believe that it is an appropriate prior since we need to pay attention to all areas. The superharmonic prior and the average moment matching prior seem suitable for the multiple comparison problem.

In Section 5, we propose a Monte Carlo method that generates the hyperparameters from the posterior distribution. The generated hyperparameters are used in the implementation of our proposed multiple comparison method. We note that for the Fay-Herriot model we can save some computing time by considering a Monte Carlo method instead of the usual Monte Carlo Markov Chain method.

In Section 6, we analyze two well-known data sets which were analyzed by Morris and Christiansen (1995) for a different problem. These two examples clearly demonstrate the utility of an appropriate multiple comparison method. We also note that in both the examples, the average moment matching prior and the super-harmonic prior yield similar results. This is because the weights that generate the super-harmonic prior are not very variable in these two examples.

2. The Fay-Herriot Model

In order to estimate the per-capita income of small places (population less than 1000), Fay and Herriot (1979) used the following two-level Bayesian model:

The Fay-Herriot model:

- Level 1: $y_i|\theta_i \overset{\text{ind}}{\sim} N(\theta_i, D_i),\ i = 1, \cdots, m;$
- Level 2: $\theta_i \overset{\text{ind}}{\sim} N(x_i'\beta, \psi),\ i = 1, \cdots, m.$

In the above model, level 1 is used to account for the sampling variability of the regular survey estimates y_i of true small area means. Level 2 links the true small area means θ_i to a vector of k known auxiliary variables x_i, often obtained from various administrative and census records. The parameters β and ψ are generally unknown and are estimated from the available data. In order to estimate the sampling variability D_i, Fay and Herriot (1979) employed the generalized variance function (GVF; see Wolter 1985, Chapter 5) method that uses some external information from the survey.

We note that the Fay-Herriot model can be viewed as an area level mixed regression model:

$$y_i \ = \ \theta_i + e_i = x_i'\beta + v_i + e_i,\ i = 1, \cdots, m,$$

where v_i's and e_i's are independent with $v_i \overset{iid}{\sim} N(0, \psi)$ and $e_i \overset{ind}{\sim} N(0, D_i)$. Note that the area-specific random effect v_i is used to relate the true per-capita income (θ_i) to the auxiliary variables (x_i) obtained from the census, housing and Internal Revenue Service records. In other words, Fay and Herriot (1979) used random effects in order to capture the additional area-specific effects not explained by the area-specific auxiliary variables. This is achieved at the expense of an additional unknown variance component ψ to be estimated from the data. In contrast, the corresponding regression model without random effects fails to capture this additional area-specific variability. Using the U.S. census data, Fay and Herriot (1979) demonstrated that their EB estimator [also an empirical best linear unbiased predictor (EBLUP)] performed better than the direct survey estimator and a synthetic estimator used earlier by the U.S. Census Bureau.

There are a variety of applications of the Fay-Herriot model. In the context of census undercount, several researchers found the Fay-Herriot model useful. See, for example, Cressie (1992), Dick (1995), among others. Estimates of the number of poor school-age children for the U.S. counties and states are produced using the Fay-Herriot model. See the report prepared by the National Research Council (2000). For methodological research, the Fay-Herriot model is often used. See, for example, Glickmann and Pfeffermann (2004), Datta et al. (2005), among others. Particular cases of the Fay and Herriot (1979) can be found in the baseball data example of Efron and Morris (1975) and the false alarm probability estimation example of Carter and Rolph (1974). For example, in the baseball data example, $x_i'\beta = \mu$, i.e., θ_i were treated as exchangeable, and $D_i = 1$. In both the examples, the well-known arc-sine transformation on the sample proportion was taken, which justified the assumption of known sampling variance D_i of the transformed proportion. Like Fay and Herriot (1979), these two applications implement mixed models using the empirical Bayes method.

Suppose we are interested in finding a $100(1-\alpha)\%$ credible interval for a specific $\ell'\theta$, where ℓ is a known $m \times 1$ column vector. The Bayesian approach is straightforward. We simply find the posterior distribution of $\ell'\theta$ and use this to find the desired credible interval. To illustrate the method, first assume ψ is known, but β unknown. We put a flat prior on β, i.e. $\pi(\beta) \propto 1$. It follows that $\theta \mid y \sim N(\Lambda\nu, \Lambda)$, where $y = (y_1, \ldots, y_m)'$, $\theta = (\theta_1, \ldots, \theta_m)'$, $\nu = (\frac{y_1}{D_1}, \ldots, \frac{y_m}{D_m})'$, $X = (x_1, \ldots, x_m)'$, and $\Lambda^{-1} = diag(\frac{1}{D_1} + \frac{1}{\psi}, \ldots, \frac{1}{D_m} + \frac{1}{\psi}) -$

$\frac{X(X'X)^{-1}X'}{\psi}$.

A $100(1-\alpha)\%$ credible interval for $\ell'\theta$ is given by

$$\ell' \Lambda\nu \pm \{\ell' \Lambda\ell\chi^2_{(\alpha,1)}\}^{1/2}, \tag{1}$$

where $\chi^2_{(\alpha,1)}$ is the upper α percentage point of the chi-squared distribution with 1 degree of freedom.

When ψ is unknown, we need to put priors on both β and ψ. We assume $\pi(\beta,\psi) = \pi(\beta)\pi(\psi) \propto \pi(\psi)$ (see Section 4 for a discussion on prior selection for ψ). In this case, a closed-form density for

$$T^{(1)} = \frac{\{\ell'(\theta - E(\theta \mid y))\}^2}{\ell' Var(\theta \mid y)\ell} \mid y$$

cannot be obtained. Hence, a Monte Carlo method is used to construct a credible interval for $\ell'\theta$. The method is as follows: For large R, independently simulate $(\theta_{(1)},\beta_{(1)},\psi_{(1)}),\ldots,(\theta_{(R)},\beta_{(R)},\psi_{(R)}) \sim f(\theta,\beta,\psi \mid y)$. Then $E(\theta \mid y)$ and $Var(\theta \mid y)$ are approximated by

$$\begin{aligned} E(\theta \mid y) &= \bar{\theta}_{(.)} = \frac{1}{R}\sum_{i=1}^{R}\theta_{(i)}, \\ Var(\theta \mid y) &= \frac{1}{(R-1)}\sum_{i=1}^{R}(\theta_{(i)} - \bar{\theta}_{(.)})(\theta_{(i)} - \bar{\theta}_{(.)})'. \end{aligned}$$

Also, $T^{(1)}_\alpha$, the upper α percentage point of the distribution of $T^{(1)}$, is given by the upper α percentage point of the ordered values $T^{(1)}_{(i)}$ $(i = 1,\ldots,R)$, where

$$T^{(1)}_{(i)} = \frac{\{\ell'(\theta_{(i)} - E(\theta \mid y))\}^2}{\ell' Var(\theta \mid y)\ell}.$$

When ψ is unknown, a $100(1-\alpha)\%$ credible interval for $\ell'\theta$ is given by

$$\ell' E(\theta \mid y) \pm \{\ell' Var(\theta \mid y)\ell T^{(1)}_\alpha\}^{1/2}. \tag{2}$$

3. Multiple Comparison

We are interested in constructing simultaneous $100(1-\alpha)\%$ credible intervals, say I_ℓ, for $\ell'\theta$ for all $\ell \in L$, where $L \subset R^m$, the $m-$dimensional Euclidean space. That is, we want

$$P[\ell'\theta \in I_\ell \ for\ all\ \ell \in L|y] = 1-\alpha,$$

where the probability is with respect to the posterior distribution of $\theta = (\theta_1,\ldots,\theta_m)'$ given $y = (y_1,\ldots,y_m)'$.

If one were to use (1) [when ψ is known] or (2) [when ψ is unknown] for multiple comparison, then the overall coverage probability will be much lower than the nominal $100(1-\alpha)\%$. Hence the need for our method. We can, of course, propose efficient Bayesian multiple comparison procedures. The efficiency of the procedure depends on the nature of the class L. In the following three subsections, we discuss multiple comparison procedures for three useful classes.

3.1 Pairwise comparison

Here we are only interested in constructing simultaneous credible intervals for all pairwise comparisons. We will restrict attention to the case where ψ is unknown. A Bayesian version of the Tukey's simultaneous confidence intervals can be used. Define

$$T^{(2)} \equiv \max_k\{(\theta_k - E(\theta_k \mid y)) \mid y\} - \min_k\{(\theta_k - E(\theta_k \mid y)) \mid y\}.$$

Note that $\forall i,j$,

$$|\{(\theta_i - E(\theta_i \mid y)) \mid y\} - \{(\theta_j - E(\theta_j \mid y)) \mid y\}| \le T^{(2)}.$$

Hence,

$$P\{\forall i,j, \quad |(\theta_i - E(\theta_i \mid y)) - (\theta_j - E(\theta_j \mid y))| \le T^{(2)}_\alpha \mid y\} \ge 1-\alpha,$$

where $T^{(2)}_\alpha$ is the upper α percentage point of the distribution of $T^{(2)}$. Simultaneous $100(1-\alpha)\%$ credible intervals for all pairwise comparisons, $\theta_i - \theta_j$, are given by

$$E(\theta_i \mid y) - E(\theta_j \mid y) \pm T^{(2)}_\alpha,$$

where, as before, Monte Carlo is used to compute $E(\theta_i \mid y)$, $E(\theta_j \mid y)$, $T^{(2)}_\alpha$.

3.2 Multiple comparison for all contrasts

Here we concentrate on all possible contrasts in θ (i.e. $\ell'\theta$ s.t. $\sum_{i=1}^m \ell_i = 0$). Define

$$T^{(3)} \equiv (\theta - E(\theta \mid y))'\Big\{\{Var(\theta \mid y)\}^{-1} - \frac{\{Var(\theta \mid y)\}^{-1}11'\{Var(\theta \mid y)\}^{-1}}{1'\{Var(\theta \mid y)\}^{-1}1}\Big\}(\theta - E(\theta \mid y)) \,|\, y.$$

Note that subject to the constraint $\sum_{i=1}^m \ell_i = 0$,

$$\max_\ell \frac{(\ell'(\theta - E(\theta \mid y))^2}{\ell' Var(\theta \mid y)\ell} \,|\, y = T^{(3)}, \tag{3}$$

where 1 is a $m \times 1$ column vector of 1's. When ψ is known, $T^{(3)} \sim \chi^2_{(m-1)}$. Thus simultaneous $100(1-\alpha)\%$ credible intervals for all $\ell'\theta$ such that $\sum_{i=1}^m \ell_i = 0$ are given by

$$\ell' \Lambda\nu \pm \{\ell' \Lambda\ell\chi^2_{(\alpha,m-1)}\}^{1/2}.$$

When ψ is unknown, Monte Carlo is used to compute $E(\theta \mid y)$, $Var(\theta \mid y)$, $T^{(3)}_\alpha$, and in this case simultaneous $100(1-\alpha)\%$ credible intervals for all $\ell'\theta$ s.t. $\sum_{i=1}^m \ell_i = 0$ are given by

$$\ell' E(\theta \mid y) \pm \{\ell' Var(\theta \mid y)\ell T^{(3)}_\alpha\}^{1/2}.$$

3.3 Multiple comparison for all $\ell'\theta$

Note that

$$T^{(4)} \equiv (\theta - E(\theta \mid y))'\{Var(\theta \mid y)\}^{-1}(\theta - E(\theta \mid y)) \,|\, y = \max_{\ell} \frac{(\ell'(\theta - E(\theta \mid y))^2)}{\ell' Var(\theta \mid y)\ell} \,|\, y.$$

When ψ is known, $T^{(4)} \sim \chi^2_{(m)}$. Thus simultaneous $100(1-\alpha)\%$ credible intervals for $\ell'\theta$ for all $\ell \in R^m$ are given by

$$\ell' \Lambda\nu \pm \{\ell' \Lambda \ell \chi^2_{(\alpha,m)}\}^{1/2}.$$

When ψ is unknown, Monte Carlo is used to compute $E(\theta \mid y)$, $Var(\theta \mid y)$, $T^{(4)}_{\alpha}$, and in this case simultaneous $100(1-\alpha)\%$ credible intervals for all $\ell'\theta$ for all $\ell \in R^m$ are given by

$$\ell' E(\theta \mid y) \pm (\ell' Var(\theta \mid y)\ell T^{(4)}_{\alpha})^{1/2}.$$

4. Prior Selection

There are several ways one can choose the prior distribution for ψ. A popular choice is Stein's superharmonic prior distribution given by

$$\pi(\psi) \propto 1.$$

The above choice of prior is non-informative and is known to provide an admissible procedure in the context of point estimation [Morris and Christiansen 1995]. The superharmonic prior was also used by Morris (1983a) in obtaining a suitable measure of uncertainty of his empirical Bayes estimator. In this section, we provide a class of priors satisfying a good frequentist property and show that the superharmonic prior is a member of this class, providing an interesting interpretation of the superharmonic prior.

Given $\{w_i \geq 0,\ i = 1, \cdots, m, \ni \sum_{i=1}^m w_i = 1\}$, we seek a prior $\pi(\psi)$ which satisfies the following condition:

$$\sum_{i=1}^m w_i E\{V(\theta_i \mid y) - MSE[\hat{\theta}_i(\hat{\psi})]\} \;=\; o(1/m), \quad (4)$$

where $V(.|y)$ is the variance under the prior $\pi(\psi)$, $E(\cdot)$ and $MSE(\cdot)$ are taken with respect to the Fay-Herriot model; $\hat{\theta}_i(\hat{\psi})$ is the EBLUP of θ_i, i.e. $\hat{\theta}_i(\hat{\psi}) = x_i'\tilde{\beta}(\hat{\psi}) + \frac{\hat{\psi}}{(\hat{\psi}+D_i)}(y_- x_i'\tilde{\beta}(\hat{\psi}))$, $\tilde{\beta}(\psi) = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$ is the BLUE of β, $\Sigma = diag(D_1+\psi, \ldots, D_m+\psi)$, and $\hat{\psi}$ is the REML estimator of ψ.

In order to satisfy (4), $\pi(\psi)$ must satisfy the following differential equation [see Ganesh (2007)]

$$\frac{d\pi(\psi)}{d\psi}\frac{1}{\pi(\psi)} - 2\frac{\sum_{i=1}^m w_i D_i^2/(D_i+\psi)^3}{\sum_{i=1}^m w_i\{D_i/(D_i+\psi)\}^2} + 2\frac{\sum_{i=1}^m 1/(D_i+\psi)^3}{\sum_{i=1}^m 1/(D_i+\psi)^2} = 0. \quad (5)$$

It can be checked that the solution to (5) is given by

$$\pi(\psi) \propto \frac{\sum_{i=1}^m 1/(D_i+\psi)^2}{\sum_{i=1}^m w_i\{D_i/(D_i+\psi)\}^2}. \quad (6)$$

When the prior is given by (6), it can be easily checked that for $m+2 > rank(X)$ the posterior distribution of θ is proper. It is interesting to note that Stein's super-harmonic prior is a special case of (6), simply take $w_i = \frac{1/D_i^2}{\sum_{j=1}^m 1/D_j^2}$. By taking $w_i = 1/m$ (for $i = 1, \ldots, m$), we get the average moment matching prior given by

$$\pi(\psi) \propto \frac{\sum_{i=1}^m 1/(D_i+\psi)^2}{\sum_{i=1}^m \{D_i/(D_i+\psi)\}^2}. \quad (7)$$

The prior given by (7) has the property that the average posterior variance of θ_i is second-order unbiased for the average MSE of the EBLUP of θ_i. Also, by taking $w_j = 1$ for $j = i$, and $w_j = 0$ for $j \neq i$, we get a prior obtained by Datta et al. (2005). Their main motivation was to choose a prior distribution for ψ such that the posterior variance of θ_i is second-order unbiased for the mean squared error of the EBLUP of θ_i, i.e.

$$E\{V(\theta_i \mid y)\} = MSE\{\hat{\theta}_i(\hat{\psi})\} + o(1/m). \quad (8)$$

Datta et al. (2005) showed that the prior which satisfies (8) is given by

$$\pi(\psi) \propto \ (D_i+\psi)^2 \sum_{j=1}^m \frac{1}{(D_j+\psi)^2}. \quad (9)$$

Note that unless for $i = 1, \ldots, m$, $D_i = D$, the prior for ψ is area specific and hence it is not possible to select a prior which satisfies (9) simultaneously for $i = 1, \ldots, m$.

5. Implementation of the proposed method by Monte Carlo

It is easy to show that

$$f_{\psi|y}(\psi \mid y) \propto \pi(\psi)\,\frac{\exp(-\frac{1}{2}y'(\Sigma^{-1} - \Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1})y)}{|X'\Sigma^{-1}X|^{1/2}\prod_{i=1}^m (D_i+\psi)^{1/2}}$$

$$\beta \mid \psi, y \sim N((X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y, (X'\Sigma^{-1}X)^{-1})$$
$$\theta \mid \beta, \psi, y \sim N(\Gamma\delta, \Gamma)$$

where $\Sigma = diag(D_1+\psi, \ldots, D_m+\psi)$, $\Gamma = diag(\frac{D_1\psi}{D_1+\psi}, \ldots, \frac{D_m\psi}{D_m+\psi})$, $\delta = \frac{X\beta}{\psi} + diag(\frac{1}{D_1}, \ldots, \frac{1}{D_m})y$ and $|X'\Sigma^{-1}X|$ is the determinant of $X'\Sigma^{-1}X$.

We need to generate $(\theta_*, \beta_*, \psi_*)$ from $f(\theta, \beta, \psi \mid y)$. To this end, note that

$$f(\theta, \beta, \psi \mid y) \propto f_{\psi|y}(\psi \mid y) f(\beta \mid \psi, y) f(\theta \mid \beta, \psi, y).$$

Hence $(\theta_*, \beta_*, \psi_*)$ will be generated as follows: $\psi_* \sim f_{\psi|y}(\psi \mid y)$, $\beta_* \sim f(\beta \mid \psi_*, y)$, $\theta_* \sim f(\theta \mid \beta_*, \psi_*, y)$. Simulating $\beta_* \sim f(\beta \mid \psi_*, y)$ and $\theta_* \sim f(\theta \mid \beta_*, \psi_*, y)$ is straightforward. To simulate $\psi_* \sim f_{\psi|y}(\psi \mid y)$, use the following accept-reject method:

1. Simulate $z \sim \chi^2_{(m-k-2)}$ [where $k = rank(X)$].

2. Compute $u = \frac{y^{'}(I-X(X^{'}X)^{-1}X^{'})y}{z} - \sigma^2$. If $u \geq 0$, then $u \sim f_U(u)$, where

$$f_U(u) \propto \frac{\exp(-\frac{1}{2(\sigma^2+u)}y^{'}(I - X(X'X)^{-1}X^{'})y)}{(\sigma^2+u)^{(m-k)/2}} I_{[u\geq 0]}.$$

   σ^2 is chosen s.t. the acceptance rate in the accept-reject method is maximized.

3. Generate $v \sim$ Unif$[0, 1]$.

4. Check if $\frac{1}{M}\frac{f_{\psi|y}(u|y)}{f_U(u)} \geq v$, where $M = \max_t \frac{f_{\psi|y}(t|y)}{f_U(t)}$. If true, then $u \sim f_{\psi|y}(\psi \mid y)$.

6. Data Analysis

In this section, we use two well-known data sets to illustrate to what extent the theoretically valid methods for multiple comparison differ from the naive comparison based on individual confidence intervals. In our study, we include both pairwise comparisons and comparisons of general contrasts. The other purpose of our study is to compare the average moment matching prior (7) with that of the Stein's superharmonic prior.

In our data analysis, we use the baseball run scoring data and the hospital graft failure data given in Morris and Christiansen (1995). The baseball data set (Table 1) gives the average runs scored per game and sample standard deviation of 14 baseball teams in the American League for the year 1993. Each of the teams played 162 games, and y_i denotes the average runs scored over those 162 games. A good approximation for the variance of runs scored for a single game is given by $V(\mu) = (1.375\mu)^{1.2}$, where μ is the mean runs scored for a single game. For the 162 games played, the variance D_i for the i^{th} team is then approximated by $D_i = \frac{V(y_i)}{162} = \frac{(1.375y_i)^{1.2}}{162}$, and is assumed to be known without error. The normality assumption for y_i is justified by the central limit theorem. The estimates of the true runs per game θ_i and its standard error given in Table 1 was computed using $20,000$ independent samples for each of the two priors.

The second data set (Table 2) gives a sample of 23 hospitals (of the 219 hospitals) which had at least 50 kidney transplants during a 27 month period. The y_i's are graft failure rates for kidney transplant operations, i.e. y_i = number of graft failures/n_i, where n_i is the number of kidney transplants at hospital i over the period of interest. The variance for graft failure rate D_i is approximated by $(0.2)(0.8)/n_i$, where 0.2 is the observed failure rate for all hospitals. Again, D_i is assumed to be known. In addition, a severity index x_i is available for each hospital. x_i is the average fraction of females, blacks, children and extremely ill kidney recipients at hospital i. It is thought that this severity index increases graft failure rate, and hence is included as a covariate. Once again the central limit theorem is used to approximate the distribution of y_i. The estimates of the true graft failure rates θ_i and its standard error given in Table 2 was computed using $20,000$ independent samples for each of the two priors.

For the Baseball data set, Tables 3-4 give 95% credible intervals for a few contrasts of interest. Note that when an appropriate multiple comparison method is used, the coverage probability holds simultaneously for all contrasts or pairwise comparisons. If instead, before looking at the data, a practitioner decides that a specific $\ell'\theta$ is the only contrast of interest, then a much shorter interval can be obtained by using (2). As can be seen from Tables 3-4, in a number of instances, after looking at the data if a practitioner were to naively use (2), he/she would incorrectly reject the null hypothesis $H_o : \ell^{'}\theta = 0$ when it should be accepted.

It is interesting to note that the average moment matching prior gives very similar results to the ones obtained when the superharmonic prior is used. This is because in both data sets there is little variability in the sample variances. Hence, the weights $w_i = \frac{1/D_i^2}{\sum_{j=1}^m 1/D_j^2}$ that generate the superharmonic prior are more or less uniform across areas.

7. Concluding Remarks

The paper offers a simple Bayesian solution to the multiple comparison problem for small area estimation, a problem not addressed in the small area literature. We have discussed the important problem of prior selection and obtained a new class of prior distributions for the hyperparameter ψ. In this context, we find an interesting frequentist validation of Stein's superharmonic prior, a prior frequently used in Bayesian analysis. We have not addressed the classical solution to the multiple comparison problem for small area estimation. This will be a good topic for future research.

REFERENCES

Berger, J. (1985), *Statistical Decision Theory and Bayesian Analysis,* 2nd ed. New York: Springer-Verlag.

Carter, G.M., and Rolph, J.E. (1974), Empirical Bayes methods applied to estimating fire alarm probabilities, *J. Amer. Statist. Assoc.*, **69**, 880-885.

Chen, S. (2001), Empirical Best Prediction and Hierarchical Bayes Methods in Small-Area Estimation, Ph.D. Dissertation, Department of Mathematics and Statistics, University of Nebraska-Lincoln.

Cressie, N. (1992), REML Estimation in Empirical Bayes Smoothening of Census Undercount, *Survey Methodology*, **18**, 75-94.

Datta, G.S., Rao, J.N.K. and Smith, D.D. (2005), On measuring the variability of small area estimators under a basic area level model, *Biometrika*, **92**, 183-196.

Dick, P. (1995), Modeling Net Undercoverage in the 1991 Census, *Survey Methodology*, **21**, 45-54.

Efron, B., and Morris, C. (1975), Data analysis using Stein's estimator and its generalizations, *J. Amer. Statist. Asso.*, **70**, 311-319.

Fay, R. E. and Herriot, R. A. (1979), Estimates of income for small places: an application of James-Stein procedures to census data, *J. Amer. Statist. Assoc.*, **74**, 269-277.

Ganesh, N. (2007), Small area estimation and prediction problems: spatial models, Bayesian multiple comparison and robust MSE estimation, Ph.D. Dissertation, Department of Mathematics, University of Maryland at College Park.

Ghosh, M. and Meeden, G. (1997), *Bayesian Methods for Finite Population Sampling*, London: Chapman and Hall.

Table 1: Estimates of the true runs/game and its s.e., using the superharmonic prior (columns 5 & 6) and average moment matching prior (columns 7 & 8).

Obs	Team	y_i	$\sqrt{D_i}$	θ_i	s_i	θ_i^*	s_i^*
1	Det	5.549	0.266	5.287	0.250	5.290	0.250
2	Tor	5.228	0.257	5.070	0.227	5.073	0.230
3	Tex	5.154	0.254	5.022	0.225	5.021	0.226
4	NY	5.068	0.252	4.962	0.221	4.961	0.221
5	Cle	4.877	0.246	4.827	0.214	4.829	0.214
6	Bal	4.852	0.245	4.808	0.212	4.809	0.211
7	Chi	4.790	0.243	4.765	0.210	4.764	0.211
8	Sea	4.531	0.235	4.570	0.205	4.573	0.205
9	Mil	4.525	0.235	4.569	0.206	4.567	0.206
10	Oak	4.414	0.232	4.483	0.207	4.486	0.205
11	Min	4.278	0.227	4.379	0.205	4.381	0.205
12	Bos	4.235	0.226	4.346	0.205	4.348	0.205
13	Cal	4.222	0.226	4.336	0.204	4.337	0.205
14	KC	4.167	0.224	4.293	0.208	4.294	0.205

Table 2: Estimates of graft failure rates and its s.e., using the superharmonic prior (columns 5 & 6) and average moment matching prior (columns 7 & 8).

Obs	y_i	$\sqrt{D_i}$	x_i	θ_i	s_i	θ_i^*	s_i^*
1	0.302	0.055	0.112	0.225	0.037	0.223	0.037
2	0.140	0.053	0.206	0.193	0.035	0.194	0.033
3	0.203	0.052	0.104	0.191	0.032	0.191	0.032
4	0.333	0.052	0.168	0.250	0.037	0.248	0.037
5	0.347	0.047	0.337	0.294	0.039	0.292	0.039
6	0.216	0.046	0.169	0.210	0.030	0.211	0.029
7	0.156	0.046	0.211	0.195	0.032	0.197	0.031
8	0.143	0.046	0.195	0.186	0.032	0.189	0.031
9	0.220	0.044	0.221	0.222	0.030	0.223	0.029
10	0.205	0.044	0.077	0.189	0.031	0.188	0.030
11	0.209	0.042	0.195	0.213	0.028	0.213	0.028
12	0.266	0.041	0.185	0.236	0.030	0.235	0.030
13	0.240	0.041	0.202	0.228	0.029	0.227	0.028
14	0.262	0.036	0.108	0.224	0.030	0.223	0.030
15	0.144	0.036	0.204	0.182	0.029	0.183	0.029
16	0.116	0.035	0.072	0.145	0.029	0.146	0.028
17	0.201	0.033	0.142	0.200	0.025	0.200	0.025
18	0.212	0.032	0.136	0.205	0.024	0.204	0.024
19	0.189	0.031	0.172	0.198	0.024	0.198	0.024
20	0.212	0.029	0.202	0.214	0.023	0.214	0.023
21	0.166	0.029	0.087	0.172	0.023	0.172	0.023
22	0.173	0.027	0.177	0.187	0.023	0.188	0.022
23	0.165	0.025	0.072	0.169	0.021	0.169	0.021

Table 3: (Baseball data) Credible intervals for selected contrasts using the superharmonic prior.

Contrast	All contrasts	Pairwise	Individual
$\theta_1 - \theta_{14}$	(-0.691,2.680)	(-0.026,2.015)	(0.341,1.682)
$\theta_2 - \theta_{14}$	(-0.785,2.339)	(-0.244,1.797)	(0.173,1.417)
$\theta_4 - \theta_{12}$	(-0.887,2.120)	(-0.404,1.637)	(0.034,1.228)
$\theta_5 - \theta_{13}$	(-0.965,1.947)	(-0.529,1.511)	(-0.070,1.084)
$\frac{1}{2}(\theta_2 + \theta_3) - \theta_{13}$	(-0.627,2.045)	not pairwise	(0.189,1.251)
$\frac{1}{3}(\theta_1 + \theta_2 + \theta_3 - \theta_{12} - \theta_{13} - \theta_{14})$	(-0.250,1.852)	not pairwise	(0.382,1.219)

Table 4: (Baseball data) Credible intervals for selected contrasts using the average moment matching prior.

Contrast	All contrasts	Pairwise	Individual
$\theta_1 - \theta_{14}$	(-0.666,2.657)	(-0,027,2.018)	(0.348,1.667)
$\theta_2 - \theta_{14}$	(-0.785,2.343)	(-0.244,1.801)	(0.176,1.419)
$\theta_4 - \theta_{12}$	(-0.900,2.125)	(-0.410,1.636)	(0.034,1.230)
$\theta_5 - \theta_{13}$	(-0.965,1.949)	(-0.531,1.515)	(-0.077,1.088)
$\frac{1}{2}(\theta_2 + \theta_3) - \theta_{13}$	(-0.636,2.055)	not pairwise	(0.187,1.253)
$\frac{1}{3}(\theta_1 + \theta_2 + \theta_3 - \theta_{12} - \theta_{13} - \theta_{14})$	(-0.252,1.854)	not pairwise	(0.379,1.217)

Glickman, H. and Pfeffermann, D. (2004), Mean square error approximation in small area estimation by use of parametric and nonparametric bootstrap. In *Proceedings of the Survey Research Methods Section, American Statistical Association*.

Hochberg, Y. and Tamhane, A.C. (1987), *Multiple Comparison Procedures*, New York: John Wiley & Sons.

Jiang, J., and Lahiri, P. (2006), Mixed model prediction and small area estimation (with discussions), *Test*, **15**, 1, 1-96.

Longford, N.T. (2005), *Missing data and small-area estimation*, Springer.

Miller, R.G., Jr. (1991), *Simultaneous Statistical Inference*, New York: Springer-Verlag.

Morris, C.N., and Christiansen, C.L. (1995), Hierarchical Models for Ranking and for Identifying Extremes with Applications, *Bayes Statistics 5*, Oxford: Oxford University Press.

Morris, C.N. (1983a), Parametric empirical Bayes inference: theory and applications, *J. Amer. Statist. Asso.*, **78**, 47-59.

Morris, C.N. (1983b), Parametric empirical Bayes confidence intervals, In *Proceeding on the Conference on Scientific Inference, Data Analysis and Robustness*, New York: Academic Press, 25-50.

National Research Council (2000), Small-Area Estimates of School-Age Children in Poverty: Evaluation of Current Methodology, C.F. Citro and G. Kalton (Eds.) Committee on National Statistics, Washington, DC: National Academy Press.

Pfeffermann, D. (2002), Small area estimation - new developments and directions, *Int. Statist. Rev.*, **70**, 125-143.

Pfeffermann, D. (2006) Discussion of the paper “Mixed model prediction and small area estimation,” *Test*, **15**, 1, 1-96.

Rao, J. N. K. (2003), *Small Area Estimation*, Wiley, New York.

Rao, J.N.K. (2005), Inferential issues in small area estimation: some new developments, *Statistics in Transition*, **7**, 513-526.

Wolter, K. (1985), *Introduction to Variance Estimation*, New York: Springer-Verlag.