Les Cahiers du GERAD ISSN: 0711–2440

A Unified Bayesian Approach to Small Area Estimation of Mean Parameters in Generalized Linear Models

P.J. Farrell, B. MacGibbon,
T.J. Tomberlin

G–2002–53

October 2002

A Unified Bayesian Approach to Small Area Estimation of Mean Parameters in Generalized Linear Models

Patrick J. Farrell
School of Mathematics and Statistics, Carleton University
Ottawa (Ontario) Canada, pfarrell@math.carleton.ca

Brenda MacGibbon
GERAD et Département de mathématiques, Université du Québec à Montréal
Montréal (Québec) Canada, macgibbon.brenda@uqam.ca

Thomas J. Tomberlin
John Molson School of Business, Concordia University
Montréal (Québec) Canada

October, 2002

Les Cahiers du GERAD

G–2002–53

Abstract

In this study, we present a unified Bayesian approach to small area estimation of mean parameters in generalized linear models. The basic idea consists of incorporating into such a model nested random effects that reflect the complex structure of the data in a multistage sample design. However, as compared to the ordinary linear regression model, it is not feasible to obtain a closed form expression for the posterior distribution of the parameters. The approximation most commonly used in empirical Bayes studies is that proposed by Laird (1978), where the posterior is expressed as a multivariate normal distribution having its mean at the mode and covariance matrix equal to the inverse of the information matrix evaluated at the mode. Inspired by the work of Zeger and Karim (1991) and Gu and Li (1998), we also study a stochastic simulation method to approximate the posterior distribution. Alternatively, a hierarchical Bayes approach based on Gibbs sampling similar to Farrell (2000) can also be employed. We present here the results of a Monte Carlo simulation study to compare point and interval estimates of small area proportions based on these three estimation methods.

Key Words: Logistic Regression, Overdispersion, Random Effects, Small Area Proportions.

Résumé

Dans cette étude, on présente une approche bayésienne unifiée pour l’estimation des paramètres de moyenne d’un modèle linéaire généralisé avec de petites aires. L’idée de base consiste à incorporer dans un tel modèle des effets aléatoires emboîtés qui reflètent la structure complexe des données dans un plan d’échantillonnage multiphase. Cependant, en comparaison avec le modèle de régression linéaire, il n’est pas possible d’obtenir une expression analytique pour la distribution a priori des paramètres. L’approximation la plus utilisée dans les études bayésiennes empiriques est celle proposée Laird (1978), où la distribution a priori est approximée par une normale multivariée ayant comme moyenne le mode et comme matrice de covariance l’inverse de la matrice d’information évaluée au mode. Inspirés par les travaux de Zeger et Karim (1991) et de Gu et Li (1998), on étudie aussi une méthode de simulation stochastique pour approximer la distribution a priori. Alternativement on peut aussi employer une approche bayesienne hiérarchique basée sur des échantillonnages de Gibbs similaire à celle de Farrell (2000). On présente ici des résultats d’une étude de simulation de Monte Carlo pour comparer les estimations ponctuelles et par intervalle des proportions de petites aires basées sur ces trois méthodes d’estimation.

1 INTRODUCTION

The estimation of parameters for small areas has received considerable attention of late. Model-based estimators have gained acceptance over direct survey estimators, since the latter are unstable due to the small or nonexistent sample sizes that result from small areas. Excellent summaries of these methodologies can be found in Ghosh and Rao (1994) and Rao (1999).

Model-based estimators borrow strength from related areas and are therefore less variable. Synthetic estimation proposed by Gonzales (1973) was the first of these approaches. However, the resulting estimator has a tendency to be biased. Other model-based estimators have been developed since then in order to deal with the above deficiencies, including those based on empirical and hierarchical Bayes estimators. Ghosh and Rao (1994) illustrated that small area estimates based on such approaches often outperformed other methods such as classical unbiased and synthetic estimation. Farrell et al (1997) reached a similar conclusion in a study that compared empirical Bayes, synthetic, and direct survey estimators. The proven optimality properties of empirical Bayes methods and their documented successful performance have made them popular (See Efron 1998). Farrell et al (1997) noted, however, that interval estimates based on an empirical Bayes technique do not attain the desired level of coverage when a naive approach is employed since the variability arising from estimating the parameters of the prior distribution is not acknowledged. They proposed the use of bootstrap techniques for adjusting these estimates.

Fay and Herriot (1979) were among the first to apply empirical Bayes methods based on linear models to the problem of small area estimation, while Datta and Ghosh (1991) investigated hierachical Bayes models. A nubmer of studies have focused on the specification of the prior distribution. Lahiri and Rao (1995) robustified the Fay-Herriot model by relaxing the assumption of a normal prior distribution. Datta and Lahiri (1995) developed a robust hierarchical Bayes approach for handling outliers.

Several authors have considered the problem of estimating small area rates and binomial parameters using empirical and hierarchical Bayes approaches. See for example, Dempster and Tomberlin (1980), Wong and Mason (1985), Tomberlin (1988), MacGibbon and Tomberlin (1989), Farrell et al (1994, 1997), Stroud (1994), Malec et al (1997), and Farrell (2000). Following the approach of MacGibbon and Tomberlin (1989), Farrell et al (1994) obtained empirical Bayes point estimates of small area proportions, and also provided guidelines for the choice of prior distribution. Farrell et al (1997) extended this approach to interval estimation, while Farrell (2000) considered analogous hierarchical Bayes estimation approaches.

In this study, we present unified empirical and hierarchical Bayes approaches to small area estimation of mean parameters in generalized linear models. This should be contrasted with the work of McCullagh and Nelder (1989) and Breslow and Clayton (1993) who study quasi-likelihood methods of approximate inference for these models. The basic idea considered here consists of incorporating into such a model nested random effects that reflect the complex structure of the data in a multistage sample design. However, as compared to the ordinary linear regression model, it is not feasible to obtain a closed form expression for the posterior distribution of the parameters. The approximation most commonly used in empirical Bayes studies is that originally proposed by Laird (1978), where the posterior is expressed as a multivariate normal distribution having its mean at the mode and covariance matrix equal to the inverse of the information matrix evaluated at the mode. Inspired by the work of Zeger and Karim (1991) on Gibbs sampling and of Gu and Li (1998) on stochastic approximation computing techniques, we also study a stochastic simulation method to approximate the posterior distribution by assuming a normal prior. Alternatively, a hierarchical Bayes approach based on Gibbs sampling similar to Farrell (2000) can also be employed to estimate the posterior. We present here the results of a Monte Carlo simulation study to compare point and interval estimates of small area proportions based on these three estimation methods that is based on data from a two-stage sample design. A normal prior distribution is specified for all three approaches; however this assumption is not necessary.

2 THE MODEL AND ESTIMATION PROCEDURES ON GIBBS SAMPLING

We follow the framework of Dempster and Tomberlin (1980) used previously for small area estimation by MacGibbon and Tomberlin (1989), Farrell et al (1994, 1997), and Farrell (2000). This basic framework is re-formulated to be consistent with the notation of McCullagh and Nelder (1989) for generalized linear models. In these models, the expected value of a variable of interest is expressed as a function of K covariates as well as random sampling characteristics. For example, consider the case of a two-stage cluster sample consisting of samples of individuals within each of I primary sampling units or local areas.

Specifically, let Y_{ij} represent a random variable for the characteristic of interest for the j-th individual within the i-th local area, and let y_{ij} represent a realization of Y_{ij}. If μ_{ij} is the expected value of Y_{ij}, then following McCullagh and Nelder (1989) the μ_{ij} are related to a linear function of covariates and sampling characteristics via a link function g with differentiable inverse h, that is $\eta_{ij} = g(\mu_{ij})$ and $\mu_{ij} = h(\eta_{ij})$. We concentrate here on the estimation of small area characteristics. In this framework, we might be interested in the total of y_{ij}'s for each local area, that is

$$T = \sum_{j} y_{ij} \tag{1}$$

These types of parameters we propose to estimate using a prediction approach based on the linear model

$$\eta_{ij} = \beta_0 + \sum_{k=1}^{K} \beta_k X_{ijk} + \phi_i \tag{2}$$

where ϕ_i represents a random effect associated with the i-th local area, and β_k represents the regression coefficient associated with the k-th covariate.

Let us henceforth illustrate the techniques on a problem where the objective is the development of point and interval estimates for the proportion of individuals, p_i, in each of I small areas that possess a characteristic of interest. The data to be used in obtaining these estimates will be obtained using a two stage sample design, where individuals are sampled from selected small areas. For example, imagine that two choices of surgical procedure, A and B, are available for treatment of a certain disease, and that interest is in the proportion of patients at each of I hospitals (the small areas) that select procedure A. A sample is drawn from each selected hospital, and information is recorded on the choice of procedure for each patient in the sample. In addition, suppose that covariate information on gender and age is available for all patients in each hospital for which point and interval estimates are desired, regardless of whether the individual is sampled or not.

In the framework of the sample design proposed for this example, p_i can be written as

$$p_i = \sum_j y_{ij} \,/\, N_i \tag{3}$$

where N_i is the population size of the i-th local area (hospital), and y_{ij} takes on a value of zero or one, depending upon whether or not the j-th individual in the i-th local area possesses the characteristic of interest. Using a predictive model-based approach proposed by Royall (1970), an estimator for p_i is

$$\hat{p}_i = \Big(\sum_{j \varepsilon S} y_{ij} + \sum_{j \varepsilon S'} \hat{y}_{ij}\Big) \,/\, N_i \tag{4}$$

where the sum over $j \ \varepsilon \ S$ of y_{ij} is the sum of the outcome variable for sampled individuals from the i-th local area, and the sum over $j \ \varepsilon \ S'$ of $\hat{y}_{ij}$ is the sum of the estimated outcome variables for nonsampled individuals in the i-th local area. Values for $\hat{y}_{ij}$ are obtained by initially specifying a model to describe the probability π_{ij} that the j-th individual within the i-th local area possesses the characteristic of interest. Specifically, the model is given by

$$y_{ij} \big| \pi_{ij} \sim i.i.d. \text{ Bernoulli } (\pi_{ij}), \quad \text{logit}\,(\pi_{ij}) = \beta_0 + \sum_{k=1}^{K} \beta_k X_{ijk} + \phi_i = X_{ij}^T \beta + \phi_i \tag{5}$$

so that

$$\pi_{ij} = [1 + \exp(\beta_0 + \sum_{k=1}^{K} \beta_k X_{ijk} + \phi_i)]^{-1} = [1 + \exp(X_{ij}^T \beta + \phi_i)]^{-1} \tag{6}$$

In the notation of McCullagh and Nelder (1989), $\pi_{ij} = \mu_{ij}$, and as indicated above, ϕ_i represents a random effect associated with the ***i***-th small area, which is assumed to follow some specified prior probability distribution. These random effects are included in the model to account for the influence of unobserved covariates on the variation in the p_i. For purposes of the simulation study conducted here, we will assume that

$$\phi_i \sim \textit{i.i.d.}\ \text{Normal}(0, \sigma^2) \tag{7}$$

Once estimates for β and ϕ_i have been determined, π_{ij} is estimated by

$$\hat{\pi}_{ij} = [1 + \exp(\hat{\beta}_0 + \sum_{k=1}^{K} \hat{\beta}_k X_{ijk} + \hat{\phi}_i)]^{-1} = [1 + \exp(X_{ij}^T \hat{\beta} + \hat{\phi}_i)]^{-1} \tag{8}$$

The estimates $\hat{\pi}_{ij}$ in conjunction with (4) ultimately allow for the development of point and interval estimates for the p_i. The approaches employed are described next.

2.1 Empirical Bayes Model Parameter Estimates Based on Classical EM Algorithm

Typically, the random effects variance, σ^2, is unknown. Suppose that a value is assigned to this variance (that will be updated and possibly altered later). For this value of σ^2, the distribution of the data is given by

$$f(y \mid \beta, \phi, \sigma^2)\ \alpha\ \prod_{ij} \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{1 - y_{ij}} \tag{9}$$

where ϕ is a vector containing the random effects ϕ_i. If a flat prior is placed upon β, then the prior distribution of the parameters is

$$f(\beta, \phi \mid \sigma^2)\ \alpha\ \frac{1}{\sigma^I} \exp(-\sum_i \frac{\phi_i^2}{2\sigma^2}) \tag{10}$$

Thus, the joint distribution of the data and the parameters is given by

$$f(y, \beta, \phi \mid \sigma^2)\ \alpha\ \prod_{ij} \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{1 - y_{ij}}\ \frac{1}{\sigma^I} \exp(-\sum_i \frac{\phi_i^2}{2\sigma^2}) \tag{11}$$

so that the posterior distribution of the parameters is

$$f(\beta, \phi \mid y, \sigma^2) = P(y, \beta, \phi \mid \sigma^2) / P(y \mid \sigma^2) \tag{12}$$

It is not possible to obtain a closed form for the expression in (12) due to the intractable integration required to evaluate the denominator on the right hand side. Therefore, according to a proposal by Laird (1978), we approximate (11) by a multivariate normal having its mean at the mode and covariance matrix equal to the inverse of the information matrix evaluated at the mode.

The resulting estimates are conditional on the initially specified value of σ^2, say $\hat{\sigma}^2_{\{0\}}$. Let these estimates be represented by $\hat{\beta}_{\{0\}}$ and $\hat{\phi}_{\{0\}}$. The EM algorithm proposed by Dempster, Laird, and Rubin (1978) can be used to find a maximum likelihood estimate for σ^2. Specifically, using the estimates obtained with $\hat{\sigma}^2_{\{0\}}$, an updated value for σ^2, $\hat{\sigma}^2_{\{1\}}$ is determined using $\hat{\sigma}^2_{\{1\}} = \sum_{i=1}^{I}[\hat{\phi}^2_{i\{0\}} + V\hat{a}r(\hat{\phi}_{i\{0\}})]/I$. If $\hat{\sigma}^2_{\{1\}}$ is approximately equal to $\hat{\sigma}^2_{\{0\}}$, then $\hat{\beta}_{\{0\}}$, $\hat{\phi}_{\{0\}}$, and the associated covariance matrix serve as the empirical Bayes estimates of the model parameters. Alternatively, if these two successive estimates of the random effects variance are deemed to be different, another set of estimates for the fixed and random effects parameters would be computed using the Laird approximation with $\hat{\sigma}^2_{\{1\}}$ as the value specified for the random effects variance. This iterative procedure would continue until successive estimates for σ^2 converge. As above, at the $\{n + 1\}$-th iteration, $\hat{\sigma}^2_{\{n\}}$ would be updated using

$$\hat{\sigma}^2_{\{n+1\}} = \sum_{i=1}^{I}[\hat{\phi}^2_{i\{n\}} + V\hat{a}r(\hat{\phi}_{i\{n\}})]/I \tag{13}$$

2.2 Empirical Bayes Model Estimates Based on Stochastic Simulation

Gu and Li (1998) proposed an alternative iterative approach for estimating the random effects variance that also makes use of the assumption that the posterior distribution in (12) is multivariate normal. It consisted of a stochastic approximation based on the Robbins and Monro (1951) procedure. Here we use a simpler method inspired by the work of Zeger and Karim (1991). At the $\{n + 1\}$-th iteration, $(\hat{\beta}_{\{n\}}, \hat{\phi}_{\{n\}})_{\{1\}}$, $(\hat{\beta}_{\{n\}}, \hat{\phi}_{\{n\}})_{\{2\}}$, …, $(\hat{\beta}_{\{n\}}, \hat{\phi}_{\{n\}})_{\{T\}}$ are generated from this multivariate normal. Then

$$Q(\sigma^2 \big| \hat{\sigma}^2_{\{n\}}) = \iint \log f(y \big| \beta, \phi, \sigma^2) f(\beta, \phi \big| y, \sigma^2 = \hat{\sigma}^2_{\{n\}}) d\beta d\phi \approx \frac{1}{T}\sum_{t=1}^{T} \log[f(y \big| (\hat{\beta}_{\{n\}}, \hat{\phi}_{\{n\}})_{\{t\}}, \hat{\sigma}^2_{\{n\}})] \tag{14}$$

is maximized over σ^2 to obtain $\hat{\sigma}^2_{\{n+1\}}$. If it is assumed that the estimates for σ^2 have converged at iteration *n*, then the *T* values of $(\hat{\beta}_{\{n\}}, \hat{\phi}_{\{n\}})_{\{t\}}$ are averaged to produce the empirical Bayes estimates of the fixed and random effects parameters. The inverse of the information matrix is then evaluated at this average in order to obtain a covariance matrix.

2.3 Empirical Bayes Local Area Estimates

Once the empirical Bayes estimates of the model parameters have been obtained according to Section 2.1 or 2.2, (8) is used to determine a value for $\hat{\pi}_{ij}$ for all $j \ \varepsilon \ S'$ in the i-th local area, then (4) is used to obtain empirical Bayes point estimates of small area proportions by setting $\sum \hat{y}_{ij} = \sum \hat{\pi}_{ij}$. To develop empirical Bayes interval estimates, we consider the mean square error of $\hat{p}_i$ as a predictor for p_i. When $\sum \hat{y}_{ij}$ in (4) is replaced by $\sum \hat{\pi}_{ij}$, this mean square error can be estimated as

$$M\hat{S}E(\hat{p}_i) = V\hat{a}r\left(\frac{\sum_{j\varepsilon S'} \hat{\pi}_{ij}}{N_i}\right) + \frac{\sum_{j\varepsilon S'} \hat{\pi}_{ij}(1-\hat{\pi}_{ij})}{N_i^2} \qquad (15)$$

For sampled local areas, where the sample size, n_i, is greater than zero, the first term in (15) is of order $1 / n_i$, while the second term is of order $1 / N_i$. In this study, the approximation of the mean square error of $\hat{p}_i$ is based on the first term only, yielding a useful approximation so long as N_i is large compared to n_i.

To develop an expression for the variance of $\hat{p}_i$, we let Z_{ij} represent a vector of fixed effects predictor variables for the *ij*-th individual augmented by a series of binary variables, each indicating whether or not the *ij*-th individual belongs to a particular local area. We also let $\hat{\Gamma}$ be the vector containing the estimates of the fixed and random effects parameters. Then $Z_{ij}^T \hat{\Gamma} = X_{ij}^T \hat{\beta} + \hat{\phi}_i$ where $\hat{\beta}$ and $\hat{\phi}_i$ are the empirical Bayes estimates of β and ϕ_i. To obtain an expression for the variance of $\hat{p}_i$, a first order multivariate Taylor series expansion of (4) with $\sum \hat{y}_{ij}$ replaced by $\sum \hat{\pi}_{ij}$ is taken with respect to the realized values of the fixed and random effects estimates, yielding an approximate expression that describes $\hat{p}_i$ as a linear function of these estimates. If the variance of this expression is taken, the result is

$$V\hat{a}r(\hat{p}_i) = \left[\sum_{j\varepsilon S'} Z_{ij}^T \hat{\pi}_{ij}(1-\hat{\pi}_{ij})\right]\left(\frac{\hat{\Sigma}}{N_i^2}\right)\left[\sum_{j\varepsilon S'} Z_{ij}\hat{\pi}_{ij}(1-\hat{\pi}_{ij})\right] \qquad (16)$$

where $\hat{\Sigma}$ represents the covariance matrix of the estimated vector $\hat{\Gamma}$. A 100(1 - α)% naive empirical Bayes confidence interval for p_i can be determined using $\hat{p}_i \pm z_{(1-\alpha/2)}\sqrt{V\hat{a}r(\hat{p}_i)}$, where $z_{(1-\alpha/2)}$ is the 100(1 - α/2) percentile of a standard normal distribution.

However, note that estimates of posterior variances given by (16) do not include the uncertainty due to estimating the prior parameters; hence empirical Bayes confidence intervals based on these variances are often too short to achieve the desired level of coverage. A number of methods for addressing this shortcoming are available. Farrell, MacGibbon, and Tomberlin (1997) made use of a parametric Type III bootstrap proposed by Laird and Louis (1987). Carlin and Gelfand (1991) proposed a modification to this bootstrap. It is this modification that is used in the simulation study in Section 3. For a complete description of the procedure, see Farrell (2000).

2.4 Hierarchical Bayes Model Parameter and Local Area Estimates

To develop hierarchical Bayes estimates for the parameters in the model in (5) requires that a distribution for the random effects variance be specified. If a flat prior is chosen for σ^2, then the joint distribution $P(y, \beta, \phi, \sigma^2)$ is identical to $P(y, \beta, \phi \mid \sigma^2)$ in (11). Using (11) then allows for the determination of marginal posterior distributions of each parameter up to a constant of proportionality (See Gilks, Best, and Tan 1995); the evaluation of the actual distribution is not possible due to the intractable integration required to obtain $P(\beta, \phi, \sigma^2 \mid y)$. Specifically the posterior distribution for any given parameter is proportional to the product of all terms in (11) that contain it, thus yielding the following

$$f(\beta_0 \,\big|\, y, \beta_1, \ldots, \beta_K, \phi, \sigma^2) \;\alpha\; \prod_{ij} \pi_{ij}^{y_{ij}} (1-\pi_{ij})^{1-y_{ij}}$$

$$f(\beta_k \,\big|\, y, \beta_0, \beta_1, \ldots, \beta_{k-1}, \beta_{k+1}, \ldots, \beta_K, \phi, \sigma^2) \;\alpha\; \prod_{ij} \pi_{ij}^{y_{ij}} (1-\pi_{ij})^{1-y_{ij}}$$

$$f(\phi_i \,\big|\, y, \beta, \phi_1, \ldots, \phi_{i-1}, \phi_{i+1}, \ldots, \phi_I, \sigma^2) \;\alpha\; \prod_{ij} \pi_{ij}^{y_{ij}} (1-\pi_{ij})^{1-y_{ij}} \;\; \exp(-\sum_i \frac{\phi_i^2}{2\sigma^2})$$

$$f(\sigma^2 \,\big|\, y, \beta, \phi) \;\alpha\; \frac{1}{\sigma^I} \exp(-\sum_i \frac{\phi_i^2}{2\sigma^2})$$

Under Gibbs sampling, an initial set of values would be assumed as the estimates for β, ϕ, and σ^2, say $\hat{\beta}_{\{0\}}$, $\hat{\phi}_{\{0\}}$, and $\hat{\sigma}^2_{\{0\}}$. An updated estimate for β_0, say $\hat{\beta}_{0\{1\}}$, is obtained by sampling from the distribution $f(\beta_0 \,\big|\, y, \hat{\beta}_{1\{0\}}, \ldots, \hat{\beta}_{K\{0\}}, \hat{\phi}_{\{0\}}, \hat{\sigma}^2_{\{0\}})$. Sampling from $f(\beta_1 \,\big|\, y, \hat{\beta}_{0\{1\}}, \hat{\beta}_{2\{0\}}, \ldots, \hat{\beta}_{K\{0\}}, \hat{\phi}_{\{0\}}, \hat{\sigma}^2_{\{0\}})$ based on $\hat{\beta}_{0\{1\}}$ yields the revised estimate $\hat{\beta}_{1\{1\}}$ for β_1. The completion of a first iteration is realized once the revised estimates $\hat{\beta}_{\{1\}}$, $\hat{\phi}_{\{1\}}$, and $\hat{\sigma}^2_{\{1\}}$ are obtained. This procedure of sampling from full conditional distributions using the most up-to-date revised estimates continues until the estimates of each parameter are deemed to have stabilized from one iteration to the next. See Geman and Geman (1984) and Gelfand and Smith (1990) for a general discussion on Gibbs sampling, and Gelman and Rubin (1992) for methods of convergence.

Note that a different full conditional distribution must be sampled every time a new estimate is obtained, regardless of which parameter is being estimated. Since many iterations are usually needed to ensure that estimates for each parameter have stabilized, efficient methods for constructing full conditional distributions and sampling from them are required. For log-concave distributions, this can be accomplished through adaptive rejection sampling (See Gilks and Wild, 1992). For applications where the full conditional distributions are not log-concave, Gilks, Best, and Tan (1995) propose appending a Hasting-Metropolis algorithm step to the adaptive rejection sampling scheme. They suggest using the resulting adaptive rejection Metropolis sampling scheme within the Gibbs sampling algorithm. We follow this approach here.

Specifically, suppose that the Gibbs sampler has been applied to the full conditional distribution of the parameter θ, $f(\theta \mid y, \hat{\psi})$, to obtain an updated estimate, say $\hat{\theta}_{CUR}$. Here, $\hat{\psi}$ contains the most recent updated estimates for all other parameters with associated full conditional distributions. For example, one possibility is that $\theta = \beta_0$, $\hat{\psi} = \{\hat{\beta}_{1\{10\}}, \ldots, \hat{\beta}_{K\{10\}}, \hat{\phi}_{\{10\}}, \hat{\sigma}^2_{\{10\}}\}$, so that $\hat{\theta}_{CUR} = \hat{\beta}_{0\{11\}}$. In what follows, the various distributions referred to are conditional upon y and $\hat{\psi}$; however we will suppress the conditioning, writing $f(\theta \mid y, \hat{\psi})$ as $f(\theta)$, for example. Let $S_M = \{\theta_i;\ i = 0, 1, \ldots, M + 1\}$ denote a set of values in ascending order for θ at which $f(\theta)$ is to be evaluated, where θ_0 and θ_{M+1} are possibly infinite lower and upper limits. Further, for $1 \le i \le j \le M$, let $L_{ij}(\theta; S_M)$ denote the straight line through the points $[\theta_i, \ln f(\theta_i)]$ and $[\theta_j, \ln f(\theta_j)]$; for other (i, j) assume that $L_{ij}(\theta; S_M)$ is undefined. Under adaptive rejection Metropolis sampling, in order to determine if $\hat{\theta}_{CUR}$ is to be kept or replaced when applying the Gibbs sampler to the full conditional of the next parameter, we proceed as follows:

(1) Sample θ from $g_M(\theta) = \frac{1}{v_M} \exp[h_M(\theta)]$ where $v_M = \int \exp[h_M(\theta)] d\theta$, and $h_M(\theta)$ is a piecewise linear function given by $h_M(\theta) = \max[L_{i,\,i+1}(\theta; S_M), \min\{L_{i-1,\,i}(\theta; S_M), L_{i+1,\,i+2}(\theta; S_M)\}]$, $\theta_i \le \theta \le \theta_{i+1}$.

(2) Sample W_1 from a uniform (0, 1) distribution.

(3) If $W_1 > f(\theta) / \exp[h_M(\theta)]$, set $S_{M+1} = S_M \cup \{\theta\}$, ensure that all values for θ in S_{M+1} are arranged in increasing order, increment M, and go back to (1). Otherwise, set $\theta_A = \theta$, and continue.

(4) Sample W_2 from a uniform (0, 1) distribution.

(5) If $W_2 > \min\left[1, \frac{f(\theta_A)\min\{f(\hat{\theta}_{CUR}), \exp[h_M(\hat{\theta}_{CUR})]\}}{f(\hat{\theta}_{CUR})\min\{f(\theta_A), \exp[h_M(\theta_A)]\}}\right]$, then use $\hat{\theta}_{CUR}$ when applying the Gibbs sampler to the next full conditional distribution. Otherwise, use θ_A instead.

When making use of adaptive rejection Metropolis sampling within the Gibbs sampler here, for each parameter $S_M = \{\theta_i;\ i = 0, 1, \ldots, M + 1\}$ initially comprised six θ_i values based on the 5th, 30th, 45th, 55th, 70th, and 95th percentiles of $h_M(\theta)$ from the previous Gibbs iteration. This adaptive rejection Metropolis sampling scheme is applied immediately following each time a full conditional distribution is sampled via the Gibbs sampler.

Following Gilks, Best, and Tan (1995) when running the above algorithm here for a particular data set, it was executed twice, using 15,000 iterations each time. A different set of starting values for the parameter estimates was used for each of the two runs. To construct the posterior distributions of the parameters, the last 3,000 iterations in each run were used in order to ensure proper convergence. The method of Gelman and Rubin (1992) was used to assess convergence. This approach yields 6,000 sets of estimates for the fixed and random effects parameters in the model, where a set is linked by the iteration number at which the estimates were produced. For each set of estimates, (8) is used to determine values for $\hat{\pi}_{ij}$ for all $j \varepsilon S'$ in the i-th local area. A value for $\hat{y}_{ij}$ where $j \varepsilon S'$ is then generated from a Bernoulli distribution with parameter $\hat{\pi}_{ij}$. The resulting values of $\hat{y}_{ij}$ are then used in (4) to determine a value for $\hat{p}_i$. There would be 6,000 such estimates for $\hat{p}_i$, one associated with each of the 6,000 sets of model estimates. These estimates for $\hat{p}_i$ are then treated as an empirical distribution. If a point estimate for the proportion of the i-th local area is desired the median of this distribution could be used. In addition, if a $100(1 - \alpha)\%$ interval estimate is required, then the $100(\alpha/2)$ and $100(1 - \alpha/2)$ percentiles of this distribution can be taken as the lower and upper limits, respectively.

3 SIMULATION STUDY

A simulation study was conducted in order to compare the performance of estimators for small area proportions based on a hierarchical Bayes estimation approach with those based on empirical Bayes techniques that use the classical EM method with a Laird approximation and a stochastic normal approximation. We imagine that two choices of surgical procedure, A and B, are available for treatment of a certain disease, and that interest is in the proportions of patients at particular hospitals that select procedure A. Suppose also that when sampling takes place in order to estimate these proportions, covariate information on gender and age is collected along with the surgical procedure selected.

Treating the hospitals as local areas, a population of 2,500 patients was created for each. To develop these populations, for each hospital a random effect value, ϕ_i, was first generated from a normal distribution with mean zero and standard deviation 0.25. Then, for the j-th patient at the i-th hospital, outcomes for gender, X_{ij1}, and age, X_{ij2}, were generated from a Bernoulli distribution with parameter 0.5, and from a continuous uniform distribution ranging from 25 to 65, respectively. The indicator variable for gender took on a value of one for females, and zero for males. For the *j*-th patient at the *i*-th hospital, the probability π_{ij} of choosing surgical procedure A was then computed according to the model

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \phi_i$$

where β_0 = -0.5, β_1 = -0.5, and β_2 = 0.02. Using the π_{ij}, the small area proportion of individuals at the *i*-th hospital opting for procedure A was determined as

$$p_i = \sum_{j=1}^{2500} \pi_{ij} \,/\, 2500$$

This yielded the small area proportions given in Table 1, which range from 0.35 to 0.59. For the *j*-th patient at the *i*-th hospital, a response variate value for Y_{ij} of 1 or 0 to indicate selection of procedure A or B respectively was generated from a Bernoulli distribution with parameter π_{ij}.

In order to study the properties of small area estimators of proportions using the three estimation methodologies, samples of 50 individuals were drawn from each of the twenty hospitals, yielding a total sample size of 1,000. For each approach, a point estimate and 95% confidence interval was determined for the proportion of patients at each hospital opting for procedure A. This process was repeated 1,000 times to allow for a comparison of the three approaches over repeated realizations of the sampling design.

The results over the entire 1,000 replications are presented in Table 1. Included for each hospital are summaries for each of the three estimation methods that reflect an average point estimate for the proportion of interest, an average confidence interval width, and a coverage rate. The average point estimates suggest that the design bias in the small area estimators associated with the three estimation approaches is quite small for most hospitals. In order to compare the three procedures, the mean absolute difference between the small area proportions and the average estimated proportions was determined for each approach. The hierarchical Bayes technique produced a mean absolute difference of 0.0028, as compared to 0.0058 and 0.0041 for the empirical Bayes methods based on classical and stochastic approximation approaches, respectively. The hierarchical Bayes approach resulted in the smallest absolute difference for half of the twenty hospitals, while the empirical Bayes method based on stochastic approximation had the smallest absolute difference in seven others.

Table 1: A comparison of the three estimation methods

Info	True p	EBayes	SS	HBayes	LengthEB	LengthSS	LengthHB	CovEB	CovSS	CovHB
Small Area Size:	0.35	0.356421	0.345471	0.347934	0.2136	0.2112	0.2130	92.4	95.2	94.8
2500	0.40	0.405710	0.411713	0.405080	0.2183	0.2206	0.2170	94.3	94.3	95.2
	0.47	0.464604	0.475181	0.470655	0.2245	0.2225	0.2230	96.8	95.5	94.7
Sample Size:	0.49	0.489121	0.492494	0.487850	0.2244	0.2226	0.2231	93.3	93.9	95.5
50	0.51	0.509368	0.504117	0.507329	0.2225	0.2233	0.2198	92.7	95.3	95.0
	0.51	0.507591	0.505628	0.502191	0.2243	0.2266	0.2240	95.4	93.6	96.0
Replications:	0.51	0.518245	0.517595	0.518074	0.2237	0.2224	0.2239	94.2	93.2	95.2
1000	0.52	0.533892	0.512017	0.518343	0.2243	0.2222	0.2237	94.5	95.5	95.6
	0.52	0.516117	0.522476	0.521422	0.2255	0.2250	0.2267	96.8	95.9	95.6
Normal Prior:	0.53	0.539127	0.526676	0.526501	0.2243	0.2267	0.2248	92.5	93.7	94.5
sd 0.25	0.54	0.547885	0.547268	0.538503	0.2223	0.2213	0.2219	93.1	96.0	94.5
	0.54	0.540483	0.544656	0.539384	0.2241	0.2262	0.2231	93.7	94.2	95.5
	0.55	0.540491	0.551902	0.550977	0.2224	0.2218	0.2227	96.9	94.1	94.6
	0.55	0.562211	0.553931	0.550421	0.2228	0.2235	0.2211	96.9	96.7	95.6
	0.56	0.549658	0.553354	0.558175	0.2230	0.2236	0.2204	95.4	95.2	95.4
	0.56	0.555629	0.555698	0.565672	0.2244	0.2237	0.2262	95.2	95.9	94.1
	0.58	0.583034	0.568609	0.576189	0.2204	0.2187	0.2224	94.4	94.9	95.6
	0.58	0.577056	0.585506	0.575505	0.2228	0.2203	0.2213	94.9	95.6	95.0
	0.59	0.594618	0.582714	0.593583	0.2196	0.2212	0.2198	94.2	94.8	95.6
	0.59	0.594970	0.588955	0.592129	0.2190	0.2171	0.2187	94.0	94.5	95.0

There is little difference in the average interval lengths obtained using the various procedures. In addition, the average coverage rates over the twenty hospitals are similar for the three approaches, and very close to the 95% nominal rate. The average coverage rates for the empirical Bayes approaches based on classical and stochastic approximation procedures are 94.58% and 94.90% respectively, while the average coverage rate for the hierarchical Bayes intervals is 95.15%. However, of note is the difference in the variability of the twenty coverage rates obtained for the three methods. The standard deviation of the coverage rates for the hierarchical Bayes intervals is only 0.497%, as compared to 1.458% and 0.937% for the empirical Bayes intervals based on classical and stochastic approximation approaches, respectively.

4 CONCLUSION AND DISCUSSION

In the context of the simulation study conducted here for estimating small area proportions, both the empirical and hierarchical Bayes procedures yielded point estimates with a small design bias, with the latter approach being slightly better than the others. The average coverage rates for the three methods are similar and very close to the 95% nominal rate. However, the variability in the individual local area coverage rates based on the hierarchical Bayes intervals is noticeably smaller than that in counterparts based on the two empirical Bayes approaches. This is clearly an advantage of the hierarchical Bayes approach. Of note also is the reduction in variability in the individual coverage rates when an empirical Bayes stochastic simulation is used instead of the classical approach. Unfortunately, this reduction in coverage rate variability as one opts for classical empirical Bayes, stochastic simulation, and hierarchical Bayes is tempered by sizeable increases in the computing time necessary to obtain the results.

A future simulation study will be done by relaxing the normality assumption on the prior distribution. The three methods suggested here will be compared to a non-normal stochastic simulation method similar to the one proposed by Zeger and Karim (1991).

5 ACKNOWLEDGEMENT

This research was supported by the Natural Sciences and Engineering Research Council of Canada.

6 REFERENCES

Breslow, N.E., and Clayton, D.G. (1993), "Approximate Inference in Generalized Linear Models", *Journal of the American Statistical Association*, 88, 9 – 25.

Carlin, B.P., and Gelfand A.E. (1990), "Approaches for Empirical Bayes Confidence Intervals", *Journal of the American Statistical Association*, 85, 105 – 114.

Datta, G.S., and Ghosh, M. (1991), "Bayesian Prediction in Linear Models: Application to Small Area Estimation", *Annals of Statistics*, 19, 1748 – 1770.

Datta, G.S., and Lahiri, P. (1995), "Robust Hierarchical Bayes Estimation of Small Area Characteristics in the Presence of Covariates and Outliers", *Journal of Multivariate Analysis*, 54, 310 – 328.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977), "Maximum Likelihood from Incomplete Data via the EM Algorithm", *Journal of the Royal Statistical Society*, *Series B*, 39, 1 – 38.

Dempster, A.P., and Tomberlin, T.J. (1980), "The Analysis of Census Undercount from a Postenumeration Study", *Proceedings of the Conference on Census Undercount*, 88 – 94.

Efron, B. (1998), "R. A. Fisher in the 21st century " (invited paper presented at the 1996 R. A. Fisher Lecture), with comments and a rejoinder by the author, *Statistical Science*, 13, 95-122.

Farrell, P.J. (2000), "Bayesian Inference for Small Area Proportions", *Sankhya*, *Series B*, 62, 402 – 416.

Farrell, P.J., MacGibbon, B., and Tomberlin, T.J. (1994), "Protection Against Outliers in Empirical Bayes Estimation", *Canadian Journal of Statistics*, 22, 365 – 376.

Farrell, P.J., MacGibbon, B., and Tomberlin, T.J. (1997), "Empirical Bayes Estimators of Small Area Proportions in Multistage Designs", *Statistica Sinica*, 7, 1065 – 1083.

Fay, R.E., and Herriot, R.A. (1979), "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data", *Journal of the American Statistical Association*, 74, 269 – 277.

Gelfand, A.E., and Smith, A.F.M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities", *Journal of the American Statistical Association*, 85, 398 – 409.

Gelman, A., and Rubin, D.B. (1992), "Inference from Iterative Simulation Using Multiple Sequences", *Statistical Science*, 7, 457 – 472.

Geman, S., and Geman, D. (1984), "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images", *IEEE Trans. Pattn Anal. Mach. Intell.*, 6, 721 – 741.

Ghosh, M., and Rao, J.N.K. (1994), "Small Area Estimation: An Appraisal", *Statistical Science*, 9, 55 – 93.

Gilks, W.R., Best, N.G., and Tan, K.K.C. (1995), "Adaptive Rejection Metropolis Sampling Within Gibbs Sampling", *Applied Statistics*, 44, 455 – 472.

Gilks, W.R., and Wild, P. (1992), "Adaptive Rejection Sampling for Gibbs Sampling", *Applied Statistics*, 41, 337 – 348.

Gu, M., and Li, S. (1998), "A Stochastic Approximation Algorithm for Maximum Likelihood Estimation with Incomplete Data", *Canadian Journal of Statistics*, 26, 567 – 582.

Lahiri, P., and Rao, J.N.K. (1995), "Robust Estimation of Mean Square Error of Small Area Estimators", *Journal of the American Statistical Association*, 90, 758 – 766.

Laird, N.M., and Louis, T.A. (1987), "Empirical Bayes Confidence Intervals Based on Bootstrap Samples", *Journal of the American Statistical Association*, 82, 739 – 750.

MacGibbon, B., and Tomberlin, T.J. (1989), "Small Area Estimates of Proportions via Empirical Bayes Techniques", *Survey Methodology*, 15, 237 – 252.

Malec, D., Sedransk, J., Moriarity, C.L., and LeClere, F.B. (1997), "Small Area Inference for Binary Variables in the National Health Interview Survey", *Journal of the American Statistical Association*, 92, 815 – 826.

McCullagh, P., and Nelder, J.A. (1989), *Generalized Linear Models*, London: Chapman and Hall.

Rao, J.N.K. (1999), "Some Recent Advances in Model-Based Small Area Estimation", *Survey Methodology*, 25, 175 – 186.

Robbins, H. and Monro S. (1951), "A Stochastic Approximation Method", *Annals of Mathematical Statistics*, 22, 400 – 407.

Royall, R.M. (1970), "On Finite Population Sampling Theory under Certain Linear Regression Models", *Biometrika*, 57, 377 – 387.

Stroud, T.W.F. (1994), "Bayesian Analysis of Binary Survey Data", *Canadian Journal of Statistics*, 22, 33 – 45.

Wong, G. Y. and Mason, W. M. (1985), "The Hierarchical Logistic Regression Model for Multilevel Analysis", *Journal of the American Statistical Association*, 80, 513 – 524.

Zeger, S. L. and Karim, M. R. (1991), "Generalized Linear Models with Random Effects: A Gibbs Sampling Approach", *Journal of the American Statistical Association*, 86, 79 – 86.