Journal of Statistical Research *ISSN 0256 - 422 X*
2010, Vol. 44, No. 1, pp. 167-186
Bangladesh

A COMPARISON OF COMPUTATIONAL APPROACHES TO BAYESIAN SMALL AREA ESTIMATION OF PROPORTIONS IN HIERARCHICAL LOGISTIC MODELS

Patrick J. Farrell

School of Mathematics and Statistics, Carleton University,
1125 Colonel By Drive, Ottawa, Ontario, Canada, K1S 5B6
Email: pfarrell@math.carleton.ca

Brenda MacGibbon

GERAD and Université du Québec à Montréal, Québec, Canada

Thomas J. Tomberlin

Carleton University, Ottawa, Ontario, Canada

SUMMARY

In this study, we are interested in comparing various computational approaches to Bayesian small area estimation of proportions in logistic regression 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. 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. However, the proven optimality properties of empirical Bayes methods and their documented successful performance have made them popular (cf. Efron 1998). The EM algorithm has proven to be an extremely useful computational tool here for empirical Bayes estimation. The approximation often used in the M step 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), Wei and Tanner (1990), Gu and Li (1998) and Nielsen (2000) we also study a stochastic simulation method to approximate the posterior distribution. Alternatively, a hierarchical Bayes approach based on Gibbs sampling 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. As the empirical Bayes estimators obtained are known to be biased, we use the bootstrap to correct for this.

Keywords and phrases: Logistic Regression, Generalized Linear Models, Overdispersion, Random Effects, Stochastic Simulation, EM algorithm, Gibbs Sampling.

© Institute of Statistical Research and Training (ISRT), University of Dhaka, Dhaka 1000, Bangladesh.

1 Introduction

The estimation of parameters for small areas has received considerable attention. This is especially true in health research as evidenced by the use of institutional profiling (cf. Normand and Shahian, 2007). 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. 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. Many other model-based estimators have also been developed, in particular, for the linear model. In a seminal theoretical paper Stein (1955) showed that the usual estimator of the mean vector of a multivariate normal distribution was inadmissible by exhibiting an estimator which dominated the usual one in mean square error. Such estimators came to be known as James-Stein estimators. Efron and Morris (1971, 1972a,b, 1973, 1975, 1977) in a series of papers showed the versatility and applicability of the Stein method for data analysis. Their 1977 paper gives insightful examples of the improvement in estimation possible with Stein's estimator. Fay and Herriot (1979) were among the first to apply James-Stein estimation techniques based on linear models to the problem of small area estimation. Such estimators could also be viewed as empirical Bayes estimators. Datta and Ghosh (1991) investigated hierarchical Bayes models for this problem and a number of studies 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 and Datta and Lahiri (1995) developed a robust hierarchical Bayes approach for handling outliers. Empirical and hierarchical Bayes techniques were soon adapted to estimation of mean parameters in hierarchical generalized linear models. Excellent summaries of these methodologies can be found in Ghosh and Rao (1994), Rao (1999), and Rao (2003).

Here we concentrate on the study of empirical and hierarchical Bayes approaches to small area estimation of proportions in hierarchical logistic regression models. The basic idea considered consists of incorporating into such a model nested random effects that reflect the complex structure of the data in a multistage sample design and to compare several computational methods to estimate the random effects and other parameters. Many authors have considered this problem from both an empirical Bayes and a hierarchical Bayes perspective. See, for example, Dempster and Tomberlin (1980), Wong and Mason (1985), Tomberlin (1988), MacGibbon and Tomberlin (1989), Albert and Chib (1993), Farrell et al. (1994, 1997a,b), Ghosh and Rao (1994), Stroud (1994), Malec et al. (1997), Farrell (2000), Holmes and Knorr-Held (2003), Holmes and Held (2006), Browne and Draper (2006), Wu and Rao (2009a,b). One method used is the classical EM algorithm of Dempster et al. (1977) combined with the approximation in the M step 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. Other researchers introduced various computational innovations on the classical EM algorithm proposed by Dempster et al. (1977). Wei and Tanner (1990), and Zeger and Karim (1991) introduced the Monte Carlo EM algorithm, which was used successfully by Chan and Ledolter (1995) in a study of a time series of counts. Rudd (1991) also gave some interesting extensions of the EM algorithm with econometric applications. Gu and Li (1998) proposed an alternative iterative approach consisting of a stochastic approximation based on the Robbins and Monro (1951) procedure. Nielsen (2000)

gave asymptotic results for the stochastic EM algorithm originally suggested by Celeux and Diebolt (1986). Holmes and Knorr-Held (2003) and Holmes and Held (2006) showed that conjugate priors were possible for Bayesian logistic regression by using a scale mixture of normals representation and additional auxiliary variables. Tanner (1996) gives a nice overview of some of these computational methods.

Our main goal here is to compare in a Monte Carlo simulation study the behaviour of three different computational approaches to small area estimation of proportions in a random-effects logistic model. The methods we have chosen to compare are: the classical empirical Bayes method using the EM algorithm proposed by Dempster et al. (1977) combined with the Laird (1978) approximation; a stochastic simulation method, that is, a hybrid stochastic EM algorithm using the Laird (1978) approximation in the M step and the averaging log-likelihood modification of the stochastic EM algorithm proposed by Nielsen (2000) for the E step; and Gibbs sampling (cf. Geman and Geman, 1984 and Gelfand and Smith, 1990) using Markov chain Monte Carlo (MCMC) techniques. This comparative study was motivated by our previous work in analyzing data using different computational techniques. For example, in Farrell et al. (1997b), we chose to use the EM algorithm in a random-effects logistic model to compare with classical unbiased and synthetic estimation approaches in order to obtain local area labour force participation rates for females in the United States using US census data. More recently, we had access to a very interesting data set originally retrieved by Simons et al. (1997) from the University of Southern California Cancer Surveillance Program, consisting of all patients who underwent surgery for rectal cancer in Los Angeles county for a specific five-year period, in order to analyze and assess the factors that affect the choice of surgery for rectal cancer. The data set consisted of over 1000 patients with several patient characteristics as well as characteristics of the hospital where the patient had the surgery. Although we considered our previous limited computer experiments with the EM algorithm, stochastic simulation and Gibbs sampling for small area estimation and applied it, with hospital as a 'small area' or random effect in a logistic model to this data set, Farrell et al. (2009) chose to use Gibbs sampling rather than either of the other two methods in order to study the problem of detecting hospitals that are outliers. This decision was based on our belief that, regardless of the many advantages of the classical EM algorithm including its ease of implementation, its computational scalability and its good convergence properties, intuitively Gibbs sampling should give the more accurate estimates. Although in our experience stochastic simulation seemed a very promising technique, we felt the need to compare the performances of these three methods in a more formal way before analyzing a real data set. Thus, our main contribution in this paper is the implementation of a relatively simple Monte Carlo simulation study in order to compare the three techniques on an application involving a random-effects logistic regression model. We also derive the bias-corrected estimates by adapting a parametric bootstrap method originally proposed by Laird and Louis (1987) and including a modification of these bootstrap techniques proposed by Carlin and Gelfand (1991).

The paper is organized as follows. Section 2 consists of the description of the model and the different estimation procedures used. In Section 3, the simulation study is described and the results presented. This is followed by a discussion and conclusions in Section 4.

2 The Model and Estimation Procedures

We follow the framework of Dempster and Tomberlin (1980) used previously for small area estimation by MacGibbon and Tomberlin (1989), and Farrell et al. (1994, 1997a,b). Although we concentrate on the hierarchical logistic model here, other generalized linear models could be studied in an analogous fashion; thus, the basic framework is re-formulated to be consistent with the notation of McCullagh and Nelder (1989) for generalized linear models. However, the estimation methods used here 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. 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

$$T_i = \sum_j y_{ij}. \tag{2.1}$$

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

$$\eta_{ij} = \beta_0 + \sum_{k=1}^{K} \beta_k X_{ijk} + \phi_i, \tag{2.2}$$

where ϕ_i represents a random effect associated with the i-th local area, β_k represents the regression coefficient associated with the k-th covariate and η_{ij} represents the link function.

Let us henceforth illustrate the techniques on a problem where the objective is the development of point and interval estimates in the hierarchical logistic model 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 based on a two stage sample design, where individuals are sampled from selected small areas. For example, imagine that there are two outcomes, A and B, and that interest is in the proportion of individuals in each of I small areas with outcome A. A sample is drawn from each selected small area, and information is recorded on the outcome for each individual in the sample. In addition, suppose that covariate information is available for all individuals in each small area 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

$$p_i = \sum_j y_{ij}/N_i, \tag{2.3}$$

where N_i is the population size of the i-th local area, 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

$$\hat{p}_i = \Big(\sum_{j\in S} y_{ij} + \sum_{j\in S'} \hat{y}_{ij}\Big)/N_i, \tag{2.4}$$

where the sum over $j \in S$ of y_{ij} is the sum of the outcome variable for sampled individuals from the i-th local area, and the sum over $j \in S'$ 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, setting $\eta_{ij} = \text{logit}(\pi_{ij})$, the model is given by

$$y_{ij} \mid \pi_{ij} \sim i.i.d.\text{Bernoulli}(\pi_{ij}), \text{logit}(\pi_{ij}) = \beta_0 + \sum_{k=1}^{K} \beta_k X_{ijk} + \phi_i = X_{ij}^T\beta + \phi_i \tag{2.5}$$

so that

$$1 - \pi_{ij} = [1 + \exp(\beta_0 + \sum_{k=1}^{K} \beta_k X_{ijk} + \phi_i)]^{-1} = [1 + \exp(X_{ij}^T\beta + \phi_i)]^{-1}. \tag{2.6}$$

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

$$\phi_i \sim i.i.d.\text{Normal}(0, \sigma^2). \tag{2.7}$$

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

$$1 - \hat{\pi}_{ij} = [1 + \exp(\hat{\beta}_0 + \sum_{k=1}^{K} \hat{\beta}_k X_{ijk} + \hat{\phi}_i)]^{-1} = [1 + \exp(X_{ij}^T\hat{\beta} + \hat{\phi}_i)]^{-1}. \tag{2.8}$$

The estimates $\hat{\pi}_{ij}$ in conjunction with (2.4) 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

$$f(y \mid \beta, \phi, \sigma^2) \propto \prod_{ij} \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{1-y_{ij}}, \tag{2.9}$$

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

$$f(\beta, \phi \mid \sigma^2) \propto \frac{1}{\sigma^I} \exp\Big(-\sum_i \frac{\phi_i^2}{2\sigma^2}\Big). \tag{2.10}$$

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

$$f(y, \beta, \phi \mid \sigma^2) \propto \prod_{ij} \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{1-y_{ij}} \frac{1}{\sigma^I} \exp\Big(- \sum_i \frac{\phi_i^2}{2\sigma^2} \Big) \tag{2.11}$$

so that the posterior distribution of the parameters is

$$f(\beta, \phi \mid y, \sigma^2) = f(y, \beta, \phi \mid \sigma^2) / f(y \mid \sigma^2). \tag{2.12}$$

It is not possible to obtain a closed form for the expression in (2.12) due to the intractable integration required to evaluate the denominator on the right hand side. Therefore, according to a proposal by Laird (1978), we can approximate (2.12) 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 et al. (1977) 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

$$\hat{\sigma}^2_{\{1\}} = \sum_{i=1}^{I} [\hat{\phi}^2_{i\{0\}} + \hat{Var}(\hat{\phi}_{i\{0\}})]/I, \tag{2.13}$$

where I is the total number of sampled local areas. 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

$$\hat{\sigma}^2_{\{n+1\}} = \sum_{i=1}^{I} [\hat{\phi}^2_{i\{n\}} + \hat{Var}(\hat{\phi}_{i\{n\}})]/I. \tag{2.14}$$

2.2 Empirical Bayes Model Estimates Based on Stochastic Simulation

As indicated in the previous section, the EM algorithm involves two steps, the E step and the M step. Various authors have proposed different methods for the calculations involved in each of these steps. The maximization in the M step is often done directly or by the usual methods such as those proposed by Wei and Tanner (1990), Chan and Ledolter (1995), Nielsen (2000), and others. These include such methods as gradient ones or scoring. For the E-step, Wei and Tanner (1990) used a Monte Carlo technique to calculate the expected likelihood. For random effects generalized linear models, Zeger and Karim (1991) proposed using either a Gibbs sampler or Monte Carlo methods to calculate this expectation. They also considered both the cases of normal and non-normal prior distributions. Chan and Ledolter (1995) successfully used Monte Carlo techniques for problems involving time series of counts. Gu and Li (1998) proposed an alternative iterative approach for

estimating the random effects variance in a generalized linear model that also makes use of the assumption that the prior distribution in equation (2.10) is multivariate normal. It consists of a stochastic approximation based on the Robbins and Monro (1951) procedure.

Both Tanner (1996) and Nielsen (2000) remark, that in order to obtain asymptotic convergence of the estimated likelihood obtained by Monte Carlo techniques to the true likelihood, the sample size used must increase towards infinity. Nielsen (2000) obtained asymptotic convergence results with the stochastic EM algorithm using for the M step, one of the usual maximization methods given above and for the E step, the method of stochastic simulation originally proposed by Celeux and Diebolt (1986), where at each iteration a sample of size one is randomly generated.

Here we use a hybrid method where, influenced by Nielsen (2000), we use the average likelihood method which takes a fixed number of iterations at each step in order to calculate the expected likelihood, while retaining the method proposed by Laird (1978) for the M step. We continue to refer to it as the stochastic simulation method here. We now describe it more precisely.

Specifically, the initialization step is as described in the previous section. We then fix a value T for the number of iterations we will use to average the likelihood. In the E step 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 the multivariate normal found by the Laird (1978) approximation in the previous M step. Then

$$\begin{aligned} Q(\sigma^2 \mid \hat{\sigma}^2_{\{n\}}) &= \iint \log[f(y \mid \beta, \phi, \sigma^2) f(\beta, \phi \mid y, \hat{\sigma}^2_{\{n\}})] d\beta d\phi \\ &\approx \frac{1}{T} \sum_{t=1}^{T} \log\{f[y \mid (\hat{\beta}_{\{n\}}, \hat{\phi}_{\{n\}})_{\{t\}}, \sigma^2] f(\beta, \phi \mid y, \hat{\sigma}^2_{\{n\}})\} \end{aligned} \tag{2.15}$$

is maximized over σ^2 using the approach of Laird (1978) to obtain $\hat{\sigma}^2_{\{n+1\}}$. If, however, it is assumed that the estimates for σ^2 have converged at iteration n, then another iteration is not needed and 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 to obtain the covariance matrix.

2.3 Empirical Bayes Local Area Estimates

Once the empirical Bayes estimates of the model parameters have been obtained according to the methods given in Section 2.1 or 2.2, equation (2.8) is used to determine a value for $\hat{\pi}_{ij}$ for all $j \epsilon S'$ in the i-th local area. Then (2.4) 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 $\hat{p}_i$. When $\sum \hat{y}_{ij}$ in equation (2.4) is replaced by $\sum \hat{\pi}_{ij}$, this mean square error can be estimated as

$$\hat{MSE}(\hat{p}_i) = \hat{Var}\left(\frac{\sum_{j \varepsilon S'} \hat{\pi}_{ij}}{N_i}\right) + \frac{\sum_{j \varepsilon S'} \hat{\pi}_{ij}(1 - \hat{\pi}_{ij})}{N_i^2}. \tag{2.16}$$

For sampled local areas, where the sample size, n_i, is greater than zero, the first term in (2.16) 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

$$Z_{ij}^T\hat{\Gamma} = X_{ij}^T\hat{\beta} + \hat{\phi}_i \tag{2.17}$$

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 (2.4) 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. Taking the variance of this expression yields

$$\hat{Var}(\hat{p}_i) = \left[\sum_{j\varepsilon S'} Z_{ij}^T\hat{\pi}_{ij}(1-\hat{\pi}_{ij})\right]\left(\frac{\hat{\Sigma}}{N_i^2}\right)\left[\sum_{j\varepsilon S'} Z_{ij}\hat{\pi}_{ij}(1-\hat{\pi}_{ij})\right], \tag{2.18}$$

where $\hat{\Sigma}$ represents the estimated covariance matrix of $\hat{\Gamma}$. A $100(1-\alpha)\%$ naive empirical Bayes confidence interval for p_i can be determined using

$$\hat{p}_i \pm z_{(1-\alpha/2)}\sqrt{\hat{Var}(\hat{p}_i)}, \tag{2.19}$$

where $z_{(1-\alpha/2)}$ is the $100(1-\alpha/2)$ percentile of a standard normal distribution. Note that provided that information is available on all Z_{ij} vectors in a local area that was not sampled, it is possible to use equations (2.4) and (2.19) to develop point and interval estimates for such a domain (See Farrell et al. 1997b, for example). However, we do not consider this situation here.

Let us also note that estimates of posterior variances given by (2.18) 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 when a naive approach is employed since the variability arising from estimating the parameters of the prior distribution is not incorporated (Ghosh and Rao, 1994; Farrell et al. 1994, 1997b;). A number of methods for addressing this shortcoming are available. One of the most compelling ways to account for such variability is by using the bootstrap originally introduced by Efron (1979). (Excellent expositions of the bootstrap can be found in Efron and Gong (1983) and Efron and Tibshirani (1993).)

For confidence intervals of the estimates of the small area proportions, Farrell et al. (1997a,b) used the parametric Type III bootstrap techniques proposed by Laird and Louis (1987) and also included a modification of these bootstrap techniques proposed by Carlin and Gelfand (1991). They demonstrated via a simulation study that the modification produced estimates for the nominal rate of coverage for individual local areas that were closer than those achieved with the simple Laird and Louis(1987) approach. It should be noted that the Type III bootstrap differs from the Rao-Wu bootstrap which is most often employed in complex survey data for constructing confidence intervals for nonlinear statistics (See Rao and Wu, 1988). The Rao-Wu bootstrap has been used effectively in many situations. Recently, Wu and Rao (2009a,b) have introduced a bootstrap calibrated pseudo empirical likelihood confidence interval, and a Bayesian pseudo empirical likelihood interval for data from complex designs.

The methodology for Laird and Louis (1987) Type III bootstrap requires the generation of a number of bootstrap samples, N_B, from each sample selected from the population under consideration. The procedure for generating a single bootstrap sample is as follows:

(i) For a given set of sample data, obtain empirical Bayes estimates of the regression coefficients and the random effects for the model in (2.5), along with an estimate of the prior distribution of the random effects using the procedure described in Section 2.1 or 2.2.

(ii) For each sampled local area, generate a random effect using the estimated prior distribution of the random effects obtained in (i).

(iii) For each sample observation, compute an estimated probability, $\hat{\pi}^*_{ij}$, using the estimated regression coefficients in (i) and the random effects generated in (ii).

(iv) For each sample observation, generate y^*_{ij} from a Bernoulli distribution with parameter $\hat{\pi}^*_{ij}$.

(v) The values obtained for y^*_{ij}, along with the vectors Z_{ij} constitute the data for a bootstrap sample.

For the b-th bootstrap sample, an empirical Bayes point estimate for the proportion of local area i, $\hat{p}^*_{bi}$, along with an associated estimate of variability $\hat{Var}(\hat{p}^*_{bi})$, are determined using the estimation procedures described in Section 2.1 or 2.2, along with equation (2.18) above. These quantities are used to calculate

$$\hat{Var}^*(\hat{p}_i) = \frac{\sum_{b=1}^{N_B} \hat{Var}(\hat{p}^*_{bi})}{N_B} + \frac{\sum_{b=1}^{N_B} (\hat{p}^*_{bi} - \hat{p}^*_i)^2}{N_B - 1} \tag{2.20}$$

where $\hat{p}^*_i = \sum_{b=1}^{N_B} \hat{p}^*_{bi}/N_B$. A bootstrap-adjusted empirical Bayes confidence interval for p_i based on the Laird and Louis (1987) approach can then be determined using

$$\hat{p}_i \pm z_{(1-\alpha/2)} \sqrt{\hat{Var}^*(\hat{p}_i)}. \tag{2.21}$$

Carlin and Gelfand (1991) proposed a modification to the Type III bootstrap, which conditions on the data from the local area in question. Their procedure, when applied to the generation of a single bootstrap sample for the i-th local area is as follows:

(i) For a given set of sample data, obtain empirical Bayes estimates of the regression coefficients and the random effects for the model in (2.5), along with an estimate of the prior distribution of the random effects using the procedure described in Section 2.1 or 2.2.

(ii) For the k-th sampled local area, where $k \neq i$, generate a random effect ϕ^*_k using the estimated prior distribution of the random effects obtained in (i).

(iii) For each observation from the k-th sampled local area, where $k \neq i$, compute an estimated probability, $\hat{\pi}^*_{kj}$, using the estimated regression coefficients in (i) and the random effects generated in (ii).

(iv) For each observation from the k-th sampled local area, where $k \neq i$, generate y^*_{kj} from a Bernoulli distribution with parameter $\hat{\pi}^*_{kj}$.

(v) The y_{ij} associated with the i-th local area, the generated y^*_{kj} values, where $k \neq i$, as well as the Z vector for each sampled local area constitute the data for a bootstrap sample.

Suppose that N_B bootstrap samples are drawn in this fashion, with the b-th sample, using the estimation procedures described in Section 2.1 or 2.2 and equation (2.18) above, producing the estimates $\hat{p}^*_{Cbi}$ and $\hat{Var}(\hat{p}^*_{Cbi})$. These estimates can then be used to determine a modified bootstrap-adjusted estimate of the variability in $\hat{p}_i$ as

$$\hat{Var}^*_C(\hat{p}_i) = \frac{\sum_{b=1}^{N_B} \hat{Var}(\hat{p}^*_{Cbi})}{N_B} + \frac{\sum_{b=1}^{N_B} (\hat{p}^*_{Cbi} - \hat{p}^*_{Ci})^2}{N_B - 1} \tag{2.22}$$

where $\hat{p}^*_{Ci} = \sum_{b=1}^{N_B} \hat{p}^*_{Cbi}/N_B$. A modified bootstrap-adjusted empirical Bayes confidence interval for p_i is given by

$$\hat{p}_i \pm z_{(1-\alpha/2)} \sqrt{\hat{Var}^*_C(\hat{p}_i)}. \tag{2.23}$$

Farrell et al. (1997a) demonstrated via a simulation study that the Carlin and Gelfand (1991) modification produced estimates for the nominal rate of coverage for individual local areas that were closer than those achieved with the simple Laird and Louis (1987) approach. We therefore used the modification in the simulation study presented in Section 3.

2.4 Hierarchical Bayes Model Parameter and Local Area Estimates

The development of hierarchical Bayes estimates for the model parameters in (2.5) requires that a distribution for the random effects variance be specified. In theory, the method is only applicable if the joint posterior of all parameters is proper. However, as a first approximation here, we have chosen a flat prior for σ^2 so that $P(y, \beta, \phi, \sigma^2)$ is identical to $P(y, \beta, \phi|\sigma^2)$ in (2.11). Using (2.11) then allows for the determination of marginal posterior distributions of each parameter up to a constant of proportionality (Gilks, Best, and Tan, 1995); the evaluation of the actual distribution is not possible due to the intractable integration required to obtain $P(\beta, \phi, \sigma^2|y)$. Specifically the posterior distribution for any given parameter is proportional to the product of all terms in (2.11) that contain it, which yields

$$\left.\begin{array}{ll} f(\beta_0 \mid y, \beta_1, \ldots, \beta_K, \phi, \sigma^2) & \propto \prod_{ij} \pi_{ij}^{y_{ij}} (1-\pi_{ij})^{1-y_{ij}}, \\ f(\beta_k \mid y, \beta_0, \ldots, \beta_{k-1}, \beta_{k+1}, \ldots, \beta_K, \phi, \sigma^2) & \propto \prod_{ij} \pi_{ij}^{y_{ij}} (1-\pi_{ij})^{1-y_{ij}}, \\ f(\phi_i \mid y, \beta, \phi_1, \ldots, \phi_{i-1}, \phi_{i+1}, \ldots, \phi_I, \sigma^2) & \propto \prod_{ij} \pi_{ij}^{y_{ij}} (1-\pi_{ij})^{1-y_{ij}} \exp(-\sum_i \frac{\phi_i^2}{2\sigma^2}), \\ f(\sigma^2 \mid y, \beta, \phi) & \propto \frac{1}{\sigma^I} \exp(-\sum_i \frac{\phi_i^2}{2\sigma^2}). \end{array}\right\} \tag{2.24}$$

Obviously, $f(\sigma^2 \mid y, \phi, \beta)$ is improper. To avoid the problems this may cause, we use a truncated version of this function for our sampling.

The computational method typically used to determine hierarchical Bayes estimates for the model parameters in (2.5) is Gibbs sampling. Although software such as WinBUGS is now readily available for such Bayesian computations and the model proposed here could be implemented in it, we include a description of the method we used for simulating the posterior distributions of the model parameter estimates.

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 $f(\beta_0 \mid y, \hat{\beta}_{1\{0\}}, \ldots, \hat{\beta}_{K\{0\}}, \hat{\phi}_{\{0\}}, \hat{\sigma}^2_{\{0\}})$. Sampling from $f(\beta_1 \mid y, \hat{\beta}_{0\{1\}}, \hat{\beta}_{2\{0\}}, \ldots, \hat{\beta}_{K\{0\}}, \hat{\phi}_{\{0\}},$

$\hat{\sigma}^2_{\{0\}}$) 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 step using the Metropolis-Hastings algorithm, an algorithm originally conceived by Hastings (1970), 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 $\theta, f(\theta \mid y, \hat{\psi})$, 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 $1 \leq i \leq j \leq M$, let $L_{ij}(\theta; S_M)$ denote the straight line through the points $[\theta_i, lnf(\theta_i)]$ and $[\theta_j, lnf(\theta_j)]$; 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 $h_M(\theta) = \max[L_{i,i+1}(\theta; S_M), \min\{L_{i-1,i}(\theta; S_M), L_{i+1,i+2}(\theta; S_M)\}]$, $\theta_i \leq \theta \leq \theta_{i+1}$.

(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 5^{th}, 30^{th}, 45^{th}, 55^{th}, 70^{th}, and 95^{th} 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 $j \epsilon S'$ in the i-th local area. A value for $\hat{y}_{ij}$ where $j \epsilon S'$ 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 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 the stochastic simulation method discussed here. For purposes of this study, we decided to create randomly generated data using a model defined by equations (2.5) through (2.7) containing two covariates; one categorical, and the other continuous. However, as noted in the introduction, we have in previous work applied the empirical Bayes techniques using the EM algorithm as well as hierarchical Bayes techniques using MCMC methodology described here to real data sets. For example, Farrell et al. (1997b) applied the classical EM algorithm using the Laird approximation to US census data and Farrell et al. (2009) applied hierarchical Bayes procedures using MCMC methodology to a cancer surveillance database. We did consider some computer experiments comparing classical EM, the Gibbs sampler and stochastic simulation but did not attempt a systematic analysis. Such an analysis is the motivation for the study here.

In order to lend a context to our randomly generated data, we imagine that two choices of surgical procedure, A and B, are available for 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 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

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \phi_i, \tag{3.1}$$

where $\beta_0 = -0.5, \beta_1 = -0.5$, and $\beta_2 = 0.02$. Using the π_{ij}, the small area proportion of individuals at the i-th hospital opting for procedure A was determined as

$$p_i = \sum_{j=1}^{2500} \pi_{ij}/2500. \tag{3.2}$$

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 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 approaches is quite small for most hospitals. To compare the three procedures, the mean absolute difference between the small area proportions and the average estimates was obtained 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 simulation 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 simulation had the smallest in seven others.

There is little difference in the average interval lengths obtained using the various procedures. The average coverage rates over the twenty hospitals are also similar for the three approaches, and very close to the 95% nominal rate. For the empirical Bayes approaches based on classical and stochastic simulation procedures these averages are 94.58% and 94.90% respectively, as compared

to 95.15% for the hierarchical Bayes intervals. 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 simulation approaches, respectively.

4 Conclusion and Discussion

In the context of the study conducted here for estimating small area proportions, both the empirical and hierarchical Bayes procedures yielded point estimates with small design bias, with the latter approach being slightly better. 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 these rates when 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 computing time. We find, however, in the context studied here, that the stochastic simulation method is a good compromise solution.

As a normal prior was used exclusively in this study, an interesting avenue of future research would be to perform a simulation study which relaxes this assumption. The study could compare the methods investigated here with analogous ones without the normality hypothesis, including a non-normal stochastic simulation method similar to the Monte Carlo one proposed by Zeger and Karim (1991). Since Holmes and Knorr-Held (2003) and Holmes and Held (2006) showed that conjugate priors are possible for Bayesian logistic regression, comparisons could also be made with their method.

An interesting extension of the computational methods we used here would be to be able to apply them to longitudinal data where generalized estimation equations (GEE) become necessary. Rao and Tausi (2004), influenced by the work of Hu and Kalbfleisch (2000) in the non-survey case, introduced an estimating function (EF) bootstrap to handle longitudinal as well as cross-sectional data for complex survey designs. Roberts et al. (2006) were able to apply the EF bootstrap to marginal logistic regression models with binary response data from a longitudinal survey with a complex design. An interesting survey on bootstrap methods for the analysis of complex sample survey data is given by Rao (2006).

Of course, in reaching any definitive conclusion we must also take into consideration the general knowledge we have about the performance of each of these methods. It is well-known that the EM algorithm is an easily implementable and computationally scalable method. Its simplicity was also noted by Rydén (2008) in his study of the EM algorithm versus MCMC methods for the estimation of hidden Markov models, although he concluded that the Bayesian methods were more useful for complex models. Browne and Draper (2006) used simulation studies with realistic designs for random-effects logistic regression multilevel models to compare (MCMC) estimation with adaptive hybrid Metropolis Gibbs sampling and quasi-likelihood methods. They concluded

that the Bayesian methods with diffuse priors clearly outperformed the likelihood methods in terms of calibration in point and interval estimation. However, it should also be noted that the Gibbs sampler is a sequential method and therefore, slower than stochastic simulation. For very large data sets, stochastic simulation would have the advantage of the possibility of having the computations done by parallel computing.

We do find that this simulation study has demonstrated the usefulness of the stochastic simulation method for the hierarchical logistic regression model and we feel that extensions of this computational method can be developed to handle more complex situations. However a good practical rule to follow in applications involving a random-effects logistic model would be to consider using all three techniques on the available data when feasible.

Acknowledgements

This research was supported by the Natural Sciences and Engineering Research Council of Canada. The authors are grateful to the referee for helpful comments which improved this manuscript.

References

[1] Albert, J. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association,* **88**, *669–679.*

[2] Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear models. *Journal of the American Statistical Association,* **88**, *9–25.*

[3] Browne, W.J. and Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis,* **1**, *473–514.*

[4] Carlin, B.P. and Gelfand A.E. (1990). Approaches for empirical Bayes confidence intervals. *Journal of the American Statistical Association,* **85**, *105–114.*

[5] Celeux, G. and Diebolt, J. (1986). The SEM algorithm: A probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational Statistics Quarterly,* **2**, *73–82.*

[6] Chan, K.S. and Ledolter, J. (1995). Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association,* **90**, *242–252.*

[7] Datta, G.S. and Ghosh, M. (1991). Bayesian prediction in linear models: Application to small area estimation. *Annals of Statistics,* **19**, *1748–1770.*

[8] 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.*

[9] 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.*

[10] 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.*

[11] Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics,* **7**, *1–26.*

[12] Efron, B. (1998). R.A. Fisher in the 21st century. *Statistical Science,* **13**, *95–122.*

[13] Efron, B. and Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross validation. *The American Statistician,* **37**, *36–48.*

[14] Efron, B. and Morris, C.N. (1971). Limiting the risk of Bayes and empirical Bayes estimators — Part I: The Bayes case. *Journal of the American Statistical Association,* **66**, *807–815.*

[15] Efron, B. and Morris, C.N. (1972a). Limiting the risk of Bayes and empirical Bayes estimators — Part II: The empirical Bayes case. *Journal of the American Statistical Association,* **67**, *130–139.*

[16] Efron, B. and Morris, C.N. (1972b). Empirical Bayes on vector observations: An extension of Stein's method. *Biometrika,* **59**, *335–347.*

[17] Efron, B. and Morris, C.N. (1973). Stein's estimation rule and its competitors — An empirical Bayes approach. *Journal of the American Statistical Association,* **68**, *117–130.*

[18] Efron, B. and Morris, C.N. (1975). Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association,* **70**, *311–319.*

[19] Efron, B. and Morris, C.N. (1977). Stein's paradox in statistics. *Scientific American,* **236**, *119–127.*

[20] Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap.* Chapman and Hall, New York.

[21] Farrell, P.J. (2000). Bayesian inference for small area proportions. *Sankhya, Series B,* **62**, *402–416.*

[22] Farrell, P.J., Groshen, S., MacGibbon, B., Tomberlin, T.J. (2009). Outlier detection for a hierarchical Bayes model in a study of hospital variation in surgical procedures. *Statistical Methods in Medical Research*, in press.

[23] Farrell, P.J., MacGibbon, B. and Tomberlin, T.J. (1994). Protection against outliers in empirical Bayes estimation. *Canadian Journal of Statistics,* **22**, *365–376.*

[24] Farrell, P.J., MacGibbon, B. and Tomberlin, T.J. (1997a). Bootstrap adjustments for empirical Bayes interval estimates of small area proportions. *Canadian Journal of Statistics,* **25**, *75–89.*

[25] Farrell, P.J., MacGibbon, B. and Tomberlin, T.J. (1997b). Empirical Bayes estimators of small area proportions in multistage designs. *Statistica Sinica,* **7**, *1065–1083.*

[26] 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.*

[27] 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.*

[28] Gelman, A. and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science,* **7**, *457–472.*

[29] Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **6**, *721–741.*

[30] Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: An appraisal. *Statistical Science,* **9**, *55–93.*

[31] Gilks, W.R., Best, N.G. and Tan, K.K.C. (1995). Adaptive rejection metropolis sampling within Gibbs sampling. *Applied Statistics,* **44**, *455–472.*

[32] Gilks, W.R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics,* **41**, *337–348.*

[33] Gonzales, M.E. (1973). Use and evaluation of synthetic estimation. *Proceedings of the American Statistical Association, Social Statistics Section, 33–36.*

[34] Gu, M. and Li, S. (1998). A stochastic approximation algorithm for maximum likelihood estimation with incomplete data. *Canadian Journal of Statistics,* **26**, *567–582.*

[35] Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika,* **57**, *97–109.*

[36] Holmes, C.C. and Held, L. (2006) Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis,* **1**, *145–168.*

[37] Holmes, C.C. and Knorr-Held, L. (2003). Efficient simulation of Bayesian logistic regression models. *LMU Publikationen.* Ludwig-Maximilians-Univeristy, Munich.

[38] Hu, F. and Kalbfleisch, J.D. (2000). The estimating function bootstrap. *Canadian Journal of Statistics,* **28**, *449–499.*

[39] 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.*

[40] Laird, N.M. (1978). Empirical Bayes methods for two-way contingency tables. *Biometrika,* **65**, *581–591.*

[41] 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.*

[42] MacGibbon, B. and Tomberlin, T.J. (1989). Small area estimates of proportions via empirical Bayes techniques. *Survey Methodology,* **15**, *237–252.*

[43] 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.*

[44] McCullagh, P. and Nelder, J.A. (1989). *Generalized linear models.* London: Chapman and Hall.

[45] Nielsen, S.F. (2000). The stochastic EM algorithm: estimation and asymptotic results. *Bernouilli,* **6**, *457–489.*

[46] Normand, S.-L.T. and Shahian, D.M. (2007). Statistical and clinical aspects of hospital outcomes profiling. *Statistical Science,* **22**, *206–226.*

[47] Rao, J.N.K. (1999). Some recent advances in model-based small area estimation. *Survey Methodology,* **25**, *175–186.*

[48] Rao, J.N.K. (2003). *Small Area Estimation: Methods and Applications.* Wiley: New York.

[49] Rao, J.N.K. (2006). Bootstrap methods for analyzing complex sample survey data. *Proceedings of Statistics Canada Symposium 2006: Methodological Issues in Measuring Population Health.*

[50] Rao, J.N.K. and Tausi, M. (2004). Estimating function jackknife variance estimators under stratified multistage sampling. *Communications in Statistics,* **33**, *2087–2095.*

[51] Rao, J.N.K. and Wu, C.F.J. (1988). Resampling inference with complex survey data. *Journal of the American Statistical Association,* **83**, *231–241.*

[52] Roberts, G., Ren, Q. and Rao, J.N.K. (2004) Using marginal mean models with data from a longitudinal survey having a complex design: Some advances in methods. In *Methodology of Longitudinal Surveys*, Chapter 20, Wiley, Chichester.

[53] Robbins, H. and Monro S. (1951). A stochastic approximation method. *Annals of Mathematical Statistics,* **22**, *400–407.*

[54] Royall, R.M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika,* **57**, *377–387.*

[55] Rudd, P.A. (1991). Extensions of estimation methods using the EM algorithm. *Journal of Econometrics,* **49**, *305–341.*

[56] Rydén, T. (2008). EM versus Markov chain Monte Carlo for estimation of hidden Markov models: A computational perspective. (with discussion). *Bayesian Analysis,* **4**, *659–715.*

[57] Simons, A.J., Ker, R., Groshen, S., Gee, C., Anthone, G.J., Ortega, A.E., Vukasin, P., Ross, R.K. and Beart, R.W. (1997). Variations in Treatment of Rectal Cancer: The Influence of Hospital Type and Caseload. *Diseases of the Colon and Rectum,* **40***, 641–646.*

[58] Stein, C. (1955). Inadmissibility of the usual estimator for the mean vector of a multivariate normal distribution. *Proceedings of the Third Berkeley Symposium in Mathematical Statistics and Probability,* **1***, 197–206.*

[59] Stroud, T.W.F. (1994). Bayesian analysis of binary survey data. *Canadian Journal of Statistics,* **22***, 33–45.*

[60] Tanner, M. A. (1996). *Tools for Statistical Inference, 3rd Edition.* Springer, New York.

[61] Wei, G.C.G. and Tanner, M.A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithm. *Journal of American Statistical Association,* **85***, 699–704.*

[62] 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.*

[63] Wu, C. and Rao, J.N.K. (2009a). Bootstrap procedures for the pseudo empirical likelihood method in sample surveys. University of Waterloo Department of Statistics and Actuarial Science Working Paper Series 2009–02.

[64] Wu, C. and Rao, J.N.K. (2009b). Bayesian pseudo empirical likelihood intervals for complex surveys. University of Waterloo Department of Statistics and Actuarial Science Working Paper Series 2009–03.

[65] 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.*

Table 1: Average point estimates, average interval widths, and coverage rates for the classical empirical Bayes (EB), stochastic simulation (SS), and hierarchical Bayes (HB) approaches over the 1000 replications conducted in the simulation study. The true proportions are also given and the nominal coverage rate is 95%.

True p	Point EB	Point SS	Point HB	Width EB	Width SS	Width HB	Cov EB	Cov SS	Cov HB
0.35	0.3564	0.3455	0.3479	0.2136	0.2112	0.2130	92.4	95.2	94.8
0.40	0.4057	0.4117	0.4051	0.2183	0.2206	0.2170	94.3	94.3	95.2
0.47	0.4646	0.4752	0.4707	0.2245	0.2225	0.2230	96.8	95.5	94.7
0.49	0.4891	0.4925	0.4879	0.2244	0.2226	0.2231	93.3	93.9	95.5
0.51	0.5094	0.5041	0.5073	0.2225	0.2233	0.2198	92.7	95.3	95.0
0.51	0.5076	0.5056	0.5022	0.2243	0.2266	0.2240	95.4	93.6	96.0
0.51	0.5182	0.5176	0.5181	0.2237	0.2224	0.2239	94.2	93.2	95.2
0.52	0.5339	0.5120	0.5183	0.2243	0.2222	0.2237	94.5	95.5	95.6
0.52	0.5161	0.5225	0.5214	0.2255	0.2250	0.2267	96.8	95.9	95.6
0.53	0.5391	0.5267	0.5265	0.2243	0.2267	0.2248	92.5	93.7	94.5
0.54	0.5479	0.5473	0.5385	0.2223	0.2213	0.2219	93.1	96.0	94.5
0.54	0.5405	0.5447	0.5394	0.2241	0.2262	0.2231	93.7	94.2	95.5
0.55	0.5405	0.5519	0.5510	0.2224	0.2218	0.2227	96.9	94.1	94.6
0.55	0.5622	0.5539	0.5504	0.2228	0.2235	0.2211	96.9	96.7	95.6
0.56	0.5496	0.5534	0.5582	0.2230	0.2236	0.2204	95.4	95.2	95.4
0.56	0.5556	0.5557	0.5657	0.2244	0.2237	0.2262	95.2	95.9	94.1
0.58	0.5830	0.5686	0.5762	0.2204	0.2187	0.2224	94.4	94.9	95.6
0.58	0.5770	0.5855	0.5755	0.2228	0.2203	0.2213	94.9	95.6	95.0
0.59	0.5946	0.5827	0.5936	0.2196	0.2212	0.2198	94.2	94.8	95.6
0.59	0.5950	0.5890	0.5921	0.2190	0.2171	0.2187	94.0	94.5	95.0