Likelihood inference for small area estimation using data cloning

Mahmoud Torabi[a], Subhash R. Lele[b], Narasimha G.N. Prasad[b]

[a]*Department of Community Health Sciences, University of Manitoba, Winnipeg, MB, R3E 0W3, Canada*
[b]*Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, T6G 2G1, Canada*

Abstract

Policy decisions regarding allocation of resources to subgroups in a population, called small areas, are based on reliable predictors of their underlying parameters. However, in sample surveys, the information to estimate reliable predictors is often insufficient at the level of the small areas. Hence, parameters of the subgroups are often predicted based on the coarser scale data. In view of this, there is a growing demand for reliable small area predictors by borrowing information from other areas. These models are commonly based on either linear mixed models (LMMs) or generalized linear mixed models (GLMMs). The frequentist analysis of LMM, a special case of GLMM, is computationally difficult. On the other hand, the advent of the Markov chain Monte Carlo algorithm has made the Bayesian analysis of LMM and GLMM computationally convenient. Recently developed data cloning method provides a frequentist approach to complex mixed models which is also computationally convenient. Data cloning which yields to max-

Email address: torabi@cc.umanitoba.ca; Fax: +1-204-789-3905 (M. Torabi)
()

imum likelihood estimation is used to conduct frequentist analysis of small area estimation for Normal and non-Normal responses. It is shown that for the Normal and non-Normal responses, data cloning leads to predictions and prediction intervals of small area parameters that have reasonably good coverage.

Keywords: Bayesian computation; Hierarchical model; Prediction interval and exponential family; Random effect

---

1. Introduction

Sample surveys are conducted with the purpose of providing reliable predictors for the finite population parameters such as totals or means. Methods used in deriving such predictors (direct survey predictors) are based on total sample size. However in the past few decades, there has been increasing demand to use sample survey data to get predictions for sub-populations, such as counties or gender-age groups. Such sub-populations for which reliable predictions are needed are called small areas. The traditional area-specific direct predictors tend to have inadequate precision because of the small samples sizes corresponding to each small area. In health services research, policy decisions about implementing programs or projects in these small areas are made using predictions of underlying parameters. Hence, survey researchers are developing methods to provide more reliable predictions for small areas. To this end, model based estimators (Rao, 2003; Jiang and Lahiri, 2006; Jiang, 2010) have been proposed to borrow strength from stated areas by introducing random effects. Depending on the nature of the data, either the linear mixed model (LMM) (Searle et al., 1992) or the generalized linear

mixed model (GLMM) (McCulloch and Searle, 2001) is most often used for small area estimation (Fay and Herriot, 1979; Battese et al., 1988; Kass and Steffey, 1989; MacGibbon and Tomberlin, 1989; Prasad and Rao, 1990; Malec et al., 1997; Ghosh et al., 1998; Singh et al., 1998; Datta et al., 2000; Ghosh et al., 2009). Among other approaches, parameters of the LMM can be estimated using either the maximum likelihood (ML) or restricted ML (REML). It is straightforward to predict the small area mean under the LMM by using the best linear unbiased predictor (BLUP), however obtaining its prediction error and associated prediction interval are difficult. Both overall parameters estimation and prediction of small area parameters under the GLMM are computationally difficult under the frequentist approach. The Bayesian approach has become quite popular because of its computational convenience and the ability to provide not just the point predictors but also the associated prediction intervals. However, the Bayesian approach to prediction of small area parameters crucially depends on the specification of the prior. Non-informative or vague priors are often used to possibly get more information from the data. However, lack of unique definition of non-informative prior leads to many different suggestions for such priors. It is well known that the choice of prior affects the predicted values. Hence, implementation of the Bayesian approach requires substantial care. For example, use of an inappropriate prior distribution can lead to improper posterior distribution making the inferential statements somewhat questionable (e.g., Natarajan and McCulloch, 1995; Hobert and Casella, 1996).

Recently, Lele et al. (2007) introduced an alternative approach, called data cloning, to compute the ML estimates and their standard errors for

general hierarchical models. Similar to the Bayesian approach, data cloning avoids high dimensional numerical integration and requires neither maximization nor differentiation of a function. Because these estimators are ML estimators, unlike the Bayesian estimators, they are independent of the choice of the priors. By applying the data cloning approach, non-estimable parameters are also flagged automatically and possibility of improper posterior distribution is completely avoided. Extending this work to the GLMM situation, Lele et al. (2010) described an approach to compute prediction and prediction intervals of the random effects. Thus, the data cloning approach is well suited to address the issues in small area estimation using the frequentist paradigm.

In this paper, we use data cloning in the context of small area estimation. In the next section, we describe the small area estimation problem in general and describe how data cloning can be used to obtain prediction and prediction intervals of small area parameters. In Section 3, we use three real datasets to evaluate the performance of data cloning under cross-sectional (normal and binomial mixed models) as well as cross-sectional and time-series (normal mixed model). The data cloning is also evaluated through simulation studies (Section 4). Concluding remarks are given in Section 5.

2. Small area estimation using data cloning

The basic model in small area estimation can be described as follows. Let y_{ij} be the variable of interest for the jth unit within the ith area ($j = 1, ..., n_i; i = 1, ..., m$). The y_{ij} are assumed to be conditionally independent

with exponential family p.d.f.

$$f(y_{ij}|\theta_{ij},\phi_{ij}) = \exp[\{y_{ij}\theta_{ij} - a(\theta_{ij})\}/\phi_{ij} + b(y_{ij},\phi_{ij})]. \quad (1)$$

The density (1) is parameterized with respect to the canonical parameters θ_{ij}, known scale parameters ϕ_{ij} and functions $a(\cdot)$ and $b(\cdot)$. The natural parameters θ_{ij} are then modeled as

$$h(\theta_{ij}) = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + u_i + v_{ij},$$

where h is a strictly increasing function, $\boldsymbol{x}_{ij}(p\times 1)$ are known design vectors, $\boldsymbol{\beta}(p\times 1)$ is a vector of unknown regression coefficients, and u_i and v_{ij} are random effects with $u_i \overset{i.i.d.}{\sim} N(0,\sigma_u^2)$ and $v_{ij} \overset{i.i.d.}{\sim} N(0,\sigma_v^2)$. We also assume that the population model holds for the sample. The objective in small area estimation is to make inferences on the small area parameters θ_{ij} or function of θ_{ij}. We now explain how data cloning can be used in the context of small area estimation.

2.1. Data cloning: a brief description

Let $\boldsymbol{y} = (\boldsymbol{y}_1,...,\boldsymbol{y}_m)'$ be the observed data vector and assume that the elements of $\boldsymbol{y}$ are conditionally independent of the random effects $\boldsymbol{u} = (u_1,...,u_m,v_{11},...,v_{mn_m})'$ and drawn from a distribution in the exponential family with parameters $\boldsymbol{\alpha}_1$. It is also assumed that distribution of $\boldsymbol{u}$ depends on parameters $\boldsymbol{\alpha}_2$:

$$\boldsymbol{y}_i|\boldsymbol{u} \sim f_{\boldsymbol{y}_i|\boldsymbol{u}}(\boldsymbol{y}_i|\boldsymbol{u},\boldsymbol{\alpha}_1),$$

$$\boldsymbol{u} \sim g_{\boldsymbol{u}}(\boldsymbol{u}|\boldsymbol{\alpha}_2). \quad (2)$$

The goal of the analysis is to estimate the model parameters $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2)'$ and predict the random effects $\boldsymbol{u}$. The likelihood for (2) is given by

$$L(\boldsymbol{\alpha}; \boldsymbol{y}) = \int \prod_{i=1}^{m} f_{\boldsymbol{y}_i|\boldsymbol{u}}(\boldsymbol{y}_i|\boldsymbol{u}, \boldsymbol{\alpha}_1) g_{\boldsymbol{u}}(\boldsymbol{u}|\boldsymbol{\alpha}_2) d\boldsymbol{u}.$$

To illustrate the data cloning approach, we start with the standard Bayesian approach to inference for hierarchical models. Denote $\pi(\boldsymbol{\alpha})$ as a prior distribution on the parameter space. The posterior distribution $\pi(\boldsymbol{\alpha}|\boldsymbol{y})$ is given by

$$\pi(\boldsymbol{\alpha}|\boldsymbol{y}) = \frac{L(\boldsymbol{\alpha}; \boldsymbol{y})\pi(\boldsymbol{\alpha})}{C(\boldsymbol{y})}, \tag{3}$$

where $C(\boldsymbol{y}) = \int L(\boldsymbol{\alpha}; \boldsymbol{y})\pi(\boldsymbol{\alpha})d\boldsymbol{\alpha}$ is the normalizing constant. There are computational tools such as Markov chain Monte Carlo (MCMC) algorithms that facilitate generation of random variates from the posterior distribution $\pi(\boldsymbol{\alpha}|\boldsymbol{y})$ without computing the integrals in the numerator or the denominator of (3)(Gilks et al., 1996; Spiegelhalter et al., 2004).

To understand the idea behind the data-cloning algorithm, imagine a hypothetical situation where the statistical experiment underlying the observations $\boldsymbol{y}$ is repeated independently by K different individuals and all these individuals happened to have the same set of observations $\boldsymbol{y}$ called $\boldsymbol{y}^{(K)} = (\boldsymbol{y}, \boldsymbol{y}, ..., \boldsymbol{y})$. The likelihood function for the combination of the data from these K independent experiments is then given by $\{L(\boldsymbol{\alpha}; \boldsymbol{y})\}^K$. Note that this likelihood function has two important features: a) the location of the maximum of this function is exactly equal to the location of the maximum of $L(\boldsymbol{\alpha}; \boldsymbol{y})$ and b) the Fisher information matrix based on this likelihood is K times the Fisher information matrix based on $L(\boldsymbol{\alpha}; \boldsymbol{y})$. Denote $\hat{\boldsymbol{\alpha}}$ as maximum likelihood estimator (MLE) and $I(\hat{\boldsymbol{\alpha}})$ as corresponding Fisher

information matrix based on $L(\boldsymbol{\alpha}; \boldsymbol{y})$. It is assumed that the parameters are identifiable and that there is a unique mode (but possibly multiple smaller peaks) to the likelihood function. The posterior distribution of $\boldsymbol{\alpha}$ conditional on the data $\boldsymbol{y}^{(K)}$ is then given by

$$\pi_K(\boldsymbol{\alpha}|\boldsymbol{y}^{(K)}) = \frac{\{L(\boldsymbol{\alpha}; \boldsymbol{y})\}^K \pi(\boldsymbol{\alpha})}{C(\boldsymbol{y}^{(K)})}, \tag{4}$$

where $C(\boldsymbol{y}^{(K)}) = \int \{L(\boldsymbol{\alpha}; \boldsymbol{y})\}^K \pi(\boldsymbol{\alpha}) d\boldsymbol{\alpha}$ is the normalizing constant. It then follows from the standard result regarding the asymptotic behaviour of the posterior distributions (e.g., Walker, 1969) that under regularity conditions, if K is large, $\pi_K(\boldsymbol{\alpha}|\boldsymbol{y}^{(K)})$ is approximately Normal with mean $\hat{\boldsymbol{\alpha}}$ and variance equal to $I^{-1}(\hat{\boldsymbol{\alpha}})/K$. Hence, this distribution is nearly degenerate at the MLE $\hat{\boldsymbol{\alpha}}$ provided K is large. Moreover, the mean of this posterior distribution is the MLE and K times the posterior variance is the corresponding asymptotic variance of the MLE $\hat{\boldsymbol{\alpha}}$.

In reality, we do not have data from K such independent experiments. However, suppose that instead of looking at the distribution in equation (4) as the posterior distribution of $\boldsymbol{\alpha}$ given the observations from K independent experiments, we look upon it as just another distribution, defined over the parameter space $\boldsymbol{\Omega}$, with probability function (4). This distribution is simply a function of the single set of observations $\boldsymbol{y}$ and the model components $f(\cdot), g(\cdot)$ and $\pi(\cdot)$. Lele et al. (2010) proved that under regularity conditions, as K becomes large, this distribution is nearly degenerate at the MLE $\hat{\boldsymbol{\alpha}}$, the mean of the probability distribution (4) converges to $\hat{\boldsymbol{\alpha}}$, and for continuous parameters, its variance is approximately $\boldsymbol{I}^{-1}(\hat{\boldsymbol{\alpha}})/K$. It then follows by generating random variates $\boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_B$ (say B is large) from (4)

and then computing its mean and variance to obtain the MLE $\hat{\boldsymbol{\alpha}}$ and its asymptotic variance, respectively. In fact, such generation of random variates from $\pi_K(\boldsymbol{\alpha}|\boldsymbol{y}^{(K)})$ is very straightforward using the MCMC technique. The experiment described above can be implemented using WinBUGS software. The $K-$cloned dataset, $\boldsymbol{y}^{(K)}$, is created by repeating the observed data vector K times. It is pretended that these data are obtained from K independent experiments. The standard MCMC method is used to generate random variates from the posterior distribution $\pi_K(\boldsymbol{\alpha}|\boldsymbol{y}^{(K)})$. The MLE of the parameter $\boldsymbol{\alpha}$ is simply the mean of these random variates provided K is large. The variance of the MLE, the inverse of the Fisher information, for the original data is K times the variance (or, variance-covariance matrix for the multiparameter case) of these random variates provided the parameter space is continuous. The advantage of this procedure is that it avoids: 1) analytical or numerical evaluation of the high dimensional integral which is a major computational hurdle for maximum likelihood estimation of GLMM, 2) numerical optimization of a function, and 3) numerical computation of the curvature of the likelihood function. In this procedure, the number of clones (K) is under the control of the analyst and can be made as large as necessary to achieve the desired accuracy of the resultant estimates. Moreover, as long as the prior distribution is not degenerate and the model satisfies some regularity conditions, the results do not depend on the choice of the prior distribution. A prior that has large probability mass near the true MLE requires fewer clones to achieve the desired accuracy. Lele *et al.* (2010) provided various checks to determine the adequate number of clones K. For instance, one may plot the largest eigenvalue of the posterior variance as a

function of the number of clones K to determine if the posterior distribution has become nearly degenerate. As another criterion, it is approximately true that as we increase the number of clones K,

$$(\boldsymbol{\alpha} - \bar{\boldsymbol{\alpha}})' \boldsymbol{V}^{-1} (\boldsymbol{\alpha} - \bar{\boldsymbol{\alpha}}) \sim \chi^2_p, \tag{5}$$

where $\bar{\boldsymbol{\alpha}}$ and $\boldsymbol{V}$ are the mean and variance of the posterior distribution, respectively, and p is the dimension of $\boldsymbol{\alpha}$. One may also compute the following two statistics: a) $\zeta = \frac{1}{B}\sum_{q=1}^{B}(O_q - Q_q)^2$, where O_q and Q_q are observed and estimated quantiles for χ^2_p random variable, and b) $\tilde{r}^2 = 1 - \rho^2$, where ρ is the correlation between (O_q, Q_q). If these statistics are close to zero, it indicates that the approximation (5) is reasonable.

2.1.1. Prediction of random effects

Prediction of random effects, particularly from the frequentist viewpoint, is problematic. If the parameters $\boldsymbol{\alpha}$ are known, then one can clearly use the conditional distribution of $\boldsymbol{u}$, the latent variables, given the observed data. That is, one can use $\pi(\boldsymbol{u}|\boldsymbol{y}, \boldsymbol{\alpha}^*)$ where $\boldsymbol{\alpha}^*$ is the true value of the parameter. A naive approach, when $\boldsymbol{\alpha}$ is estimated using the data, is to use $\pi(\boldsymbol{u}|\boldsymbol{y}, \hat{\boldsymbol{\alpha}})$. However, this approach does not take into account the variability introduced by the model parameters estimate. An approach that has been suggested in the literature (e.g., Hamilton, 1986; Lele et al., 2010) to take into account the variation of the estimator is to use the density:

$$\pi(\boldsymbol{u}|\boldsymbol{y}) = \frac{\int f(\boldsymbol{y}|\boldsymbol{u}, \boldsymbol{\alpha}_1) g(\boldsymbol{u}|\boldsymbol{\alpha}_2) \phi(\boldsymbol{\alpha}, \hat{\boldsymbol{\alpha}}, I^{-1}(\hat{\boldsymbol{\alpha}})) d\boldsymbol{\alpha}}{C(\boldsymbol{y})}, \tag{6}$$

where $\phi(., \mu, \sigma^2)$ denotes Normal density with mean μ and variance σ^2, which are equal to the MLE and the inverse of the Fisher information matrix here.

Harris (1989) argued for the use of bootstrap estimate of the sampling distribution instead of the asymptotic Normal distribution. Kass and Steffey (1989) and Booth and Hobert (1998) considered this predictive density and provided first-order and second-order approximations, respectively (based on Laplace approximations), for the mean and variance of it, but they did not study how good the predictive density or its approximation is compared to $\pi(\boldsymbol{u}|\boldsymbol{y}, \boldsymbol{\alpha}^*)$. In practice, especially if we have non-Normal responses, it is more informative to obtain prediction intervals than simple summary statistics such as mean and variance of the predictive distribution. In this paper, we obtain prediction intervals for small area parameters using the density in equation (6) along with MCMC sampling, which is a novel approach in small area estimation context. It is not known if the predictive density in (6) is optimal in any sense as an estimator of $\pi(\boldsymbol{u}|\boldsymbol{y}, \boldsymbol{\alpha}^*)$. It is, however, a consistent and sensible estimator. For a Normal linear regression model with fixed effects, this approach leads to the usual t-distribution based prediction intervals. The results presented in Section 4 show that the prediction intervals based on this predictive density have properties similar to the competitors in LMM, however, these competitors are not applicable in GLMM. This suggests that the use of predictive density in equation (6) is a reasonable approach to obtain prediction intervals for small area parameters. The main advantage of the predictive density formulation is that it can be applied to non-Normal responses in a straightforward fashion. The results presented in Section 4.2 show that the coverage obtained for the prediction of proportions is quite close to the nominal coverage. This suggests that the use of predictive density of equation (6), although not fully justified theoretically,

is sensible.

In this paper, for prior distributions, unless stated, we used the independent normal distribution for fixed effects with mean 0 and variance 10^6, and gamma distribution for inverse of variance components with shape and scale parameter 0.001.

3. Data analysis

3.1. County crop areas dataset

We study performance of the data cloning approach by applying it to a real dataset given by Battese et al. (1988). We consider the estimation of mean hectares of corn per segment for $m = 12$ counties (areas) in north-central Iowa. The total sample size, $n = \sum_i n_i$, for the 12 counties is 36 and the sample size n_i within each county ranged from 1 to 5. The population size N_i within each county ranged from 394 to 965. The nested error model used here is given by

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + u_i + e_{ij} (j = 1, ..., n_i; i = 1, ..., m), \qquad (7)$$

where y_{ij} is the number of hectares of corn in the jth segment of the ith county, x_{1ij} and x_{2ij} are the number of pixels classified as corn and soybeans in the jth segment of the ith county, $u_i \overset{i.i.d.}{\sim} N(0, \sigma_u^2)$ and $e_{ij} \overset{i.i.d.}{\sim} N(0, \sigma_e^2)$. The small area mean

$$\mu_i = \beta_0 + \beta_1 \bar{X}_{1i} + \beta_2 \bar{X}_{2i} + u_i \equiv \bar{\boldsymbol{X}}_i' \boldsymbol{\beta} + u_i, \qquad (8)$$

where $\bar{X}_{1i}$ and $\bar{X}_{2i}$ are the known population mean number of pixels classified as corn and soybeans in the ith county by using LANDSAT satellite readings and $\bar{\boldsymbol{X}}_i = (1, \bar{X}_{1i}, \bar{X}_{2i})'$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$.

The best estimator of the ith area mean μ_i in the class of unbiased predictors (Rao, 2003), sometimes called the "Bayes" estimator, is given by

$$\mu_i^B = E(\mu_i|\boldsymbol{y}_i, \boldsymbol{\beta}, \boldsymbol{\delta}) = \bar{\boldsymbol{X}}'_i\boldsymbol{\beta} + \gamma_i(\bar{y}_i - \bar{\boldsymbol{x}}'_i\boldsymbol{\beta}), \tag{9}$$

where $\gamma_i = \sigma_u^2/(\sigma_u^2 + \sigma_e^2/n_i), \boldsymbol{\delta} = (\sigma_u^2, \sigma_e^2)', \bar{y}_i$ and $\bar{\boldsymbol{x}}_i$ are the sample means. However, the Bayes estimator given by (9) depends on the knowledge of the regression coefficients $\boldsymbol{\beta}$ and variance parameters $\boldsymbol{\delta}$ which are unknown in practice. Replacing $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ by their estimated values $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\delta}}$, we obtain the empirical Bayes (EB) or empirical best linear unbiased prediction (EBLUP) of μ_i as

$$\hat{\mu}_i^{EB} = \bar{\boldsymbol{X}}'_i\hat{\boldsymbol{\beta}} + \hat{\gamma}_i(\bar{y}_i - \bar{\boldsymbol{x}}'_i\hat{\boldsymbol{\beta}}). \tag{10}$$

The mean squared prediction error, sometimes also called the mean squared error (MSE), of $\hat{\mu}_i^{EB}$ is $\text{MSE}(\hat{\mu}_i^{EB}) = E(\hat{\mu}_i^{EB} - \mu_i)^2$ where the expectation is with respect to the model. Prasad and Rao (1990) and Datta and Lahiri (2000) gave an approximate formula to compute the MSE of the EBLUP as well as a nearly unbiased estimator of MSE of the EBLUP (called $mse(\hat{\mu}_i^{EB})$) in the sense that $E[mse(\hat{\mu}_i^{EB})] = MSE(\hat{\mu}_i^{EB}) + o(m^{-1})$.

Table 1 shows the estimates of the model parameters by employing ML, REML, PR, HB, and data cloning (DC) methods, noting that PR stands for the method of Prasad and Rao (1990) and HB stands for hierarchical Bayesian method. To monitor the convergence of the model parameters, we used several diagnostic methods implemented in the Bayesian output analysis (BOA) program (Smith, 2007). We also used three diagnostic methods described in Section 2 and implemented in dclone package (Sólymos, 2010),

a freely available package created for R, to monitor the convergence of the model parameters in terms of number of clones (K). For this specific application, we used $K = 20$ to obtain MLE, and $5,000$ iterations for the convergence of model parameters. In PR method, the regression coefficients are estimated by the generalized least square method and variance components by fitting-of-constants method (Rao, 2003). As expected, the estimates of the model parameters for two methods ML and DC are very similar unlike the HB approach.

Table 1. Estimates (and standard errors) of the model parameters in the nested error linear regression model (7) for the ML, REML, PR, HB, and DC approaches.

Parameter	ML	REML	PR	HB	DC
β_0	50.97(24.52)	51.07(24.41)	51.05(24.58)	49.18(26.23)	51.00(23.64)
β_1	0.33 (0.06)	0.33(0.05)	0.33(0.05)	0.33(0.05)	0.33(0.05)
β_2	-0.13(0.06)	-0.14(0.06)	-0.13(0.06)	-0.13(0.06)	-0.13(0.05)
σ_u^2	121.10(72.70)	140.02(82.27)	139.68(59.73)	143.00(124.30)	121.20(74.04)
σ_e^2	137.30(39.29)	147.27(42.16)	149.56(45.09)	180.10(67.74)	137.40(39.44)

In Table 2, we present the EB (EBLUP) estimates of the mean hectares of corn per segment using model (8) and corresponding root mean squared errors (square root of mse($\hat{\mu}_i^{EB}$)) for methods ML, REML, PR, HB, and DC; noting that the mse($\hat{\mu}_i^{EB}$) for PR, ML and REML methods are calculated based on the approximations given in Prasad and Rao (1990) and Datta and Lahiri (2000), and posterior distributions are used to calculate the mse($\hat{\mu}_i^{EB}$) for HB and DC methods. It seems that the all approaches have comparable

results in terms of the estimates of small area means and associated root mean squared errors of the estimates of small area means.

Table 2. Average EB and associated root mean squared errors (rmse) over sample sizes within areas for the ML, REML, PR, HB, and DC methods in the nested error linear regression model (7).

	EB					rmse				
n_i	ML	REML	PR	HB	DC	ML	REML	PR	HB	DC
1	118.5	118.4	118.4	119.0	118.4	9.7	9.9	9.6	8.8	8.7
2	108.7	108.4	108.5	110.1	108.8	8.1	8.2	8.1	8.2	7.6
3	122.8	122.8	122.8	122.1	122.7	6.7	6.7	6.6	6.7	6.3
4	115.2	115.3	115.3	113.9	114.9	5.9	5.9	5.8	6.2	5.5
5	124.8	124.8	124.8	125.2	124.8	5.4	5.5	5.4	5.9	5.2

3.2. Health insurance of minority subpopulations dataset

We now evaluate the performance of data cloning using a real dataset that uses a binomial mixed model. Ghosh et al. (2009) considered small domain estimation of the proportion of persons without health insurance for different minority groups in the Asian population. The small domains were constructed on the basis of age, sex, race, and region where the person belonged. Ghosh et al. (2009) used the data provided by National Health Interview Survey (NHIS) for the year 2000 which report the individual level binary response, whether a person has health insurance or not, along with his or her individual level covariates.

The Asian group is composed of four categories: Chines, Filipino, Asian Indian, and Others such as Korean, Vietnamese, Japanese, Hawaiian, Samoan, Guamanian and etc. These individuals were assigned to specific domains depending on their age, gender, race, and the region they come from. There were three age-groups (0-17, 18-64, 65+), two groups for gender, four regions depending on the size of the Metropolitan Statistical Area ($< 499,999; 500,000-999,999; 1,000,000-2,499,999; > 2,500,000$), and four groups for race. The total number of domains is then $96(= 3 \times 2 \times 4 \times 4)$. The sample sizes for some domains for a targeted minority Asian population were not large enough to produce the reliable estimates. They then employed both HB and EB methodologies to obtain small domain estimates and also to find the associated measure of precision. In particular, they considered the following model

$$\log(\frac{p_{ij}}{1-p_{ij}}) = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \beta_3 x_{ij3} + u_i (j = 1, ..., n_i; i = 1, ..., 96),$$

where $u_i \overset{i.i.d.}{\sim} N(0, \sigma_u^2); p_{ij} = E(y_{ij})$ with $y_{ij} = 1$ or 0 if the jth individual in the ith small domain does not (does) have health insurance; $x_{ij1}, x_{ij2}, x_{ij3}$ are the family size, education level, and total family income of the jth unit in the ith small domain, respectively.

However, we can consider age, gender, and race as small area and region as strata. We then have the following two-level binomial mixed model

$$\log(\frac{p_{ij}}{1-p_{ij}}) = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \beta_3 x_{ij3} + u_i + v_{ij} (j = 1, ..., n; i = 1, ..., m), \tag{11}$$

with $y_{ij} \sim Binomial(n_{ij}, p_{ij})$ where n_{ij} is number of persons without health

insurance within region j and area i, $n = 4$ is the number of regions and $m = 24(= 3 \times 2 \times 4)$ is the number of small areas, $u_i \overset{i.i.d.}{\sim} N(0, \sigma_u^2)$ and $v_{ij} \overset{i.i.d.}{\sim} N(0, \sigma_v^2)$. We first consider the estimates of model parameters by applying DC and HB methods. The estimates of model parameters and associated standard errors are similar for both DC and HB methods with the exception of the variance component σ_u^2 (Table 3). To illustrate that the results of the HB method are not invariant to priors, we used uniform distribution $U(0, 1000)$, instead of the gamma distribution as a prior for the standard error (Gelman, 2006). We observed that the results of DC are invariant while HB method performed quite differently for the gamma and uniform prior distributions (Table 3).

Table 3. Estimates (and standard errors) of the model parameters in the binomial mixed model (11) for DC and HB approaches with different priors.

Parameter	DC	HB	HB
		Gamma distribution	Uniform distribution
β_0	-0.468(0.825)	-0.333(1.052)	0.015(1.126)
β_1	-0.200(0.137)	-0.208(0.165)	-0.224(0.166)
β_2	0.024(0.015)	0.022(0.017)	0.022(0.018)
β_3	0.300(0.093)	0.297(0.108)	0.267(0.115)
σ_u^2	0.001(0.002)	0.031(0.056)	0.053(0.084)
σ_v^2	0.074(0.044)	0.098(0.064)	0.125(0.076)

The estimates of rate of persons without health insurance for different area and region, and associated standard errors for both DC and HB methods are provided in Figure 1 which shows that they are similar. To compare the performance of HB method for different priors, we computed the ratio of

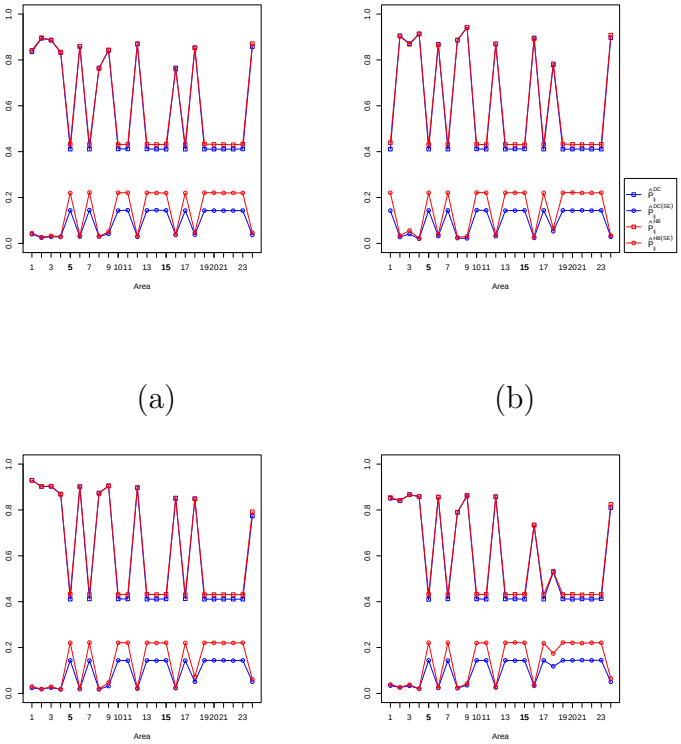
predicted probabilities of persons without health insurance for gamma distribution to uniform distribution (Figure 2). It is obvious that the performance of HB method is different for gamma and uniform distributions, illustrating the fact that the HB results are not invariant to the choice of priors.

3.3. Unemployment rate estimation

The Canadian Labour Force Survey (LFS) produces monthly estimates of the unemployment rate at national and provincial levels. The LFS also releases unemployment estimates for sub-provincial areas such as Census Metropolitan Areas (CMAs) and Census Agglomeration (CAs). However, for some sub-provincial areas, the direct estimates are not reliable since the sample size in some areas is quite small. The small area estimation in LFS concerns estimation of unemployment rates for local sub-provincial areas such as CMA/CAs using small area models. We use the 2013 January to December LFS unemployment rate estimates, y_{it}, in our data analysis. There are $m = 80$ CMA/CAs across Canada. Within each area, we consider 12 consecutive monthly estimates y_{it} from January 2013 to December 2013. For the January to December 2013 data, the overall average (over 80 CMA/CAs and 12 months) unemployment rate is 0.072. There have been some developments in the LMM to address time-series models in the context of small area estimation (Rao and Yu,1994; Datta et al., 1999; Esteban et al., 2012; Torabi, 2012). We consider the following area-level time-series model:

$$y_{it} = u_i + v_{it} + e_{it} \qquad (t = 1, ..., T = 12; i = 1, ..., m = 80), \tag{12}$$

$$v_{it} = \rho v_{i,t-1} + \epsilon_{it}, \qquad |\rho| < 1,$$



(a) (b)

(c) (d)

Figure 1: Prediction and standard error (SE) of (a) p_{i1}, (b) p_{i2}, (c) p_{i3}, and (d) p_{i4} in the binomial mixed model (11) for DC and HB approaches.

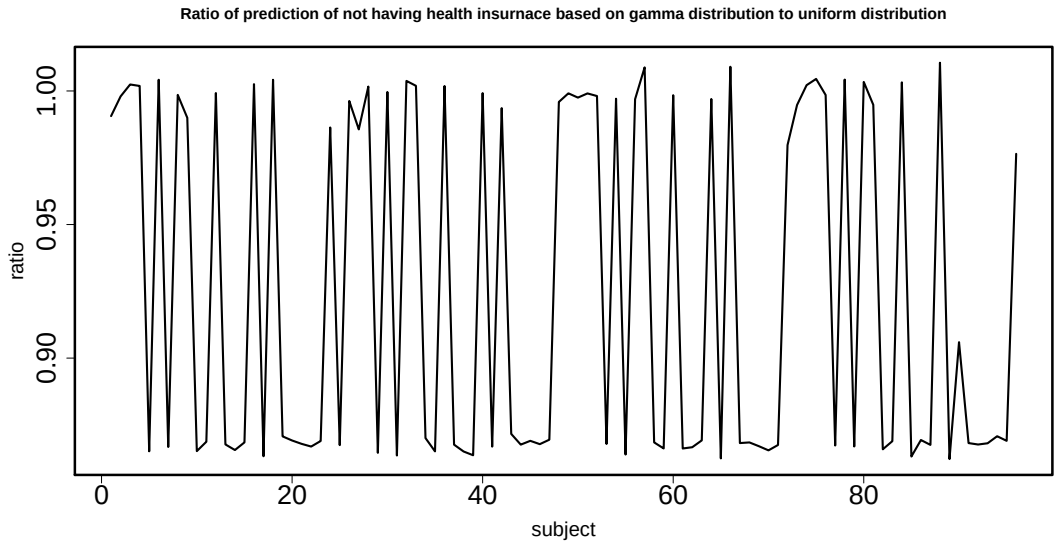


Figure 2: Ratio of prediction of probabilities of persons without health insurance for gamma distribution to uniform distribution.

where $u_i \overset{i.i.d.}{\sim} N(0, \sigma_u^2)$, v_{it} has an AR(1) model with $\epsilon_{it} \overset{i.i.d.}{\sim} N(0, \sigma_\epsilon^2)$, and the parameter of interest is $\theta_{iT} = u_i + v_{iT}$ as the true unemployment rate for area i in December 2013. Let $\boldsymbol{y}_i = (y_{i1}, ..., y_{iT})', \boldsymbol{e}_i = (e_{i1}, ..., e_{iT})', \boldsymbol{\theta}_i = (\theta_{i1}, ..., \theta_{iT})'$, where $\boldsymbol{e}_i$ follows a multivariate Gaussian with mean vector $\mathbf{0}$ and sampling covariance matrix $\boldsymbol{\Sigma}_i$. We then have $\boldsymbol{y}_i \sim N(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i), (i = 1, ..., m)$. The sampling covariance matrix $\boldsymbol{\Sigma}_i$ is unknown in the model, however, a smoothed estimator of the sampling variances is often used. There are methods to calculate the sampling covariances, for example generalized variance function (Tiller, 1989, 1992), equal coefficient of variation (CV), and equal design effects (You, 2008). In this paper, we use the equal design effects modelling approach (see You, 2008 for more details).

The model parameters estimate (and corresponding standard errors) of $\rho, \sigma_\epsilon^2, \sigma_u^2$ using the DC approach are -0.225 (0.484), 0.031(0.036) and 0.004 (0.002), respectively. Figure 3 displays the LFS direct estimates, DC and HB model-based predictions of the December 2013 unemployment rates for the 80 CMA/CAs across Canada. For the point estimates, the DC predictions lead to moderate smoothing of the direct LFS and the HB model-based estimates. For the CMAs with large population sizes and therefore large sample sizes, the direct estimates, HB, and DC predictions are very close to each other as expected.

Figure 4 displays the root mean squared error of the unemployment rate predictions in December 2013. It is clear from Figure 4 that the unemployment rates are predicted with higher precisions in the DC approach as compared to the HB approach.

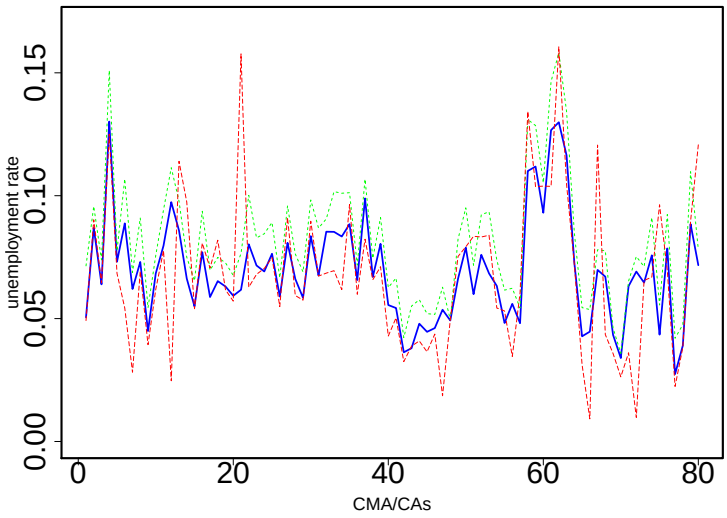


Figure 3: Unemployment rate predictions for the DC (solid blue-line), direct (dashed red-line), and HB (dashed green-line) approaches in December 2013.

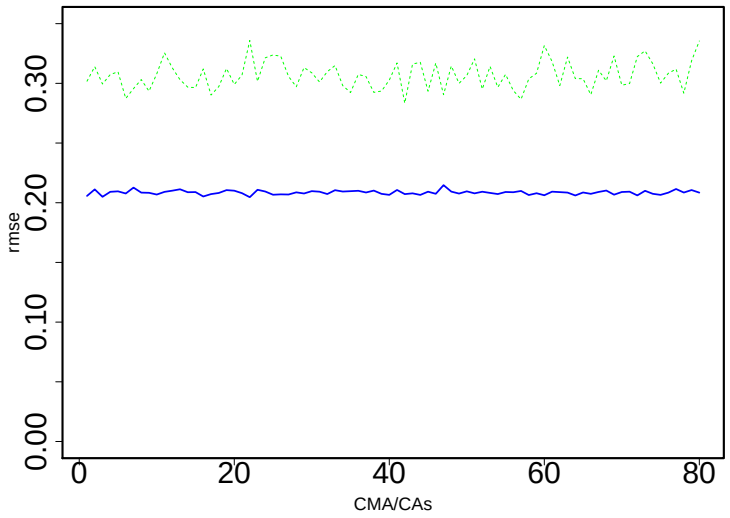


Figure 4: Root mean squared error (rmse) of the DC (solid blue-line) and HB (dashed green-line) unemployment rate predictions in December 2013.

4. Simulation study

4.1. Linear mixed model

We conduct a simulation study on the relative performance of the EB estimators in the LMM set-up. The EB estimators studied are ML, REML, PR, HB, and DC methods. We used a real dataset $\{(y_{ij}, x_{1ij}, x_{2ij}); j = 1, ..., n_i; i = 1, ..., m\}$ given by Battese et al. (1988) to simulate samples from the nested error model (7).

We first obtain the estimates of model parameters from the dataset, using for example ML method, which are $(\beta_0, \beta_1, \beta_2, \sigma_u^2, \sigma_e^2)' = (51.0, 0.329, -0.134, 121.2, 137.4)'$, and then treat them as known for the simulation study. Using those parameter values, we then approximate the true MSE of the EB estimator by drawing R=1,000 independent samples $\{y_{ij}^{(l)}; j = 1, ..., n_i; i = 1, ..., m; l = 1, ..., R\}$, as $y_{ij}^{(l)} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + u_i^{(l)} + e_{ij}^{(l)}$ where the x_{1ij} and x_{2ij} values are fixed, $u_i^{(l)}$ is generated from $N(0, \sigma_u^2)$ and $e_{ij}^{(l)}$ from $N(0, \sigma_e^2)$. Using the simulated datasets $\{(y_{ij}^{(l)}, x_{1ij}, x_{2ij}); j = 1, ..., n_i; i = 1, ..., m; l = 1, ..., R\}$, we compute the EB estimates $\hat{\mu}_i^{EB(l)}$ from (10), for each simulation run l, using different methods. In addition, $\hat{\mu}_i^{DC(l)}$ is obtained as Bayes estimator (posterior mean) of μ_i where the model parameters are obtained by the DC approach, noting that the sampling distribution of the MLE estimators are used as prior to get $\hat{\mu}_i^{DC(l)}$. The $\hat{\mu}_i^{HB(l)}$ is posterior mean of μ_i with some vague priors on the model parameters. For this simulation set-up, the average number of clones was $K = 50$ to obtain MLE, and the average number of iterations for convergence of the model parameters was about $20,000$.

The true MSE of $\hat{\mu}_i^{EB}$ is approximated as

$$\text{TMSE}_i = \frac{1}{R}\sum_{l=1}^{R}\{\hat{\mu}_i^{EB(l)} - \mu_i^{(l)}\}^2,$$

where $\mu_i^{(l)} = \beta_0 + \beta_1\bar{X}_{1i} + \beta_2\bar{X}_{2i} + u_i^{(l)}$. Table 4 shows that in terms of true MSE, $\hat{\mu}_i^{DC}$ is more efficient than other estimators particularly over HB with relative efficiency, TMSE($\hat{\mu}_i^{HB}$)/TMSE($\hat{\mu}_i^{DC}$), ranging from 105% to 120%.

Table 4. Average true MSE of $\hat{\mu}_i^{ML}, \hat{\mu}_i^{REML}, \hat{\mu}_i^{PR}, \hat{\mu}_i^{HB}$, and $\hat{\mu}_i^{DC}$ over sample sizes within areas in the nested error linear regression model (7).

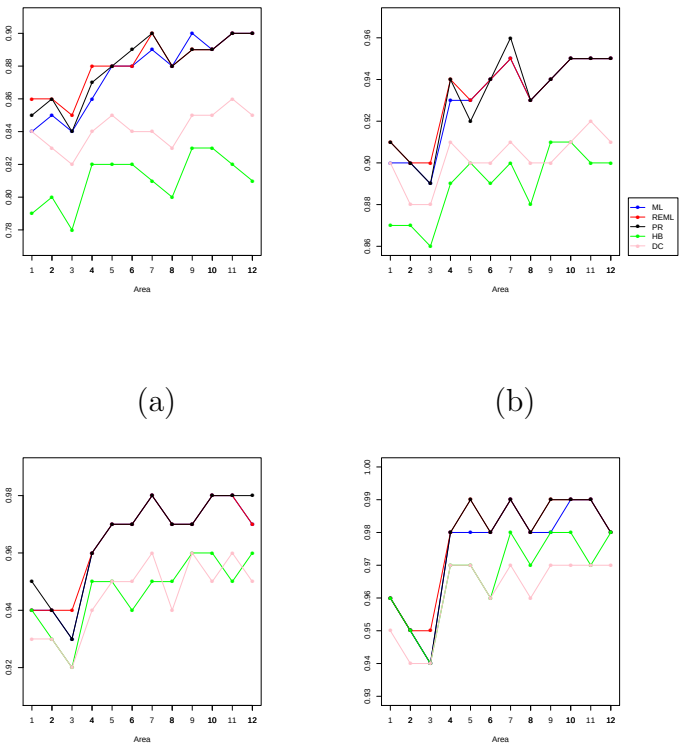
n_i	ML	REML	PR	HB	DC
1	85.3	83.8	82.5	88.7	84.5
2	59.6	58.3	57.2	65.0	59.4
3	40.9	40.2	41.6	46.0	40.9
4	31.3	30.6	32.3	36.5	31.3
5	27.6	26.9	50.8	33.2	27.7

We also study the performance of prediction intervals resulted from the DC approach. To have comparable results, we study the coverage probabilities of $\hat{\mu}_i^{ML}, \hat{\mu}_i^{REML}, \hat{\mu}_i^{PR}, \hat{\mu}_i^{DC}$ and $\hat{\mu}_i^{HB}$. To this end, for each simulation run l, we calculate $\mu_i^{(l)}$ and compute appropriate quantiles α and $(1-\alpha)$ of the posterior means $\hat{\mu}_i^{DC(l)}$ and $\hat{\mu}_i^{HB(l)}$. The coverage probabilities of $\hat{\mu}_i^{DC}$ is the proportion of the times (over $R = 1,000$) that $\mu_i^{(l)}$ falls within $(\hat{\mu}_i^{DC(l)}(\alpha), \hat{\mu}_i^{DC(l)}(1-\alpha))$. We have similar expression for the HB method. The coverage probability of $\hat{\mu}_i^{ML}, \hat{\mu}_i^{REML}$ and $\hat{\mu}_i^{PR}$ are based on the traditional closed form prediction interval

$$\text{EBLUP} \pm z_{\alpha/2}\sqrt{\text{mse}}, \tag{13}$$

where mse is a nearly unbiased estimator of MSE of the EBLUP, and $z_{\alpha/2}$ denoting the upper $100(1-\alpha/2)\%$ percentile of the standard normal distribution. The results of the coverage probabilities and average lengths of confidence intervals of $\hat{\mu}_i^{ML}, \hat{\mu}_i^{REML}, \hat{\mu}_i^{PR}, \hat{\mu}_i^{DC}$ and $\hat{\mu}_i^{HB}$ and different confidence coefficients are given in Figures 5 and 6.

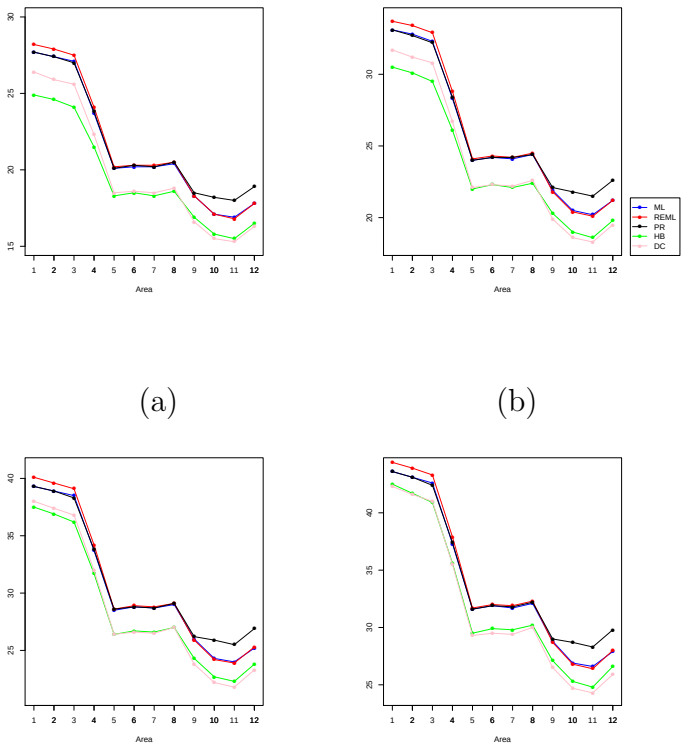
In Table 5, we compute the average coverage probabilities and average lengths of estimates of small area means over areas for the all methods. The DC method, along other methods ML, REML, and PR, performs very well in terms of coverage probabilities of the small area means and also average lengths for different confidence coefficients. The HB method also performs well in coverage probabilities of the small area means. Note that in the DC method, the inferences are based on the likelihood unlike the HB method. Most importantly, the results in the DC method are invariant to the choice of priors, while the results in HB method may depend on the priors and we may get different results for different priors. Although the methods ML, REML, and PR performed very well in terms of coverage probabilities and average lengths, these methods are not applicable in GLMM.



(a) (b)

(c) (d)

Figure 5: Coverage probabilities of confidence intervals of the $\hat{\mu}_i^{ML}, \hat{\mu}_i^{REML}, \hat{\mu}_i^{PR}, \hat{\mu}_i^{HB}$ and $\hat{\mu}_i^{DC}$ with confidence coefficients (a) 0.90, (b) 0.95, (c) 0.98, and (d) 0.99 in the nested error linear regression model (7).



(a) (b)

(c) (d)

Figure 6: Average lengths of confidence intervals of the $\hat{\mu}_i^{ML}, \hat{\mu}_i^{REML}, \hat{\mu}_i^{PR}, \hat{\mu}_i^{HB}$ and $\hat{\mu}_i^{DC}$ with confidence coefficients (a) 0.90, (b) 0.95, (c) 0.98, and (d) 0.99 in the nested error linear regression model (7).

Table 5. Average coverage probabilities (and average lengths) of $\hat{\mu}_i^{ML}, \hat{\mu}_i^{REML}$, $\hat{\mu}_i^{PR}, \hat{\mu}_i^{DC}$ and $\hat{\mu}_i^{HB}$ over areas with different confidence coefficients in the nested error linear regression model (7).

	Average coverage probabilities (and average lengths)				
Confidence coefficient	ML	REML	PR	DC	HB
0.90	0.88(21.4)	0.88(21.6)	0.88(21.7)	0.84(19.9)	0.81(19.5)
0.95	0.93(25.6)	0.93(25.8)	0.93(25.9)	0.90(23.8)	0.89(23.6)
0.98	0.96(30.4)	0.96(30.6)	0.97(30.8)	0.95(28.5)	0.95(28.5)
0.99	0.98(33.7)	0.98(33.9)	0.98(34.1)	0.96(31.7)	0.97(32.0)

4.2. Logistic mixed model

We present a simulation study to evaluate the performance of the proposed approach for non-Normal responses. To that end, we first generate $R = 3,500$ independent samples from binomial distribution:

$$y_{ij}^{(l)} \sim Binomial(n, p_{ij}^{(l)}),$$

$$\log(\frac{p_{ij}^{(l)}}{1-p_{ij}^{(l)}}) = \alpha + u_i^{(l)} + v_{ij}^{(l)} (j = 1, ..., k; i = 1, ..., m; l = 1, ..., R), \quad (14)$$

where $u_i^{(l)}$ is generated from $N(0, \sigma_u^2)$ and $v_{ij}^{(l)}$ from $N(0, \sigma_v^2)$. We set $n = 3, k = 5, m = 40, \alpha = 0$ and $\sigma_u^2 = \sigma_v^2 = 1$.

We first evaluate the performance of a method so called "lme4" which is based on Laplace approximation to get MLE in R. To our knowledge,

the lme4 package is the most popular and sophisticated package to get the MLE in the GLMM. In lme4, we can get the model parameters estimate and variance-covariance matrices for fixed effects and random effects separately. We can also get the prediction of random effects from the lme4 package. We first compare the performance of lme4 package in terms of parameters estimation and corresponding standard errors.

Using the simulated datasets $\{y_{ij}^{(l)}; j = 1, ..., k; i = 1, ..., m; l = 1, ..., R\}$, we compute the model parameters estimates from (14), for each simulation run l, for the methods DC, lme4 and HB. Our interest is to compare the performance of these three approaches and in particular two frequentist approaches DC and lme4. Table 6 presents the bias of parameters estimators. We also want to explore if the reported standard errors properly represent the true standard deviations. Towards this goal, we report the true variance of the estimated parameters and mean values of the estimated variances. It seems that the fixed parameter $\boldsymbol{\alpha}$ is well estimated by all three approaches DC, lme4 and HB. However, the variance components (σ_u^2, σ_v^2) are badly estimated by the lme4 package with relatively large biases particularly for σ_u^2. On the other hand, the DC estimators of these two variance components are reasonably unbiased and their standard errors are also estimated well. Overall, it seems that DC approach performs very well in terms of point estimates and standard errors as compared to the lme4 method. Note that in the HB method, the bias for σ_v^2 is much larger than corresponding bias value in the DC method, and also the standard errors of HB estimators and of the two variance components are relatively larger than the corresponding estimators in DC method.

Table 6. Mean values of biases and variances, and simulated variances of model parameters for the DC, lme4 and HB methods in the binomial mixed model (14).

	DC			lme4			HB		
Parameter	Bias	Simulated VAR	VAR	Bias	Simulated VAR	VAR	Bias	Simulated VAR	VAR
α	-0.003	0.041	0.040	0.007	0.036	0.034	-0.004	0.039	0.040
σ_u^2	0.038	0.179	0.193	-0.300	0.209	0.700	0.042	0.229	0.216
σ_v^2	-0.005	0.165	0.173	-0.107	0.138	0.892	0.064	0.209	0.219

We now turn to evaluate the performance of prediction of random effects for the DC approach and compare it with the lme4 and HB methods. To that end, for each simulation run l, the posterior mean $\hat{p}_{ij}^{DC(l)}$ is computed using the model parameters obtained by DC approach while $\hat{p}_{ij}^{HB(l)}$ is the posterior mean with vague priors on the model parameters; noting that in the lme4 method for each simulation run l, we can get the random effects $u_i^{(l)}$ and $v_{ij}^{(l)}$ and consequently $\hat{p}_{ij}^{lme4(l)}$. For this simulation set-up, the average number of clones was $K = 20$ to obtain MLE, and the average number of iterations for convergence of the model parameters was about $5,000$. To compute the true MSE of $\hat{p}_{ij}$, we also generate $R = 3,500$ non-sampled units:

$$y_{ijr}^{(l)} \sim Binomial\{(N-n), p_{ij}^{(l)}\},$$

$$\log(\frac{p_{ij}^{(l)}}{1-p_{ij}^{(l)}}) = \alpha + u_i^{(l)} + v_{ij}^{(l)} (j = 1, ..., k; i = 1, ..., m; l = 1, ..., R),$$

where $N = 100$ and for each simulation run l, the true small area proportions is $P_{ij}^{(l)} = N^{-1}(y_{ij}^{(l)} + y_{ijr}^{(l)})$. The true MSE of $\hat{p}_{ij}$ is then approximated as $\text{TMSE}(\hat{p}_{ij}) = R^{-1}\sum_{l=1}^{R}(\hat{p}_{ij}^{(l)} - P_{ij}^{(l)})^2, (j = 1, ..., k; i = 1, ..., m)$, for $\hat{p}_{ij}^{DC(l)}, \hat{p}_{ij}^{lme4(l)}$ and $\hat{p}_{ij}^{HB(l)}$. We also calculate the average absolute bias of $\hat{p}_{ij}$

over simulation runs. The average true MSE and average absolute bias of $\hat{p}_{ij}$ over areas and units for the DC approach are 0.028 and 0.135 respectively, while these values for the lme4 method are 0.074 and 0.221 respectively, and for the HB method are 0.028 and 0.137 respectively. Since the TMSE values are similar for each i, it is sufficient to report TMSE$(\hat{p}_{1j}), (j = 1, ..., k)$. As shown in Table 7, both $\hat{p}_{1j}^{DC}$ and $\hat{p}_{1j}^{HB}$ have similar results in terms of measure of variability unlike $\hat{p}_{1j}^{lme4}$. As a result, the performance of DC method is better than lme4 in terms of prediction of random effects based on true MSE and average absolute bias of $\hat{p}_{ij}$.

Table 7. True MSE (and average absolute bias) of $\hat{p}_{1j}(j = 1, ..., k)$ for the DC, lme4 and HB methods in the binomial mixed model (14).

Unit j	1	2	3	4	5
DC	0.027(0.133)	0.027(0.135)	0.028(0.137)	0.027(0.134)	0.028(0.136)
lme4	0.028(0.136)	0.077(0.227)	0.076(0.226)	0.076(0.224)	0.077(0.226)
HB	0.028(0.135)	0.028(0.137)	0.029(0.138)	0.028(0.136)	0.028(0.137)

We also study the relative bias (RB) of an estimator of the MSE, say mse, as

$$\text{RB}[\text{mse}(\hat{p}_{ij})] = \Big\{\frac{1}{R}\sum_{l=1}^{R}\text{mse}^{(l)}(\hat{p}_{ij}) - \text{TMSE}(\hat{p}_{ij})\Big\}/\text{TMSE}(\hat{p}_{ij}),$$

where $\text{mse}^{(l)}(\hat{p}_{ij})$ is the value of $\text{mse}(\hat{p}_{ij})$ for the lth simulation study. Since the lme4 package does not provide the variance of the $\hat{p}_{ij}$, we only compare the performance of RB of $\text{mse}(\hat{p}_{ij})$ for two methods DC and HB. The average absolute RB values over areas and units for the DC method is 5.2% while

this value for the HB approach is 9.9%. We only report the absolute RB values for DC and HB approaches across units in area 1 (Table 8).

Table 8. Percent absolute RB of mse$(\hat{p}_{1j}), (j = 1, ..., k)$, for both DC and HB methods in the binomial mixed model (14).

Unit j	1	2	3	4	5
DC	2.7	3.4	7.2	4.4	4.8
HB	7.6	8.0	11.9	9.4	9.5

We also study the coverage probabilities of $\hat{p}_{ij}^{DC}$ and $\hat{p}_{ij}^{HB}$. The results of the coverage probabilities and average lengths of confidence intervals of the $\hat{p}_{1j}^{DC}$ and $\hat{p}_{1j}^{HB}$ and different confidence coefficients are given in Table 9 for different $j(= 1, ..., k)$. The DC approach performs very well in terms of coverage probabilities of $\hat{p}_{1j}(j = 1, ..., k)$ for different confidence coefficients which are close to the nominal coverage.

Table 9. Coverage probabilities (and average lengths) of confidence intervals of the $\hat{p}_{1j}^{DC}$ and $\hat{p}_{1j}^{HB}, (j = 1, ..., k)$, with confidence coefficients 0.90, 0.95, 0.98, and 0.99 in the binomial mixed model (14).

	0.90		0.95		0.98		0.99	
Unit j	DC	HB	DC	HB	DC	HB	DC	HB
1	0.882(0.523)	0.870(0.515)	0.931(0.602)	0.927(0.595)	0.967(0.683)	0.962(0.676)	0.981(0.730)	0.978(0.725)
2	0.887(0.522)	0.872(0.514)	0.939(0.601)	0.931(0.594)	0.969(0.682)	0.966(0.675)	0.981(0.730)	0.979(0.723)
3	0.877(0.520)	0.863(0.512)	0.929(0.600)	0.921(0.592)	0.966(0.681)	0.962(0.674)	0.978(0.729)	0.976(0.722)
4	0.884(0.521)	0.869(0.512)	0.932(0.600)	0.920(0.592)	0.969(0.680)	0.963(0.674)	0.979(0.728)	0.975(0.722)
5	0.877(0.521)	0.862(0.514)	0.928(0.601)	0.918(0.594)	0.966(0.682)	0.959(0.675)	0.978(0.729)	0.975(0.724)

We compute the average coverage probabilities and average lengths of the predictions of small area proportions over areas i and units j for two methods DC and HB (Table 10). The DC method performs very well in

terms of coverage probabilities of the small area proportions for different confidence coefficients which is slightly better than the HB approach.

Table 10. Average coverage probabilities (and average lengths) of the $\hat{p}_{ij}^{DC}$ and $\hat{p}_{ij}^{HB}$ over areas and units with different confidence coefficients in the binomial mixed model (14).

	Average coverage probabilities (and average lengths)	
Confidence coefficient	DC	HB
0.90	0.880(0.522)	0.866(0.514)
0.95	0.932(0.602)	0.924(0.594)
0.98	0.967(0.682)	0.963(0.676)
0.99	0.980(0.730)	0.977(0.724)

5. Discussion

In small area estimation complex models are being used when responses are proportions or counts. In such cases, Bayesian methods are advocated because they are computationally more convenient than the maximum likelihood (ML) method. In this paper, we have shown that likelihood inference can be obtained for such models by using data cloning to overcome computational difficulties of the ML method. Under the linear mixed models, it may also lead to better inferential solutions to small area parameters. Under the generalized linear mixed model, data cloning based prediction intervals have at least as good a coverage as the Bayesian prediction intervals with appropriate priors. These appropriate priors, of course, are not known in practice.

The answers using data cloning are invariant to the choice of prior and the issue of improper posterior distribution is non-existent. Also, non-estimable parameters are flagged automatically. Note that we used the asymptotic Normal distribution to predict the random effects, however, one may use the bootstrap estimate of sampling distribution (Harris, 1989). This may lead to improvement in the prediction of random effects in terms of coverage probabilities and average lengths. To conclude, data cloning provides a better statistical inference for small area estimation, in particular for complex models on proportion or count responses.

As pointed out by a referee, probability matching priors also give approximately frequentist answers (Fraser and Reid, 2002). The problem with those priors is that they are extremely difficult to construct for any realistic model. The method of data cloning is simple and straightforward and does not need any special construction of priors. Even after using probability matching priors, we still have to do Markov chain Monte Carlo to get the posterior. We cannot use off-the-shelf programs such as WinBUGS to conduct that kind of analysis, and special programs will need to be written for each model.

Acknowledgments

We would like to thank the Editor, Associate Editor, and two referees for constructive comments and suggestions. We also thank Yong You from Statistics Canada for providing the LFS dataset. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada.

References

G.E. Battese, R.M. Harter, W.A. Fuller, 1988. An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83, 28–36.

J.G. Booth, J.P. Hobert, 1998. Standard errors of prediction in generalized linear mixed models. *Journal of the American Statistical Association*, 93, 262–272.

G.S. Datta, M. Ghosh, L.A. Waller, 2000. Hierarchical and empirical Bayes methods for environmental risk assessment, in P. K. Sen, and C. R. Rao (ed.). *Handbook of Statistics*, Volume 18, Amsterdam: Elsevier Science B.V., pp. 223–245.

G.S. Datta, P. Lahiri, 2000. A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica*, 10, 613–627.

G.S. Datta, P. Lahiri, T. Maiti, K. Lu, 1999. Hierarchical Bayes estimation of unemployment rates for the states of the US. *Journal of the American Statistical Association*, 94, 1074–1082.

M.D. Esteban, D. Morales, A. Perez, L. Sanamaria, 2012. Small area estimation of poverty proportions under area-level time series models. *Computational Statistics and Data Analysis*, 56, 2840–2855.

R.E. Fay, R.A. Herriot, 1979. Estimation of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74, 269–277.

D.A.S. Fraser, N. Reid, 2002. Strong matching of frequentist and Bayesian parametric inference. *Journal of Statistical Planning and In-*

ference, 103, 263–285.

A. Gelman, 2006. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1, 515–533.

M. Ghosh, D. Kim, K. Sinha, T. Maiti, M. Katzoff, V.L. Parsons, 2009. Hierarchical and empirical Bayes small domain estimation of the proportion of persons without health insurance for minority subpopulations. *Survey Methodology*, 35, 53–66.

M. Ghosh, K. Natarajan, T.W.F. Stroud, B.P. Carlin, 1998. Generalized linear models for small-area estimation. *Journal of the American Statistical Association*, 93, 273–282.

W.R. Gilks, S. Richardson, D.J. Spiegelhalter, (ed.), 1996. *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

J.D. Hamilton, 1986. A standard error for the estimated state vector of a state-space model. *Journal of Econometrics*, 33, 387–397.

I.R. Harris, 1989. Predictive fit for natural exponential families. *Biometrika*, 76, 675–684.

J.P. Hobert, G. Casella, 1996. The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association*, 91, 1461–1473.

J. Jiang, 2010. *Large Sample Techniques for Statistics.* New York: Springer.

J. Jiang, P. Lahiri, 2006. Mixed model prediction and small area estimation (with discussion). *TEST*, 15, 1–96.

R.E. Kass, D. Steffey, 1989. Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes

models). *Journal of the American Statistical Association*, 84, 717–726.

S.R. Lele, B. Dennis, F. Lutscher, 2007. Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. *Ecology Letters*, 10, 551–563.

S.R. Lele, K. Nadeem, B. Schmuland, 2010. Estimability and likelihood inference for generalized linear mixed models using data cloning. *Journal of the American Statistical Association*, 105, 1617–1625.

B. MacGibbon, T.J. Tomberlin, 1989. Small area estimation of proportions via empirical Bayes techniques. *Survey Methodology*, 15, 237–252.

D. Malec, J. Sedransk, C.L. Moriarity, F.B. LeClere, 1997. Small area inference for binary variables in the national health interview survey. *Journal of the American Statistical Association*, 92, 815–826.

C.E. McCulloch, S.R. Searle, 2001. *Generalized, Linear, and Mixed Models.* New York: John Wiley and Sons.

R. Natarajan, C.E. McCulloch, 1995. A note on the existence of the posterior distribution for a class of mixed models for Binomial responses. *Biometrika*, 82, 639–643.

N.G.N. Prasad, J.N.K. Rao, 1990. The estimation of mean squared error of small-area estimators. *Journal of the American Statistical Association*, 85, 163–171.

J.N.K. Rao, 2003. *Small Area Estimation.* New York: John Wiley and Sons.

J.N.K. Rao, M. Yu, 1994. Small-area estimation by combining time-series and cross-sectional data. *The Canadian Journal of Statistics*, 22, 511–528.

S.R. Searle, G. Casella, C.E. McCulloch, 1992. *Variance Components.* New York: John Wiley and Sons.

A.C. Singh, D.M. Stukel, D. Pfeffermann, 1998. Bayesian versus frequentist measures of error in small area estimation. *Journal of Royal Statistical Society Series B*, 60, 377–396.

P. Sólymos, 2010. dclone: data cloning in R. *The R Journal*, 2, 29–37.

D. Spiegelhalter, A. Thomas, N. Best, D. Lunn, 2004. *WinBUGS version 1.4 User Manual.* MRC Biostatistics unit, Institute of Public Health, London.

R. Tiller, 1989. A kalman filter approach to labour force estimation using survey data. In: *Proceedings of the Survey Research Section, American Statistical Association*, 16–25.

R. Tiller, 1992. Time series modeling of sample survey data from the U.S. current population survey. *Journal of Official Statistics*, 8, 149–166.

M. Torabi, 2012. Hierarchical Bayes estimation of spatial statistics for rates. *Journal of Statistical Planning and Inference*, 142, 358–365.

Y. You, 2008. An integrated modelling approach to unemployment rate estimation for sub-provincial areas of Canada. *Survey Methodology*, 34, 19–27.

A.M. Walker, 1969. On the asymptotic behavior of posterior distributions. *Journal of Royal Statistical Society Series B*, 31, 80–88.

00