Likelihood inference in small area estimation by combining time-series and cross-sectional data

Mahmoud Torabi[a], Farhad Shokoohi[b]

[a]*Department of Community Health Sciences, University of Manitoba, MB, R3E 0W3, Canada*
[b]*Department of Statistics, Shahid Beheshti University, Tehran, 1983963113, Iran*

Abstract

Using both time-series and cross-sectional data, a linear model incorporating autocorrelated random effects and sampling errors was previously proposed in small area estimation. However, in practice there are many situations that we have time-related counts or proportions in small area estimation; for example monthly dataset on the number of incidences in small areas. The frequentist analysis of these complex models is computationally difficult. On the other hand, the advent of the Markov chain Monte Carlo algorithm has made the Bayesian analysis of complex models computationally convenient. Recent introduction of the method of data cloning has made frequentist analysis of mixed models also equally computationally convenient. We use data cloning to conduct frequentist analysis of small area estimation for Normal and non-Normal data situations with incorporating cross-sectional and time-series data. Another important feature of the proposed approach is to predict small area parameters by providing prediction intervals. The performance of

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

the proposed approach is evaluated through several simulation studies and also by a real dataset.
Keywords: Autocorrelated errors; Bayesian computation; Hierarchical model; Prediction interval and exponential family; Random effect

---

1. Introduction

Small area estimation has received a lot of attention in recent years due to growing demand for reliable small area statistics. Rao (2003) has given a comprehensive account of model-based small area estimation. In particular, area level (Fay and Herriot, 1979) and nested error linear regression models (Battese et al., 1988; Prasad and Rao, 1990) are often used in small area estimation to obtain efficient model-based estimators of small area means.

Most of the research on small area estimation has focused on cross-sectional data at a given point in time, and the research based on time series in the context of small area estimation is limited. Scott and Smith (1974), Jones (1980) among others used time-series methods to develop efficient estimates of aggregated parameters from repeated surveys. Tiller (1992) used the idea of Kalman filter to combine a current-period state-wide estimate from the U.S. Current Population Survey with past estimate for the same state. However, non of them studied small area estimation by combining cross-sectional and time-series data.

Pfeffermann and Burck (1990) and Singh et al. (1991) among others studied cross-sectional and time-series models for small area estimation using Kalman filter by assuming specific models for the sampling errors over

time. In a pioneering paper, Rao and Yu (1994) proposed a combined cross-sectional and time-series model involving autocorrelated random effects and sampling errors with an arbitrary covariance matrix over time. Datta et al. (2002) applied same model as Rao-Yu model but replacing autoregressive (AR) random effects with random walk model. Datta et al. (1999) considered a similar model but added extra terms to reflect seasonal variation in their application. Torabi (2012) extended Datta et al. (1999) model to account for spatial variation over regions.

The main purpose of this paper is to extend the Rao-Yu model for non-Normal data using frequentist paradigm. There are many applications in small area estimation where responses are time-related counts or proportions. For example, we may be interested to analyze monthly dataset of number of incidences in small areas. Indeed, these types of models fall in the class of Generalized Additive Mixed Models (GAMMs). It is well known that the parameter estimation and prediction of small area statistics under the GAMM are extremely difficult using the frequentist approach. The Bayesian approach, especially the non-informative 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 implementation of non-informative Bayesian approach requires substantial care. The inferences may also depend on the choice of prior.

Recently, Lele et al. (2007) introduced an alternative approach, called data cloning (DC), to compute the maximum likelihood (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. Extending this work to Generalized Linear Mixed Model (GLMM) situation, Lele et al. (2010) described an approach to compute prediction and prediction intervals for the random effects. We use the idea of data cloning to extend the Rao-Yu model with incorporating cross-sectional and time-series to non-Normal data using the frequentist paradigm. Because these estimators are ML estimators, unlike the Bayesian estimators, they are independent of the choice of priors and non-estimable parameters are also flagged automatically.

In this paper, we use data cloning to propose a combined cross-sectional and time-series model with AR(1) for Normal and non-Normal data. In the next section, we describe the combined cross-sectional and time-series models. We then describe how data cloning can be used to obtain prediction and prediction intervals for small area parameters. The performance of proposed approach is reported through several simulation studies and also by an application to a real dataset. Finally, some concluding remarks are given.

2. Cross-sectional and time-series models

The basic model for the combined cross-sectional and time-series data can be described as follows. Let y_{it} be the variable of interest for the ith area in given time $t(t = 1, ..., T; i = 1, ..., m)$. The y_{it} are assumed to be conditionally independent with exponential family p.d.f.

$$f(y_{it}|\theta_{it}, \phi_{it}) = \exp[\{y_{it}\theta_{it} - a(\theta_{it})\}/\phi_{it} + b(y_{it}, \phi_{it})], \tag{2.1}$$

$(t = 1, ..., T; i = 1, ..., m)$. The density (2.1) is parameterized with respect to the canonical parameters θ_{it}, known scale parameters ϕ_{it} and functions $a(\cdot)$ and $b(\cdot)$. The exponential family (2.1) covers well-known distributions including Normal, binomial and Poisson distributions. The natural parameters θ_{it} are then modeled as

$$h(\theta_{it}) = \boldsymbol{x}'_{it}\boldsymbol{\beta} + v_i + u_{it} \ (t = 1, ..., T; i = 1, ..., m),$$

where h is a strictly increasing function, the $\boldsymbol{x}_{it}(p \times 1)$ are known design vectors, $\boldsymbol{\beta}(p \times 1)$ is a vector unknown regression coefficient, $v_i \overset{i.i.d.}{\sim} N(0, \sigma_v^2)$, and u_{it}'s are assumed to follow a common $AR(1)$ process for each i, that is,

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

with $\epsilon_{it} \overset{i.i.d.}{\sim} N(0, \sigma_\epsilon^2)$.

As special case, under Normal distribution, $h(\theta_{it}) = \theta_{it}$, the Rao-Yu model is given by

$$\hat{\theta}_{it} = \theta_{it} + e_{it} (t = 1, ..., T; i = 1, ..., m),$$

where e_{it}'s are sampling errors normally distributed, given the θ_{it}'s, with zeros means and a known block diagonal covariance matrix Ψ with blocks Ψ_i. The errors $(v_i, \epsilon_{it}, e_{it})$ are also assumed to be independent of each other. For the case of an AR(1) model with ρ known, Rao and Yu (1994) estimated σ_ϵ^2 and σ_v^2 by extending the simple transformation method of Fuller and Battese (1973). Replacing σ_ϵ^2 and σ_v^2 by their estimators $\hat{\sigma}_\epsilon^2$ and $\hat{\sigma}_v^2$, they got the empirical best linear unbiased predictor (EBLUP) of $\theta_{it}, \hat{\theta}_{it}(\rho)$, under the AR(1) model with ρ known. Rao and Yu (1994) also obtained a second-order approximation to estimator of mean squared prediction error

(MSPE) of $\hat{\theta}_{it}(\rho)$ using Taylor expansion. For the case of ρ unknown, Rao and Yu (1994) obtained a consistent estimator $\hat{\rho}$ and pointed out that this estimator often takes values outside the admissible range (-1,1), particularly for small T or small σ_ϵ^2 relative to the sampling variation. To avoid this difficulty, they proposed a naive estimator of ρ, $\hat{\rho}_N$, which is inconsistent and underestimates ρ in the presence of sampling errors. Although, the resulting EBLUP estimator $\hat{\theta}_{it}(\hat{\rho}_N)$ was unbiased, but the corresponding estimator of MSPE was not correct to terms of order $o(m^{-1})$.

3. Frequentist inference using data cloning

Let $\boldsymbol{y} = (\boldsymbol{y}_1, ..., \boldsymbol{y}_m)'$ be the observed data vector and, conditionally on the random effects, $\boldsymbol{w} = (v_1, ..., v_m, u_{11}, ..., u_{mT})'$, assume that the elements of $\boldsymbol{y}$ are independent and drawn from a distribution in the exponential family with parameters $\boldsymbol{\beta}$ where $\boldsymbol{y}_i = (y_{i1}, ..., y_{iT}), (i = 1, ..., m)$. It is also assumed that distribution for $\boldsymbol{w}$ depends on parameters $(\rho, \sigma_v^2, \sigma_\epsilon^2)$. The goal of the analysis is to estimate the model parameters $\boldsymbol{\alpha} = (\boldsymbol{\beta}, \rho, \sigma_v^2, \sigma_\epsilon^2)'$ and predict the small area parameters $\boldsymbol{\theta} = (\theta_{11}, ..., \theta_{mT})'$.

To illustrate the DC approach, we start with standard Bayesian approach to inference for our hierarchical model. Denote $L(\boldsymbol{\alpha}|\boldsymbol{y})$ as likelihood of $\boldsymbol{\alpha}$ given $\boldsymbol{y}$ and $\pi(\boldsymbol{\alpha})$ as prior distribution on the parameter space. The posterior distribution $\pi(\boldsymbol{\alpha}|\boldsymbol{y})$ is given by

$$\pi(\boldsymbol{\alpha}|\boldsymbol{y}) = \frac{L(\boldsymbol{\alpha}|\boldsymbol{y})\pi(\boldsymbol{\alpha})}{C(\boldsymbol{y})}, \tag{3.1}$$

where $C(\boldsymbol{y}) = \int L(\boldsymbol{\alpha}|\boldsymbol{y})\pi(\boldsymbol{\alpha})d\boldsymbol{\alpha}$ is the normalizing constant. There are computational tools, Markov chain Monte Carlo (MCMC) algorithms, that fa-

cilitate generation of random variates from the posterior distribution $\pi(\boldsymbol{\alpha}|\boldsymbol{y})$ without computing the integrals in the numerator or the denominator of (3.1) (Gilks et al., 1996; Spiegelhalter et al., 2004).

The DC method uses the Bayesian computational approach for frequentist purposes. In DC, the observations $\boldsymbol{y}$ is repeated independently by K different individuals and all these individuals obtain exactly the same set of observations $\boldsymbol{y}$ called $\boldsymbol{y}^{(K)} = (\boldsymbol{y}, \boldsymbol{y}, ..., \boldsymbol{y})$. The posterior distribution of $\boldsymbol{\alpha}$ conditional on the data $\boldsymbol{y}^{(K)}$ is then given by

$$\pi_K(\boldsymbol{\alpha}|\boldsymbol{y}^{(K)}) = \frac{\{L(\boldsymbol{\alpha}|\boldsymbol{y})\}^K \pi(\boldsymbol{\alpha})}{C(\boldsymbol{y}^{(K)})}, \tag{3.2}$$

where $C(\boldsymbol{y}^{(K)}) = \int \{L(\boldsymbol{\alpha}|\boldsymbol{y})\}^K \pi(\boldsymbol{\alpha}) d\boldsymbol{\alpha}$ is the normalizing constant. The expression $\{L(\boldsymbol{\alpha}|\boldsymbol{y})\}^K$ is the likelihood for K copies of the original data. Lele et al. (2007) and Lele et al. (2010) showed that, for K large enough, $\pi_K(\boldsymbol{\alpha}|\boldsymbol{y}^{(K)})$ converges to a multivariate Normal distribution with mean equal to the MLE of the model parameters and variance-covariance matrix equal to $1/K$ times the inverse of the Fisher information matrix for the MLE. This factor of $1/K$ adjusts for the fact that the cloned dataset has K times more information than the original dataset. Hence, this distribution is nearly degenerated at the MLE $\boldsymbol{\alpha}$ for large K (Walker, 1969). Moreover, the sample mean vector of the generated random numbers provides the MLE of the model parameters, and K times their sample variance-covariance matrix is an estimate of the asymptotic variance-covariance matrix for the MLE $\hat{\boldsymbol{\alpha}}$. Lele et al. (2010) also provided various checks to determine the adequate number of clones. 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,

$$(\boldsymbol{\alpha} - \bar{\boldsymbol{\alpha}})' \boldsymbol{V}^{-1} (\boldsymbol{\alpha} - \bar{\boldsymbol{\alpha}}) \sim \chi^2_p, \tag{3.3}$$

where $\boldsymbol{V}$ is the variance of the posterior distribution and p is the dimension of $\boldsymbol{\alpha}$. One may also compute the following two statistics: a) $\zeta = \frac{1}{B} \sum_{b=1}^{B} (O_b - E_b)^2$, where O_b and E_b are observed and quantiles for χ^2_p random variable, and b) $\tilde{r}^2 = 1 - \rho^2$, where ρ is the correlation between (O_b, E_b). If these statistics are close to zero, it indicates that the approximation (3.3) is reasonable.

3.1. Prediction of small area parameters

Prediction of small area parameters (random effects), particularly from the frequentist viewpoint, is somewhat of a thorny issue. If the parameters $\boldsymbol{\alpha}$ are known, then one can clearly use the conditional distribution of $\boldsymbol{\theta}$, the latent variables, given the observed data. That is, one can use $\pi(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{\alpha}^*)$ where $\boldsymbol{\alpha}^*$ is the true value of the parameter. A naive approach, when $\boldsymbol{\alpha}$ is estimated using the data, is to use $\pi(\boldsymbol{\theta}|\boldsymbol{y}, \hat{\boldsymbol{\alpha}})$. 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) to take into account the variation of the estimator is to use the density:

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{\int f(\boldsymbol{y}|\boldsymbol{\theta}, \beta) g(\boldsymbol{\theta}|\rho, \sigma_v^2, \sigma_\epsilon^2) \phi(\boldsymbol{\alpha}, \hat{\boldsymbol{\alpha}}, I^{-1}(\hat{\boldsymbol{\alpha}})) d\boldsymbol{\alpha}}{C(\boldsymbol{y})}, \tag{3.4}$$

where $\phi(., \mu, \sigma^2)$ denotes Normal density with mean μ and variance equal to the inverse of the Fisher information matrix. Harris (1989) argues for the use

of bootstrap estimate of the sampling distribution instead of the asymptotic Normal distribution. In this paper, we obtain prediction intervals for small area parameters $\boldsymbol{\theta}$ using the density in equation (3.4) along with MCMC sampling.

4. Simulation study

4.1. Linear mixed model

We conduct a simulation study to evaluate the performance of the proposed approach in the linear mixed model set up. Following simulation set up in Rao and Yu (1994), we have:

$$y_{it} = v_i + u_{it} + e_{it}(t = 1, ..., T; i = 1, ..., m),$$

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

with $\rho = 0.2$ and 0.4, $e_{it} \overset{i.i.d.}{\sim} N(0, 1)$, $v_i \overset{i.i.d.}{\sim} N(0, \sigma_v^2)$ and $\epsilon_{it} \overset{i.i.d.}{\sim} N(0, \sigma_\epsilon^2)$. We set $m = 40$ small areas and for $T = 5$, we generate $R = 5000$ independent samples $\{y_{it}^{(r)}; t = 1, ..., T; i = 1, ..., m; r = 1, ..., R\}$ for each selected pair $(\sigma_v^2, \sigma_\epsilon^2)$, and keep Ψ_i as an identity matrix. For each simulated sample, we apply the method of data cloning to get the MLE of the model parameters estimate as well as the estimator of MSPE of $\theta_{it} = v_i + u_{it}$.

In this paper, for the data cloning analysis, the proper priors are used for variance components. In particular, the proper gamma distribution was used for the inverse of variance components with shape and scale parameter 0.001. Since the data cloning is invariant to the priors, one may use different priors.

To monitor the convergence of the model parameters, we used several diagnostic methods implemented in the Bayesian output analysis (BOA) program (Smith, 2007), a freely available package created for R. We also used diagnostic methods implemented in the dclone package (Sólymos, 2010) to monitor the convergence of the model parameters in terms of number of clones K. For this simulation set up, the average number of clones was $K = 10$ to obtain MLE, and the average number of iterations for convergence of the model parameters was about 15,000.

Similar to Rao and Yu (1994), we report the estimator of MSPE for only $\hat{\theta}_{1T}$. The empirical MSPE (EMSPE) of $\hat{\theta}_{1T}$ and relative bias (RB) of an estimator of the MSPE, say mspe, are given by

$$\text{EMSPE}(\hat{\theta}_{1T}) = \frac{1}{R}\sum_{r=1}^{R}\{\hat{\theta}_{1T}^{(r)} - \theta_{1T}^{(r)}\}^2,$$

and

$$\text{RB}\{mspe(\hat{\theta}_{1T})\} = \Big\{\frac{1}{R}\sum_{r=1}^{R} mspe^{(r)}(\hat{\theta}_{1T}) - EMSPE(\hat{\theta}_{1T})\Big\}/EMSPE(\hat{\theta}_{1T}),$$

where $\hat{\theta}_{1T}^{(r)}, \theta_{1T}^{(r)}$, and $mspe^{(r)}(\hat{\theta}_{1T})$ are the values of $\hat{\theta}_{1T}, \theta_{1T}$, and $mspe(\hat{\theta}_{1T})$ for the rth simulation study, respectively. Note that $mspe(\hat{\theta}_{1T})$ is the posterior variance of $\hat{\theta}_{1T}$.

For the case of ρ known, the results of RB of $mspe(\hat{\theta}_{1T})$ are reported in Tables 1 and 2 for $\rho = 0.2$ and 0.4 respectively with different pair of $(\sigma_v^2, \sigma_\epsilon^2)$; noting that the results of Rao-Yu approach are adopted from Rao and Yu (1994). As shown in Tables 1 and 2, the estimator of MSPE performs well via data cloning approach, leading to slight underestimation for both $\rho = 0.2$

Table 1 Percent relative bias of estimators of MSPE for ρ known using data cloning (DC) and Rao-Yu (RY) approaches in the case of true value of $\rho = 0.2$, linear mixed model.

σ_v^2	$\sigma_\epsilon^2 = 0.25$		0.5		1.0		2.0	
	DC	RY	DC	RY	DC	RY	DC	RY
0.25	-6.6	2.4	-3.6	1.7	-1.9	1.8	-1.3	1.6
0.50	-6.0	2.5	-3.4	1.8	-1.7	1.7	-1.2	1.4
1.0	-5.8	2.7	-3.4	1.9	-1.7	1.7	-1.1	1.3
2.0	-6.0	2.7	-3.6	1.9	-1.9	1.7	-1.2	1.2

and 0.4. The RB is indeed decreased with increasing between-time variation. Note that Rao-Yu approach also performs well in terms of RB and even slightly better than data cloning when $\sigma_\epsilon^2 = 0.25$ for both $\rho = 0.2$ and 0.4.

We also study the performance of the prediction intervals resulted from data cloning approach. To this end, for each simulation run r, we can calculate $\theta_{it}^{(r)} = v_i^{(r)} + u_{it}^{(r)}$ and compute appropriate quantiles α and $(1-\alpha)$ of the posterior means $\hat{\theta}_{it}^{(r)}$. In particular, the coverage probabilities of the $\hat{\theta}_{1T}$ is the proportion of the times (over $R = 5000$) that $\theta_{1T}^{(r)}$ falls within $(\hat{\theta}_{1T}^{(r)(\alpha)}, \hat{\theta}_{1T}^{(r)(1-\alpha)})$. Tables 3 and 4 show the coverage probabilities of the estimates of θ_{1T} for $\rho = 0.2$ and 0.4, respectively. The data cloning method performs very well in terms of coverage probabilities of the $\hat{\theta}_{1T}$ for different confidence coefficients for both $\rho = 0.2$ and 0.4.

For the case of ρ unknown, the RB of the $mspe(\hat{\theta}_{1T})$ is also calculated for $\rho = 0.2$ and 0.4. Similar to the case of ρ known, the estimator of MSPE performs well leading to slight underestimation for both $\rho = 0.2$ and 0.4

Table 2 Percent relative bias of estimators of MSPE for ρ known using data cloning (DC) and Rao-Yu (RY) approaches in the case of true value of $\rho = 0.4$, linear mixed model.

σ_v^2	$\sigma_\epsilon^2 = 0.25$		0.5		1.0		2.0	
	DC	RY	DC	RY	DC	RY	DC	RY
0.25	-7.1	2.9	-3.5	1.3	-1.2	1.6	-0.2	1.7
0.50	-6.6	2.8	-3.7	1.2	-1.5	1.4	-0.5	1.4
1.0	-6.0	2.9	-3.6	1.3	-1.6	1.4	-0.6	1.2
2.0	-5.9	2.9	-3.6	1.3	-1.8	1.4	-0.9	1.2

(Tables 5 and 6). Note that the Rao-Yu approach also performs well in terms of RB of $mspe(\hat{\theta}_{1T})$. The data cloning method also performs very well in terms of coverage probabilities of the $\hat{\theta}_{1T}$ for different confidence coefficients for both $\rho = 0.2$ and 0.4 (Tables 7 and 8).

We should point out that in Rao-Yu approach, we need to analytically drive tedious algebra to get $mspe(\hat{\theta}_{1T})$, while in data cloning approach, not only we can easily get $mspe(\hat{\theta}_{1T})$, but also we can get the prediction interval for $\hat{\theta}_{1T}$ through MCMC. Furthermore, although the Rao-Yu method performs very well in terms of RB, this method is not applicable in GLMM.

4.2. Binomial mixed model

We also conduct a simulation study to evaluate the performance of the proposed approach in the binomial mixed model set up. To that end, we first generate $R = 5000$ independent samples:

$$y_{it,s}^{(r)} \sim Binomial(n_{it}, p_{it}^{(r)}) \tag{4.1}$$

Table 3 Coverage probabilities of the $\hat{\theta}_{1T}$ for ρ known using data cloning with different confidence coefficients in the case of true value of $\rho = 0.2$, linear mixed model.

σ_v^2	σ_ϵ^2	Confidence coefficient			
		0.90	0.95	0.98	0.99
0.25	0.25	0.868	0.929	0.965	0.981
	0.5	0.885	0.941	0.974	0.985
	1.0	0.895	0.944	0.978	0.988
	2.0	0.897	0.948	0.979	0.988
0.5	0.25	0.874	0.932	0.968	0.981
	0.50	0.889	0.940	0.975	0.987
	1.0	0.895	0.943	0.977	0.987
	2.0	0.895	0.945	0.979	0.987
1.0	0.25	0.876	0.932	0.969	0.984
	0.50	0.890	0.940	0.976	0.986
	1.0	0.894	0.943	0.976	0.987
	2.0	0.897	0.946	0.978	0.987
2.0	0.25	0.876	0.933	0.969	0.983
	0.50	0.887	0.940	0.974	0.988
	1.0	0.893	0.945	0.976	0.987
	2.0	0.896	0.947	0.978	0.988

Table 4 Coverage probabilities of the $\hat{\theta}_{1T}$ for ρ known using data cloning with different confidence coefficients in the case of true value of $\rho = 0.4$, linear mixed model.

σ_v^2	σ_ϵ^2	Confidence coefficient			
		0.90	0.95	0.98	0.99
0.25	0.25	0.868	0.928	0.966	0.978
	0.50	0.889	0.942	0.976	0.987
	1.0	0.894	0.950	0.980	0.990
	2.0	0.899	0.950	0.979	0.990
0.50	0.25	0.873	0.931	0.968	0.980
	0.50	0.886	0.940	0.976	0.987
	1.0	0.896	0.946	0.978	0.990
	2.0	0.899	0.950	0.980	0.990
1.0	0.25	0.876	0.935	0.970	0.982
	0.50	0.891	0.940	0.976	0.986
	1.0	0.897	0.945	0.980	0.990
	2.0	0.900	0.950	0.980	0.989
2.0	0.25	0.878	0.934	0.968	0.982
	0.50	0.892	0.941	0.975	0.986
	1.0	0.898	0.946	0.980	0.989
	2.0	0.901	0.950	0.979	0.989

Table 5 Percent relative bias of estimators of MSPE for ρ unknown using data cloning (DC) and Rao-Yu (RY) approaches in the case of true value of $\rho = 0.2$, linear mixed model.

σ_v^2	$\sigma_\epsilon^2 = 0.25$		0.5		1.0		2.0	
	DC	RY	DC	RY	DC	RY	DC	RY
0.25	-8.1	-1.0	-6.0	-0.7	-3.4	0.3	-2.1	0.7
0.50	-7.3	-0.8	-5.7	-0.7	-3.4	0.1	-2.1	0.5
1.0	-6.4	-0.6	-5.5	-0.6	-3.3	0.0	-2.1	0.4
2.0	-6.1	-0.4	-5.4	-0.5	-3.3	0.0	-2.1	0.3

Table 6 Percent relative bias of estimators of MSPE for ρ unknown using data cloning (DC) and Rao-Yu (RY) approaches in the case of true value of $\rho = 0.4$, linear mixed model.

σ_v^2	$\sigma_\epsilon^2 = 0.25$		0.5		1.0		2.0	
	DC	RY	DC	RY	DC	RY	DC	RY
0.25	-7.6	-4.0	-4.0	-3.8	-1.8	-1.8	-0.5	-0.3
0.50	-6.9	-4.0	-4.7	-3.9	-2.2	-2.1	-0.9	-0.5
1.0	-6.4	-3.9	-4.6	-4.0	-2.4	-2.3	-1.0	-0.7
2.0	-5.9	-3.7	-4.5	-4.0	-2.4	-2.4	-1.2	-0.9

Table 7 Coverage probabilities of the $\hat{\theta}_{1T}$ for ρ unknown using data cloning with different confidence coefficients in the case of true value of $\rho = 0.2$, linear mixed model.

σ_v^2	σ_ϵ^2	Confidence coefficient			
		0.90	0.95	0.98	0.99
0.25	0.25	0.868	0.927	0.966	0.981
	0.50	0.884	0.935	0.974	0.983
	1.0	0.893	0.942	0.977	0.987
	2.0	0.895	0.947	0.978	0.987
0.50	0.25	0.875	0.931	0.968	0.982
	0.50	0.886	0.934	0.974	0.985
	1.0	0.893	0.942	0.976	0.987
	2.0	0.897	0.947	0.978	0.987
1.0	0.25	0.880	0.932	0.969	0.984
	0.50	0.887	0.938	0.973	0.985
	1.0	0.891	0.941	0.977	0.987
	2.0	0.896	0.947	0.977	0.987
2.0	0.25	0.882	0.934	0.970	0.984
	0.50	0.886	0.937	0.973	0.985
	1.0	0.889	0.942	0.976	0.988
	2.0	0.896	0.946	0.977	0.987

Table 8 Coverage probabilities of the $\hat{\theta}_{1T}$ for ρ unknown using data cloning with different confidence coefficients in the case of true value of $\rho = 0.4$, linear mixed model.

σ_v^2	σ_ϵ^2	Confidence coefficient			
		0.90	0.95	0.98	0.99
0.25	0.25	0.875	0.928	0.967	0.978
	0.50	0.886	0.941	0.976	0.987
	1.0	0.893	0.947	0.979	0.990
	2.0	0.898	0.950	0.978	0.990
0.50	0.25	0.873	0.936	0.970	0.981
	0.50	0.885	0.938	0.973	0.985
	1.0	0.894	0.945	0.979	0.990
	2.0	0.897	0.951	0.978	0.989
1.0	0.25	0.879	0.936	0.970	0.981
	0.50	0.889	0.940	0.974	0.986
	1.0	0.896	0.943	0.979	0.990
	2.0	0.897	0.949	0.978	0.990
2.0	0.25	0.882	0.937	0.972	0.982
	0.50	0.891	0.940	0.976	0.987
	1.0	0.898	0.944	0.979	0.990
	2.0	0.896	0.949	0.979	0.990

$$\log(\frac{p_{it}^{(r)}}{1-p_{it}^{(r)}}) = v_i^{(r)} + u_{it}^{(r)} (t = 1, ..., T; i = 1, ..., m),$$

where n_{it} is sample size of ith area at time t, $v_i^{(r)} \overset{i.i.d.}{\sim} N(0, \sigma_v^2)$, and $u_{it}^{(r)}$ is generated from AR(1) with known $(\rho, \sigma_\epsilon^2)$. We also generate $R = 5000$ independent non-samples:

$$y_{it,ns}^{(r)} \sim Binomial(N_{it} - n_{it}, p_{it}^{(r)}), \tag{4.2}$$

where N_{it} is the corresponding population size for the ith area at time t; noting that the true small area proportions for each simulation run r is $P_{it}^{(r)} = N_{it}^{-1}(y_{it,s}^{(r)} + y_{it,ns}^{(r)})$. We set $N_{it} = 100, n_{it} = 5, m = 40, \rho = 0.4$, and consider $T = 5$ for each selected pair $(\sigma_v^2, \sigma_\epsilon^2)$. Using the simulated datasets $\{y_{it,s}^{(r)}; t = 1, ..., T; i = 1, ..., m; r = 1, ..., R\}$, we apply the method of data cloning to get the MLE of model parameters estimate, and also compute the small area proportions $\hat{p}_{it}$ from (4.1), for each simulation run r, called $\hat{p}_{it}^{(r)}$. For this simulation set up, the average number of clones was $K = 30$ to obtain MLE, and the average number of iterations for convergence of the model parameters was about 30,000.

The EMSPE of $\hat{p}_{it}$ and RB of $mspe(\hat{p}_{it})$ are then given by

$$EMSPE(\hat{p}_{it}) = R^{-1} \sum_{r=1}^{R} (\hat{p}_{it}^{(r)} - P_{it}^{(r)})^2 (t = 1, ..., T; i = 1, ..., m),$$

$$\text{RB}\{mspe(\hat{p}_{it})\} = \Big\{ \frac{1}{R} \sum_{r=1}^{R} mspe(\hat{p}_{it}^{(r)}) - EMSPE(\hat{p}_{it}) \Big\} / EMSPE(\hat{p}_{it}).$$

Similar to linear mixed model, we also study the coverage probabilities of $\hat{p}_{it}$. We evaluate the performance of data cloning approach for both ρ known and unknown.

Table 9 Empirical MSPE of $\hat{p}_{1T}$ for ρ known using data cloning approach, binomial mixed model.

σ_v^2	$\sigma_\epsilon^2 = 1.0$	2.0
1.0	0.020	0.021
2.0	0.018	0.020

Table 10 Percent relative bias of estimators of MSPE for ρ known using data cloning approach, binomial mixed model.

σ_v^2	$\sigma_\epsilon^2 = 1.0$	2.0
1.0	-2.6	1.3
2.0	-4.5	-0.6

For the case of ρ known, we report the EMSPE for only $\hat{p}_{1T}$ (similar to linear mixed model set up). As shown in Table 9, the EMSPE values are slightly decreased with increasing the variance of area random effects. The RB of $mspe(\hat{p}_{1T})$ is reported in Table 10. Similar to linear mixed model, the data cloning performs very well in terms of RB ($|RB|(\%) \leq 4.5$). The results of the coverage probabilities and average lengths of confidence intervals of the $\hat{p}_{1T}$ and different coefficients are given in Table 11. The data cloning also performs very well in terms of coverage probabilities and average lengths of the $\hat{p}_{1T}$ for different confidence coefficients.

For the case of ρ unknown, we also report the EMSPE of $\hat{p}_{1T}$ in Table 12 and the RB of $mspe(\hat{p}_{1T})$ in Table 13. The data cloning approach performs very well in terms of RB, leading to slight underestimation ($|RB|(\%) \leq 6.1$). The data cloning also performs very well in terms of coverage probabilities

Table 11 Coverage probabilities (and average lengths) of the $\hat{p}_{1T}$ for ρ known with different confidence coefficients using data cloning approach, binomial mixed model.

σ_v^2	σ_ϵ^2	Confidence coefficient (average lengths)			
		0.90	0.95	0.98	0.99
1.0	1.0	0.883(0.445)	0.933(0.518)	0.965(0.595)	0.977(0.643)
	2.0	0.871(0.434)	0.921(0.508)	0.954(0.589)	0.964(0.639)
2.0	1.0	0.887(0.458)	0.935(0.533)	0.966(0.614)	0.975(0.663)
	2.0	0.872(0.415)	0.925(0.485)	0.957(0.561)	0.965(0.609)

Table 12 Empirical MSPE of $\hat{p}_{1T}$ for ρ unknown using data cloning approach, binomial mixed model.

σ_v^2	$\sigma_\epsilon^2 = 1.0$	2.0
1.0	0.020	0.021
2.0	0.019	0.020

and average lengths of the $\hat{p}_{1T}$ for different confidence coefficients (Table 14).

5. Application

The performance of the data cloning is also evaluated by using a real dataset of binomial mixed model. We use a yearly dataset of childhood (age ≤ 20 years) asthma visits to hospital in the Canadian province of Manitoba during the 2000-2010 fiscal years. The population of Manitoba was stable during the study period from 1.15 million in 2000 to 1.20 million in 2010, with

Table 13 Percent relative bias of estimators of MSPE for ρ unknown using data cloning approach, binomial mixed model.

σ_v^2	$\sigma_\epsilon^2 = 1.0$	2.0
1.0	-4.0	-0.8
2.0	-6.1	-1.6

Table 14 Coverage probabilities (and average lengths) of the $\hat{p}_{1T}$ for ρ unknown with different confidence coefficients using data cloning approach, binomial mixed model.

σ_v^2	σ_ϵ^2	Confidence coefficient (average lengths)			
		0.90	0.95	0.98	0.99
1.0	1.0	0.880(0.444)	0.930(0.516)	0.963(0.594)	0.976(0.642)
	2.0	0.882(0.458)	0.932 (0.534)	0.959(0.614)	0.970(0.664)
2.0	1.0	0.867(0.413)	0.921(0.483)	0.953(0.559)	0.962(0.607)
	2.0	0.871(0.433)	0.921(0.507)	0.954(0.587)	0.964(0.637)

an average population of children of around 335,000. The province consisted of eleven Regional Health Authorities that were responsible for the delivery of health care services. These eleven regions were further sub-divided into 56 Regional Health Authorities Districts (RHAD) and these RHAD are used as area in our model. The number of children asthma visits totaled 14,690 over the study period with mean and median number of yearly cases per region of 26 and 17 (range 3 to 422), respectively. The region children population sizes varied from 290 to 175,300, with mean and median numbers of 5,998 and 2,488, respectively. We ignore the variation of geographical regions in this data analysis, and our focus is to apply our time-series and cross-sectional binomial mixed model to this dataset. The sample sizes for some regions are not large enough (even 0 in some regions) to produce the reliable estimates. In particular, we consider the following model

$$\log(\frac{p_{it}}{1-p_{it}}) = \alpha + v_i + u_{it} (t = 1, ..., 10; i = 1, ..., 56)$$

where α is overall mean over area and time, $v_i \overset{i.i.d.}{\sim} N(0, \sigma_v^2)$, and $u_{it} = \rho u_{i,t-1} + \epsilon_{it}$, with $|\rho| < 1$ and $\epsilon_{it} \overset{i.i.d.}{\sim} N(0, \sigma_\epsilon^2)$; noting that y_{it}, children asthma visits to hospital in the ith area at time t, has binomial distribution with parameters p_{it} and n_{it} where n_{it} is the corresponding population size. We first consider the estimates of model parameters by applying DC method. The estimates of the model parameters and associated standard errors are reported in Table 15. For this specific application, the number of clones was $K = 10$ to obtain MLE with number of iterations 50,000 for the convergence of the model parameters. One of the main features of the DC method is the ability to provide the prediction and prediction interval of small area parameters. We also provide 95% prediction interval of the rates of children

Table 15 Parameter estimates and standard errors (SE) of yearly children asthma visits to hospital 2000-2010 using data cloning approach, binomial mixed model.

Parameter	α	σ_v^2	ρ	σ_ϵ^2
Estimate	-5.089	0.237	0.881	0.067
SE	0.029	0.094	0.033	0.003

asthma visits to hospital for different areas in 2010 (Figure 1).

6. Concluding remarks

In small area estimation, there are many situations where observations are time-related counts or proportions. Often, for fitting complex models in small area estimation, Bayesian methods are advocated because they are computationally more convenient than the likelihood-based methods. Analysis based on data cloning overcomes the computational difficulties of the ML method. Torabi, Lele, and Prasad (2012; Unpublished work) applied the data cloning approach in the context of small area estimation with cross-sectional data in the class of GLMMs.

Using data cloning, we have proposed a generalized model involving autocorrelated random effects and sampling errors for small area estimation with utilizing both time-series and cross-sectional data. Under the linear mixed models, the data cloning approach leads to similar inferential solutions to small area parameters as Rao and Yu (1994) approach. Note that Rao and Yu (1994) had difficulties associated with the frequentist approach

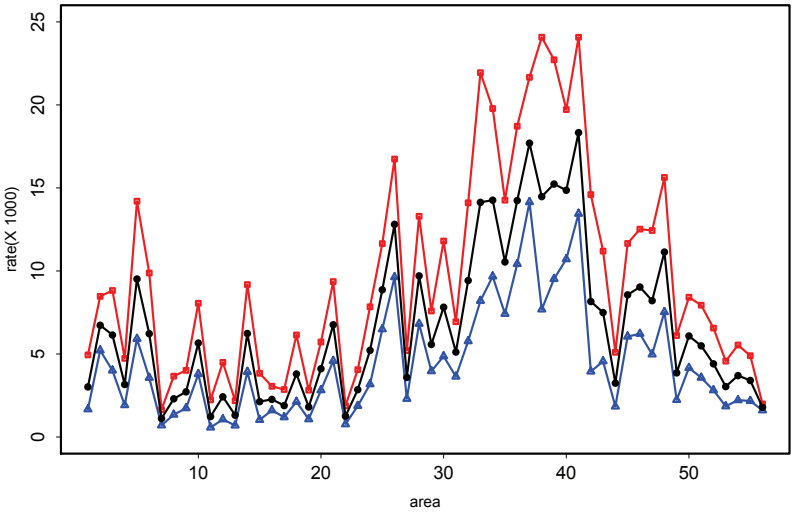

Figure 1: The 95% prediction interval of the rate of children asthma visits to hospital in 2010 using DC approach, binomial mixed model; black line is point estimate, blue and red lines are lower and upper bounds of prediction interval, respectively.

in estimating ρ, while this estimation was easily obtained via data cloning approach. Also, note that the method of Rao and Yu (1994) is not applicable for non-Normal data. Under the GLMM, our simulation results have shown that data cloning does very well in terms of relative bias of estimators of MSPE of small area parameters. The data cloning based prediction intervals also provided very good coverage probabilities and average lengths of the small area parameters. We also applied our proposed approach to a real dataset to evaluate the performance of data cloning in the binomial mixed model. The other advantage of data cloning compared to other approaches is that the non-estimable parameters are flagged automatically.

Acknowledgments

This work was done while second author visited first author as PhD student. This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. The comments of a referee is gratefully acknowledged.

Disclaimer: The interpretations, conclusions and opinions expressed in this paper are those of the authors and do not necessarily reflect the position of the Manitoba Health. This study is based in part on data provided by Manitoba Health through Manitoba Centre for Health Policy. The interpretation and conclusions contained herein are those of the researchers and do not necessarily represent the views of the government of Manitoba.

References

Battese, G., Harter, R., Fuller, W., 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.

Datta, G., Lahiri, P., Maiti, T., Lu, K., 1999. Hierarchical Bayes Estimation of Unemployment Rates for the States of the U.S. Journal of the American Statistical Association 94, 1074–1082.

Datta, G.S., Lahiri, P., Maiti, T., 2002. Empirical Bayes Estimation of Median Income of Four-Person Families by State Using Time Series and Cross-Sectional Data. Journal of Statistical Planning and Inference 102, 83 – 97.

Fay, I.R.E., 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.

Fuller, W.A., Battese, G.E., 1973. Transformations for Estimation of Linear Models with Nested-Error Structure. Journal of the American Statistical Association 68, 626–632.

Gilks, W., Richardson, S., Spiegelhalter, D., 1996. Markov Chain Monte Carlo in Practice. Chapman and Hall/CRC.

Hamilton, J.D., 1986. A Standard Error for the Estimated State Vector of a State-Space Model. Journal of Econometrics 33, 387 – 397.

Harris, I.R., 1989. Predictive Fit for Natural Exponential Families. Biometrika 76, 675–684.

Jones, R.H., 1980. Maximum Likelihood Fitting of ARMA Models to Time Series with Missing Observations. Technometrics 22, 389–395.

Lele, S.R., Dennis, B., Lutscher, F., 2007. Data Cloning: Easy Maximum Likelihood Estimation for Complex Ecological Models Using Bayesian Markov Chain Monte Carlo Methods. Ecology Letters 10, 551–563.

Lele, S.R., Nadeem, K., Schmuland, B., 2010. Estimability and Likelihood Inference for Generalized Linear Mixed Models Using Data Cloning. Journal of the American Statistical Association 105, 1617–1625.

Pfeffermann, D., Burck, L., 1990. Robust Small Area Estimation Combining Time Series and Cross-Sectional Data. Survey Methodology 16, 217–237.

Prasad, N.G.N., Rao, J.N.K., 1990. The Estimation of the Mean Squared Error of Small-Area Estimators. Journal of the American Statistical Association 85, 163–171.

Rao, J.N.K., 2003. Small Area Estimation. Wiley-Interscience.

Rao, J.N.K., Yu, M., 1994. Small-Area Estimation by Combining Time-Series and Cross-Sectional Data. The Canadian Journal of Statistics 22, 511–528.

Scott, A.J., Smith, T.M.F., 1974. Analysis of Repeated Surveys Using Time Series Methods. Journal of the American Statistical Association 69, 674–678.

Singh, A., Mantel, H., Thomas, B., 1991. Time Series Generalizations of Fay-Herriot Estimator for Small Areas, in: Proc. Survey Research Methods Section, American Statistical Association. pp. 455–460.

Smith, B.J., 2007. BOA: An R Package for MCMC Output Convergence Assessment and Posterior Inference. Journal of Statistical Software 21, 1–37.

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

Spiegelhalter, D.J., Abrams, K.R., Myles, J.P., 2004. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. Wiley.

Tiller, R.B., 1992. Time Series Modeling of Sample Survey Data from the U.S. Current Population Survey. Journal of Official Statistics 8, 149–166.

Torabi, M., 2012. Hierarchical Bayes Estimation of Spatial Statistics for Rates. Journal of Statistical Planning and Inference 142, 358 – 365.

Walker, A.M., 1969. On the asymptotic behaviour of posterior distributions. Journal of the Royal Statistical Society. Series B (Methodological) 31, 80–88.