Survey Methodology, December 2000
Vol. 26, No. 2, pp. 173-181
Statistics Canada, Catalogue No. 12-001

Hierarchical Bayes Estimation of Small Area Means Using Multi-Level Models

Yong You and J.N.K. Rao [1]

Abstract

Standard multi-level models with random regression parameters are considered for small area estimation. We also extend the models by allowing unequal error variances or by assuming random effect models for both regression parameters and error variances. We present these models in a hierarchical Bayes framework and estimate a small area mean by its posterior mean. Posterior variance of the small area mean is used as a measure of precision of the estimate. It automatically takes into account the extra uncertainty associated with the hyperparameters in the multi-level model. Gibbs sampling is used to compute the posterior means and posterior variances of small area means. Rao-Blackwellized estimators that reduce the Monte Carlo errors are obtained. Bayesian model selection and sensitivity analysis are also studied. The procedure is illustrated using data on household income in some counties (small areas) of Brazil.

Key Words: Gibbs sampling; Hierarchical Bayes; Multi-level model; Sampling error variance; Small area.

1. Introduction

Small area estimation has received a lot of attention in recent years due to growing demand for reliable small area estimators. Traditional area-specific direct estimators do not provide adequate precision because sample sizes in small areas are seldom large enough. This makes it necessary to employ indirect estimators that borrow strength from related areas; in particular, model-based indirect estimators. Battese, Harter and Fuller (1988) proposed and applied a nested error regression model to provide model-based small area estimates. The model takes the form

$$y_{ij} = x_{ij}^T \beta + v_{0i} + e_{ij},\ j = 1, \ldots, n_i; i = 1, \ldots, m, \qquad (1)$$

where y_{ij} are the observations associated with the sampled units in the i^{th} small area, $i = 1, \ldots, m$, x_{ij} is the $p \times 1$ vector of unit-level explanatory variables, β is a set of p fixed regression parameters, v_{0i} are independent area effects with $E(v_{0i}) = 0$ and $V(v_{0i}) = \sigma_v^2$. The e_{ij}'s are assumed to be independent random error variables with $E(e_{ij}) = 0$ and $V(e_{ij}) = \sigma_e^2$. v_{0i} and e_{ij} are also assumed to be independent. For the whole population, model (1) applies with n_i replaced by N_i, the small area population size. The model (1) may be expressed in matrix notation as follows

$$Y_i = X_i \beta + v_{0i} 1_{n_i} + e_i, i = 1, \ldots, m,$$

where $Y_i = (y_{i1}, \ldots, y_{i,n_i})^T$, $X_i = (x_{i1}, \ldots, x_{i,n_i})^T$ is a $n_i \times p$ matrix, $1_{n_i} = (1, \ldots, 1)^T$ is the unit vector of length n_i, and $e_i = (e_{i1}, \ldots, e_{i,n_i})^T$.

Holt and Moura (1993) extended the above framework to a multi-level model by introducing random regression coefficients and then relating them to area-level explanatory variables to explain some of the between small area variation. The model can be stated as follows:

$$Y_i = X_i \beta_i + e_i, \beta_i = Z_i \gamma + v_i \qquad (2)$$

where Z_i is the $p \times q$ design matrix of area-level variables, γ is a $q \times 1$ vector of fixed coefficients, and $v_i = (v_{i1}, \ldots, v_{ip})^T$ is a $p \times 1$ vector of random effects for the i^{th} area. The v_i's are independent across areas and have a joint distribution within each area with $E(v_i) = 0$ and $V(v_i) = \Phi$, where the variance covariance matrix Φ is unknown. Note that model (2) effectively integrates the use of unit-level and area-level covariates into a single model. Holt and Moura (1993) and Moura and Holt (1999) extended Prasad and Rao's (1990) framework to the above multi-level model to get the best linear unbiased predictor (BLUP) of the small area mean $\mu_i = \bar{X}_i^T \beta_i$ assuming that N_i is large, where $\bar{X}_i$ is the $p \times 1$ vector of known population means of the auxiliary variables for the i^{th} small area. They also obtained the empirical BLUP (EBLUP) and a second order approximation to the mean squared error (MSE) of EBLUP for the multi-level model. Using household income data in some counties (small areas) of Brazil, they demonstrated gain in efficiency of the EBLUP estimators over the EBLUP estimators obtained from nested error regression models. Ghosh and Rao (1994) and Rao (1999) provide a detailed overview of model-based methods for small area estimation.

In this paper, we study the multi-level model (2) in a hierarchical Bayes framework and extend the model to more general multi-level models which allow fixed unequal error variances or random error variances. The small area mean μ_i is estimated by its posterior mean and its precision is measured by its posterior variance. Posterior variance automatically takes into account the extra uncertainty

1. Yong You, Household Survey Methods Division, Statistics Canada, Ottawa, Ontario, Canada, K1A 0T6; J.N.K. Rao, School of Mathematics and Statistics, Carleton University, Ottawa, Ontario, Canada, K1S 5B6.

associated with the hyperparameters in the multi-level model. We use the Gibbs sampling method to compute the hierarchical Bayes estimates and the associated posterior variances. Section 2 presents the hierarchical Bayes multi-level models with different assumptions on error variances and related Gibbs sampling inference. Section 3 illustrates our methodology and studies model selection and sensitivity analysis by employing data on household incomes in some counties (small areas) of Brazil. And finally in section 4, we give some comments and concluding remarks.

2. Multi-Level Models and Gibbs Sampling Inference

2.1 Equal Error Variances

We consider a hierarchical Bayes representation of the multi-level model (2) as follows:

Model 1:

(i) Conditional on β_i and σ_e^2, y_{ij}'s are independent with

$$y_{ij}|\beta_i, \sigma_e^2 \sim N(x_{ij}^T\beta_i, \sigma_e^2), \quad (i=1,\ldots,m;\ j=1,\ldots,n_i); \tag{3}$$

(ii) Conditional on γ and Φ, β_i's are independent with

$$\beta_i|\gamma, \Phi \sim N_p(Z_i\gamma, \Phi), (i=1,\ldots,m). \tag{4}$$

To complete our Bayesian model specification, we adopt the prior distributions for parameters as follows:

(iii) Marginal prior distributions: $\gamma \sim N_q(0, D)$, $\tau_e \sim G(a, b)$ and $\Omega \sim W_p(\alpha, R)$, where $\tau_e = \sigma_e^{-2}$, $\Omega \sim \Phi^{-1}$, and D, a, b, α and R are known.

In step (iii) of Model 1, $G(a, b)$ denotes a gamma distribution with density given by $f(x) = b^a/\Gamma(a)\,x^{a-1}e^{-xb}$, $a>0, b>0, x\geq 0$, and $W_p(\alpha, R)$ is a Wishart distribution with density function

$$f(X) = \frac{|R|^{\frac{\alpha}{2}}}{2^{\alpha p/2}\Gamma_{p\left(\frac{\alpha}{2}\right)}}|X|^{\frac{\alpha-p-1}{2}}\exp\left\{-\frac{1}{2}\mathrm{tr}(RX)\right\},$$

where $X>0, R>0$ and $\Gamma_p(\alpha)$ is multivariate gamma function defined as

$$\Gamma_p(\alpha) = \pi^{\frac{p(p-1)}{4}}\prod_{j=1}^{p}\Gamma\left(\alpha+\frac{1}{2}(1-j)\right).$$

Remark 1.1: The prior distributions in step (iii) are conjugate with the sampling and population distributions given by (3) and (4) in the sense that they lead to full conditional distributions for γ, τ_e and Ω that are again normal, gamma and Wishart distribution, respectively. The Wishart distribution is the multivariate version of gamma distribution for the inverse variance covariance matrix of random effects. The importance of conjugacy may be exploited as follows: (1) In the Gibbs sampling step, without conjugacy the full conditional distribution for any parameter will be known up to normalizing constants. In this case, more sophisticated random generation will be required. (2) Closed-form full conditional distributions may be employed to find the Rao-Blackwellized estimators of the posterior means and posterior variances, and thus to improve posterior estimation. In general, for Bayesian inference, choosing priors is not a simple job because any proper prior on the model parameters is a plausible candidate. This is a limitation of Bayesian methods.

Remark 1.2: It is important to note that we have used proper priors on all the unknown parameters to ensure that all the posterior distributions are proper (Hobert and Casella 1996). Hence we do not face the problem of some posteriors being improper. Values for the parameters of the priors (*i.e.*, hyperparameters) are chosen to reflect a fairly vague knowledge of the prior distributions. Details will be given in section 3 on data analysis.

Remark 1.3: In Model 1, we assume equal error variance σ_e^2 for all small areas. In practice, however, variances of sampling error could be different for different small areas. A more general model should allow possibly different error variances. In sections 2.2 and 2.3, we will introduce unequal error variance and random error variance models.

We are interested in finding the posterior distributions of β_i's given the data $Y = (\{y_{ij}\}, i=1,\ldots,m;\ j=1,\ldots,n_i)$, and in particular in finding the posterior estimates of small area means $\mu_i = \bar{X}_i^T\beta_i$, which depend on the estimates of β_i. Direct evaluation of the joint posterior distribution involves high-dimensional numerical integration, and is not computationally feasible. Therefore, we use the Gibbs sampling method (Gelfand and Smith 1990) to generate samples from the joint posterior distributions. To implement the Gibbs sampling under Model 1, we need the full conditional distributions given by:

(i) For $i=1,\ldots,m$,

$$[\beta_i|Y,\gamma,\Omega,\tau] \overset{\text{ind}}{\sim} N_p((\tau_i X_i^T X_i+\Omega)^{-1}(\tau_i X_i^T Y_i+\Omega Z_i\gamma), (\tau_i X_i^T X_i+\Omega)^{-1})$$

(ii)
$$[\gamma|Y,\beta,\Omega,\tau_e] \sim N_q\left(\left(\sum_{i=1}^m Z_i^T\Omega Z_i + D^{-1}\right)^{-1}\left(\sum_{i=1}^m Z_i^T\Omega\beta_i\right), \left(\sum_{i=1}^m Z_i^T\Omega Z_i + D^{-1}\right)^{-1}\right)$$

(iii)
$$[\Omega|Y,\beta,\gamma,\tau] \sim W_p\left(\alpha+m, R+\frac{1}{2}\sum_{i=1}^m(\beta_i-Z_i\gamma)(\beta_i-Z_i\gamma)^T\right)$$

(iv)
$$[\tau_e|Y,\beta,\gamma,\Omega] \sim G\left(a+\frac{1}{2}\sum_{i=1}^m n_i,\ b+\frac{1}{2}\sum_{i=1}^m(Y_i-X_i\beta_i)^T(Y_i-X_i\beta_i)\right).$$

Since all the full conditional distributions have closed-form, it is easy to generate samples. Gibbs sampling method is as follows: (a) Using starting values $\gamma^{(0)}, \Omega^{(0)}$ and $\tau_e^{(0)}$, draw $\beta_i^{(1)}, i = 1, \ldots, m$, from $[\beta_i | Y, \gamma, \Omega, \tau_e]$; (b) Draw $\gamma^{(1)}$ from $[\gamma | Y, \beta, \Omega, \tau_e]$ using $\beta_i^{(1)}, i = 1, \ldots, m, \Omega^{(0)}$ and $\tau_e^{(0)}$; (c) Draw $\Omega^{(1)}$ from $[\Omega | Y, \beta, \gamma, \tau_e]$ using $\beta_i^{(1)}, i = 1, \ldots, m$, $\gamma^{(1)}$ and $\tau_e^{(0)}$; (d) Draw $\tau_e^{(1)}$ from $[\tau_e | Y, \beta, \gamma, \Omega]$ using $\beta_i^{(1)}, i = 1, \ldots, m, \gamma^{(1)}$ and $\Omega^{(1)}$. Steps (a)-(d) complete one sampling cycle. Perform a large number of cycles, say t, called "burn-in" period, until convergence and then treat $\{\beta_i^{(t+k)}, i = 1, \ldots, m; \gamma^{(t+k)}; \Omega^{(t+k)}; \tau_e^{(t+k)}; k = 1, \ldots, G\}$ as G samples from the joint posterior of $\beta_i, i = 1, \ldots, m, \gamma, \Omega$ and τ_e.

Suppose a sample of size G is obtained as $\{\beta_i^{(k)}, i = 1, \ldots, m; \gamma^{(k)}; \Omega^{(k)}; \tau_e^{(k)}; k = 1, \ldots, G\}$. To obtain an estimator of the posterior mean of β_i, one can use the sample mean of the $\{\beta_i^{(k)}\}$. Since β_i has a closed form full conditional distribution, we can use the sample mean of the conditional expectations $\{E[\beta_i | Y, \gamma^{(k)}, \Omega^{(k)}, \tau_e^{(k)}]\}$ to improve our estimation, since $E(\beta_i | Y) = E(E(\beta_i | Y, \gamma, \Omega, \tau_e))$, and $\mathrm{Var}(\beta_i | Y) \geq \mathrm{Var}(E(\beta_i | Y, \gamma, \Omega, \tau_e))$. This modification is based on the well-known Rao-Blackwell theorem and the corresponding estimator is the so-called Rao-Blackwellized estimator (Gelfand and Smith 1990, 1991). Thus we have the following two alternative estimators for β_i:

$$\hat{\beta}_i^{(E)} = \frac{1}{G} \sum_{k=1}^{G} \beta_i^{(k)} \tag{5}$$

and

$$\begin{aligned} \hat{\beta}_i^{(\mathrm{RB})} &= \frac{1}{G} \sum_{k=1}^{G} E(\beta_i | Y, \gamma^{(k)}, \Omega^{(k)}, \tau_e^{(k)}) \\ &= \frac{1}{G} \sum_{k=1}^{G} (\tau_e^{(k)} X_i^T X_i + \Omega^{(k)})^{-1} \\ &\quad (\tau_e^{(k)} X_i^T Y_i + \Omega^{(k)} Z_i \gamma^{(k)}), \end{aligned} \tag{6}$$

where $\hat{\beta}_i^{(E)}$ is the empirical estimator $\hat{\beta}_i^{(\mathrm{RB})}$ and is the Rao-Blackwellized estimator. Both $\hat{\beta}_i^{(E)}$ and $\hat{\beta}_i^{(\mathrm{RB})}$ are unbiased for the posterior mean. However, $\hat{\beta}_i^{(\mathrm{RB})}$ is better than $\hat{\beta}_i^{(E)}$ in terms of simulation standard error (Gelfand and Smith 1991).

The corresponding estimators for the small area mean μ_i are given as

$$\hat{\mu}_i^{(E)} = \bar{X}_i^T \hat{\beta}_i^{(E)} \frac{1}{G} \sum_{k=1}^{G} \bar{X}_i^T \beta_i^{(k)} \tag{7}$$

and

$$\begin{aligned} \hat{\mu}_i^{(\mathrm{RB})} = \bar{X}_i^T \hat{\beta}_i^{(\mathrm{RB})} &= \frac{1}{G} \sum_{k=1}^{G} \bar{X}_i^T (\tau_e^{(k)} X_i^T X_i + \Omega^{(k)})^{-1} \\ &\quad (\tau_e^{(k)} X_i^T Y_i + \Omega^{(k)} Z_i \gamma^{(k)}). \end{aligned} \tag{8}$$

We anticipate that both $\hat{\mu}_i^{(E)}$ and $\hat{\mu}_i^{(\mathrm{RB})}$ will give almost the same point estimates. However, it will be of interest to compute and compare the simulation standard errors of these two estimators to evaluate the effects of Rao-Blackwellization; see section 3.

To obtain the posterior variance of μ_i, we first find the posterior variance of β_i, since $V(\mu_i | Y) = \bar{X}_i^T V(\beta_i | Y) \bar{X}_i$. Note that

$$\begin{aligned} V(\beta_i | Y) &= E(V(\beta_i | Y, \gamma, \Omega, \tau_e)) + V(E(\beta_i | Y, \gamma, \Omega, \tau_e)) \\ &= E(V(\beta_i | Y, \gamma, \Omega, \tau_e)) + E(E(\beta_i | Y, \gamma, \Omega, \tau_e)^2) \\ &\quad - [E(E(\beta_i | Y, \gamma, \Omega, \tau_e))]^2. \end{aligned} \tag{9}$$

Using (9), the Rao-Blackwellized estimator of the posterior variance of β_i, denoted by $\hat{V}(\beta_i)$, can be obtained using the Gibbs samples $\{\beta_i^{(k)}, i = 1, \ldots, m; \gamma^{(k)}; \Omega^{(k)}; \tau_e^{(k)}; k = 1, \ldots, G\}$; see Appendix A1. The posterior variance of small area mean μ_i is then estimated by

$$\hat{V}(\mu_i) = \bar{X}_i^T \hat{V}(\beta_i) \bar{X}_i. \tag{10}$$

The same estimation procedure can be applied to the sampling error variance σ_e^2. Since conditionally σ_e^2 has an inverse gamma distribution, the Rao-Blackwellized estimator of the posterior mean of σ_e^2 is obtained as

$$\begin{aligned} \hat{\sigma}_e^{2(\mathrm{RB})} &= \frac{1}{G} \sum_{k=1}^{G} \left[b + \frac{1}{2} \sum_{i=1}^{m} (Y_i - X_i \beta_i^{(k)})^T (Y_i - X_i \beta_e^{(k)}) \right] \\ &\quad \left(a + \frac{1}{2} \sum_{i=1}^{m} n_i - 1 \right)^{-1}. \end{aligned} \tag{11}$$

Since we are mainly interested in estimating the small area means, calculation of the sampling variance is only for the purpose of model selection. Details on model selection will be given in section 3.2.

2.2 Unequal Error Variances

In practice, it is more realistic to allow unequal error variances for the sampling errors. Let σ_i^2 be the true sampling error variance for the i^{th} small area. A straightforward extension of Model 1 leads to the following hierarchical Bayes multi-level unequal error variance model:

Model 2:

(i) Conditional on β_i and σ_i^2, y_{ij}'s are independent with

$$\begin{aligned} &y_{ij} | \beta_i, \sigma_i^2 \sim N(x_{ij}^T \beta_i, \sigma_i^2), \\ &(i = 1, \ldots, m; j = 1, \ldots, n_i); \end{aligned} \tag{12}$$

(ii) Conditional on γ and Φ, β_i's are independent with

$$\beta_i | \gamma, \Phi \sim N_p(Z_i \gamma, \Phi), (i = 1, \ldots, m); \tag{13}$$

(iii) Marginal prior distributions: $\gamma \sim N_q(0, D)$, $\tau_i \overset{\text{ind}}{\sim} G(a_i, b_i)$, and $\Omega \sim W_p(\alpha, R)$, where $\tau_i = \sigma_i^{-2}$, $\Omega = \Phi^{-1}$, and D, a_i, b_i, α and R are known.

Remark 2.1: Model 2 reduces to Model 1 when $\sigma_i^2 = \sigma_e^2$ for all i. From a hierarchical Bayes perspective, extension from the equal error variance model to the unequal error variance model is straightforward. Also there is no difficulty in the Gibbs sampling implementation.

Remark 2.2: τ_i 's are assumed to be independent and have prior distributions $G(a_i, b_i)$, where a_i and b_i are known hyperparameters and usually chosen to be very small to reflect a vague knowledge about τ_i 's.

The full conditional distributions for Gibbs sampling under Model 2 are given by:

(i) For $i = 1, \ldots, m$,

$$[\beta_i | Y, \gamma, \Omega, \tau] \overset{\text{ind}}{\sim} N_p((\tau_i X_i^T X_i + \Omega)^{-1} (\tau_i X_i^T Y_i + \Omega Z_i \gamma), (\tau_i X_i^T X_i + \Omega)^{-1})$$

(ii) $$[\gamma | Y, \beta, \Omega, \tau] \sim N_q\left(\left(\sum_{i=1}^m Z_i^T \Omega Z_i + D^{-1}\right)^{-1} \left(\sum_{i=1}^m Z_i^T \Omega \beta_i\right), \left(\sum_{i=1}^m Z_i^T \Omega Z_i + D^{-1}\right)^{-1}\right)$$

(iii) $$[\Omega | Y, \beta, \gamma, \tau] \sim W_p\left(\alpha + m, R + \frac{1}{2}\sum_{i=1}^m (\beta_i - Z_i\gamma)(\beta_i - Z_i\gamma)^T\right)$$

(iv) For $i = 1, \ldots, m$,

$$[\tau_i | Y, \beta, \gamma, \Omega] \sim G(a_i + \frac{1}{2} n_i, b_i + \frac{1}{2} \times (Y_i - X_i \beta_i)^T (Y_i - X_i \beta_i)).$$

For Model 2, the empirical estimators of the posterior means of β_i and μ_i have the same form as (5) and (7). The Rao-Blackwellized estimators $\hat{\beta}_i^{(\text{RB})}$ and $\hat{\mu}_i^{(\text{RB})}$ are the estimators given by (6) and (8) with $\tau_e^{(k)}$ replaced by $\tau_i^{(k)}$. Estimator of posterior variance is $\hat{V}(\mu_i)$ given by (10) with $\tau_e^{(k)}$ replaced by $\tau_i^{(k)}$.

For the purpose of model selection and model comparison, we also find the Rao-Blackwellized estimator of the posterior mean of σ_i^2 under Model 2 as

$$\hat{\sigma}_i^{2(\text{RB})} = \frac{1}{G}\sum_{k=1}^G \left[b_i + \frac{1}{2}(Y_i - X_i\beta_i^{(k)})^T (Y_i - X_i\beta_i^{(k)})\right] \times \left(a_i + \frac{1}{2} n_i - 1\right)^{-1}. \quad (14)$$

2.3 Random Error Variances

In Model 2, we assumed unequal error variances for the sampling errors. Kleffe and Rao (1992) used a simple random error variance model to derive the best linear unbiased predictors for small area means. In this section we extend their model to the multi-level case. We assume random effect models on both regression coefficients β_i and sampling error variances σ_i^2, which leads to Model 3 given below.

Model 3:

(i) Same as in Model 2;

(ii) Same as in Model 2;

(iii) Conditional on η and λ, τ_i 's are independent with

$$\tau_i | \eta, \lambda \overset{\text{iid}}{\sim} G(\eta, \lambda), \quad (15)$$

where $\tau_i = \sigma_i^{-2}$;

(iv) Marginal prior distributions: $\gamma \sim N_q(0, D)$, $\Omega \sim W_p(\alpha, R)$, $\eta \sim U^+$ and $\lambda \sim U^+$, where U^+ denotes a uniform distribution over a subset of R^+ with large but finite length, D, α and R are known.

Remark 3.1: In Model 3, we assume that τ_i 's are iid gamma random variables with unknown hyperparameters η and λ. Thus we have population models for both regression coefficient β_i and sampling variance σ_i^2. In Model 1 and Model 2, we considered modelling β_i only and assumed vague proper prior distributions on σ_e^2 or σ_i^2.

Remark 3.2: Assumption (iii) may not be a good population model for all τ_i 's. Alternatively, we can model τ_i in a more realistic way, as in the case of β_i, by specifying a regression model for the logarithm of τ_i. This may require some auxiliary information related to τ_i. In the data analysis of section 3, however, we simply used $G(\eta, \lambda)$ as the population model for τ_i. Generally it is not easy to model the sampling variances when they are unknown.

The full conditional distributions for Gibbs sampling under Model 3 are given by:

(i) For $i = 1, \ldots, m$,

$$[\beta_i | Y, \tau, \gamma, \Omega, \eta, \lambda] \overset{\text{ind}}{\sim} N_p((\tau_i X_i^T X_i + \Omega)^{-1} (\tau_i X_i^T Y_i + \Omega Z_i \gamma), (\tau_i X_i^T X_i + \Omega)^{-1})$$

(ii) For $i = 1, \ldots, m$,

$$[\tau_i | Y, \beta, \gamma, \Omega, \eta, \lambda] \overset{\text{ind}}{\sim} G\left(\eta + \frac{n_i}{2}, \frac{1}{2}(Y_i - X_i\beta_i)^T (Y_i - X_i\beta_i) + \lambda\right)$$

(iii) $$[\gamma | Y, \beta, \tau, \eta, \lambda] \sim N_q\left(\left(\sum_{i=1}^m Z_i^T \Omega Z_i + D^{-1}\right)^{-1} \left(\sum_{i=1}^m Z_i^T \Omega \beta_i\right), \left(\sum_{i=1}^m Z_i^T \Omega Z_i + D^{-1}\right)^{-1}\right)$$

(iv) $$[\Omega | Y, \beta, \sigma^2, \gamma, \eta, \lambda] \sim W_p\left(\alpha + m, R + \frac{1}{2}\sum_{i=1}^m (\beta_i - Z_i\gamma)(\beta_i - Z_i\gamma)^T\right)$$

(v) $$[\eta \mid Y, \beta, \tau, \gamma, \Omega, \lambda] \propto [\Gamma(\eta)]^{-m} \lambda^{m\eta} \left(\prod_{i=1}^{m} \tau_i \right)^{\eta}$$

(vi) $$[\lambda \mid Y, \beta, \tau, \gamma, \Omega, \eta] \sim G\left(m\eta + 1, \sum_{i=1}^{m} \tau_i \right).$$

For Model 3, the posterior estimators of β_i and μ_i have the same forms as those given for Model 2. Under Model 3, the Rao-Blackwellized estimator of the posterior mean of σ_i^2 is given by

$$\hat{\sigma}_i^{2(\mathrm{RB})} = \frac{1}{G} \sum_{k=1}^{G} [\lambda^{(k)} + \frac{1}{2} (Y_i - X_i \beta_i^{(k)})^T \times (Y_i - X_i \beta_i^{(k)})] \left(\eta^{(k)} + \frac{1}{2} n_i - 1 \right)^{-1}. \quad (16)$$

Under Model 3, $[\eta \mid Y, \beta, \tau, \gamma, \Omega, \lambda]$ is known only up to a multiplicative constant. However, since $[\eta \mid Y, \beta, \tau, \gamma, \Omega, \lambda]$ is a log-concave function of η (see Appendix A2), adaptive rejection sampling method of Gilks, Best and Tan (1995) can be used in the Gibbs sampler to generate samples from the conditional distribution $[\eta \mid Y, \beta, \tau, \gamma, \Omega, \lambda]$.

3. Data Analysis

3.1 Data and Model Description

Following Holt and Moura (1993) and Moura and Holt (1999), we considered the estimation of household income in some counties (small areas) of Brazil. Holt and Moura's original data contains 140 small areas with the sampling units taken from each area by simple random sampling. The hierarchical Bayes method does not require the number of small areas to be large, unlike in the case of EBLUP method, for getting standard errors. Therefore, we used only a small part of the original data set in our data analysis for simple illustration. Our data set contains a subset of 10 small areas with 28 sampling units obtained by simple random sampling in each area.

Let y_{ij} denote the j^{th} household's income in the i^{th} small area. There are two unit level auxiliary variables, namely x_1 and x_2, where x_1 denotes the number of rooms in a household and x_2 denotes the educational attainment of Head of Household. The sampling model is given by

$$y_{ij} = x_{ij}^T \beta_i + e_{ij} = \beta_{0i} + x_{1ij} \beta_{1i} + x_{2ij} \beta_{2i} + e_{ij}, \quad (17)$$

where x_{1ij} denotes the number of rooms in the j^{th} household of small area i and x_{2ij} denotes the corresponding educational attainment of Head of Household. Values of x_{1ij} and x_{2ij} are centered around their respective overall sample means and e_{ij} is the sampling error variable with its distribution specified by the three error variance models discussed in section 2.

In the sampling model (17), β_i is the random regression coefficient corresponding to the i^{th} small area and is modelled as

$$\beta_{0i} = \gamma_0 + v_{0i}, \beta_{1i} = \gamma_{10} + \gamma_{11} z_i + v_{1i}, \beta_{2i} = \gamma_{20} + \gamma_{21} z_i + v_{2i},$$

where $\gamma = (\gamma_0, \gamma_{10}, \gamma_{11}, \gamma_{20}, \gamma_{21})^T$ is the unknown vector of fixed regression parameters, $v_i = (v_{0i}, v_{1i}, v_{2i})^T$ is the i^{th} small area random effect vector distributed as $v_i \sim N_3(0, \Phi)$, and z_i is an area level variable defined as the average number of cars per household in each small area. Value of z_i is also centered around its overall sample mean.

We used the three models discussed in section 2 for our data analysis. Vague proper prior distributions on unknown parameters are specified as follows: $\gamma \sim N_5(0, D)$ where $D = \mathrm{diag}(10^4, 10^4, 10^4, 10^4, 10^4)$, thus $\gamma_0, \gamma_{10}, \gamma_{11}, \gamma_{20}, \gamma_{21}$ are assumed to be independent normal variables with a mean of 0 and a standard deviation of 100, so that a 95% prior interval is around ± 200, and the prior will be locally uniform over the region supported by the likelihood. Alternatively a uniform prior on a suitably wide interval could be given, such as U(−200, 200). A Wishart prior $W_3(\alpha, R)$ is specified for the inverse covariance matrix $\Omega = \Phi^{-1}$. To represent vague prior knowledge, we have chosen the degrees of freedom α for this distribution to be as small as possible, *i.e.*, $\alpha = 3$, the rank of Ω (Spiegelhalter, Thomas, Best and Gilks 1996). The scale matrix R is specified with diagonal elements equal to 1 and off-diagonal elements equal to 0.001, which represents our prior guess at the order of magnitude of the covariance matrix. For Model 1 and Model 2, a gamma prior G(0.001, 0.001) is assumed for τ_e and τ_i's. For Model 3, $\tau_i \sim G(\eta, \lambda)$, and η and λ are assumed to be independently distributed as U(0, 10,000), *i.e.*, the uniform distribution over a large interval. We anticipate that the vague proper priors on the hyperparameters would approximate the flat priors reasonably well and thus would have minimal effect on the posterior estimation.

We implemented the Gibbs sampler for the three models using the BUGS program (Spiegelhalter *et al.* 1996), aided by CODA Splus function (Best, Cowles and Vines 1996) for assessing convergence. The BUGS program constructs the necessary full conditional distributions and carries out the Gibbs sampling as long as we specify our models using the BUGS language. Priors and initial values of the parameters must be specified in the program. For each model, the Gibbs sampler was first run for a "burn-in" period of 2,000 iterations, then 5,000 more iterations were run and kept for model analysis and estimation.

Our interest is to estimate the small area mean $\mu_i = \bar{X}_i^T \beta_i = \beta_{0i} + \bar{X}_{1i} \beta_{1i} + \bar{X}_{2i} \beta_{2i}$, where $\bar{X}_{1i}$ and $\bar{X}_{2i}$ are the i^{th} small area population means of the auxiliary variables x_1 and x_2, respectively. For this, we will first select a model for the data set, then we will present the model-based estimates for the small area means based on the selected model.

3.2 Model Selection

We have proposed three models in section 2 based on different assumptions on sampling variances. To examine

which model fits the data, we first obtained the posterior estimates of the sampling variances under the three models. We also calculated the ordinary least square (OLS) estimates of the sampling variances within each area using only the area-specific data. Table 1 shows the Rao-Blackwellized estimates of the sampling variances under the three models as well as the OLS estimates.

Table 1
Estimated Sampling Error Variances

Area	OLS	Model 1	Model 2	Model 3
1	38.17	76.86	40.18	63.60
2	31.75	76.86	34.24	62.13
3	81.26	76.86	94.77	79.58
4	48.73	76.86	52.01	67.27
5	115.98	76.86	121.65	87.70
6	90.74	76.86	94.35	79.78
7	101.67	76.86	101.67	82.14
8	135.65	76.86	159.94	97.96
9	59.10	76.86	63.37	70.57
10	62.86	76.86	65.72	71.22

From Table 1, the OLS estimates indicate large variations among the ten small areas. Model 1 assumes an equal error variance σ_e^2 for all areas and σ_e^2 is estimated by $\hat{\sigma}_e^{2(\mathrm{RB})} = 76.86$, which is much smaller than the OLS estimates for some areas. Model 2 assumes unequal error variance σ_i^2 across areas. Under Model 2, the estimated error variances $\hat{\sigma}_i^{2(\mathrm{RB})}$ to some extent show the feature of the areas; $\hat{\sigma}_i^{2(\mathrm{RB})}$ are consistent with the pattern of the OLS estimates. The most notable result is $\hat{\sigma}_5^{2(\mathrm{RB})} = 121.65$ and $\hat{\sigma}_8^{2(\mathrm{RB})} = 159.94$, which show that there are larger variations within small areas 5 and 8. Model 3 assumes σ_i^2's to be random variables distributed as $G(\eta, \lambda)$. Under Model 3, all $\hat{\sigma}_i^{2(\mathrm{RB})}$ tend to be equal to and have moved toward $\hat{\sigma}_e^{2(\mathrm{RB})} = 76.86$. The results in Table 1 suggest that Model 2, the unequal error variance model, could be the best model for our data set. For further investigation, we now present a cross-validation study to select a best fit model.

In order to study how the data support each model, we calculated the cross-validation predictive densities for each data point y_{ij}. The cross-validation density for y_{ij} is the conditional density $f(y_{ij}|Y_{(ij)})$, where $Y_{(ij)}$ denotes all data except y_{ij}. We looked at the value of $f(y_{ij}|Y_{(ij)})$ at the observed data point, the so called conditional predictive ordinate, or CPO, for each of the three models. That is

$$\mathrm{CPO}_{ij} = f(y_{ij,\mathrm{obs}}|Y_{(ij),\mathrm{obs}}),$$

where $y_{ij,\mathrm{obs}}$ denotes the observed data point. Since CPOs are nothing but the observed likelihoods, models with larger CPOs provide better fit to the observed data. By using the output from the Gibbs sampler, we can calculate the CPOs for all data points. For example, under Model 1, we have

$$\begin{aligned} f(y_{ij}|Y_{(ij)}) &= \frac{f(Y)}{f(Y_{(ij)})} \\ &= \frac{1}{\int \frac{f(Y_{(ij)}, \beta_i, \sigma_e^2)}{f(Y, \beta_i, \sigma_e^2)} \cdot f(\beta_i, \sigma_e^2|Y) d\beta_i d\sigma_e^2} \\ &= \frac{1}{\int \frac{1}{f(y_{ij}|Y_{(ij)}, \beta_i, \sigma_e^2)} \cdot f(\beta_i, \sigma_e^2|Y) d\beta_i d\sigma_e^2}. \end{aligned}$$

Now noting that the y_{ij}'s are conditionally independent, *i.e.*, $f(y_{ij}|Y_{(ij)}, \beta_i, \sigma_e^2) = f(y_{ij}|\beta_i, \sigma_e^2)$, the CPO values are calculated as

$$\widehat{\mathrm{CPO}}_{ij} = \frac{1}{\frac{1}{G}\sum_{k=1}^{G} \frac{1}{f(y_{ij,\mathrm{obs}}|\beta_i^{(k)}, \sigma_e^{2(k)})}} \qquad (18)$$

where $f(y_{ij}|\beta_i, \sigma_e^2)$ is the normal density function given by (3). For Model 2 and Model 3, the CPOs are calculated with $\sigma_e^{2(k)}$ replaced by $\sigma_i^{2(k)}$ in (18). More detailed discussion can be found in Gelfand (1995).

We present a CPO plot for the three models in Figure 1. Clearly Model 2 is the best model among the three, because a majority of CPO values for Model 2 are significantly larger than those for Model 1 and Model 3. Model 3 is slightly better than Model 1 in terms of CPO values. Also there are small CPO values for all three models, which indicate that our model assumptions may not be very well satisfied by our data set.

According to the sampling variance estimates given in the Table 1 and the CPO plot, we conclude that Model 2 is a good model for our data. Therefore, we used Model 2 to find model-based estimates of small area means and associated posterior standard errors.

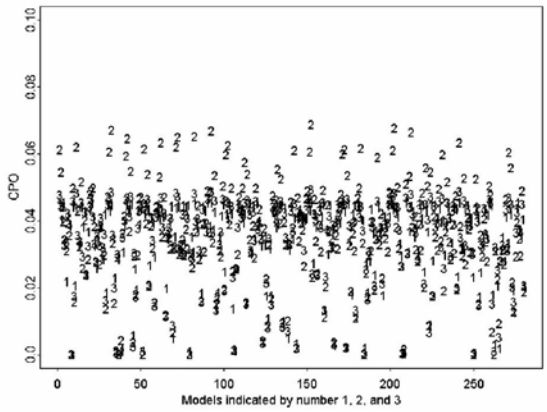

Figure 1. Model selection: CPO comparison plot.

3.3 Result of Estimation

We now present the estimates of the small area means based on Model 2 only. Table 2 presents the estimated posterior small area means and the corresponding posterior

standard errors. Our study found that the empirical estimator $\hat{\mu}_i^{(E)}$ and the Rao-Blackwellized estimator $\hat{\mu}_i^{(\mathrm{RB})}$ gave almost the same point estimates, thus we only reported the estimates obtained by using $\hat{\mu}_i^{(\mathrm{RB})}$. For comparison, we also calculated the direct estimates (sample means) and corresponding direct standard errors for the ten areas. It is clear from Table 2 that the model-based estimates are substantially more efficient than the direct estimates. The posterior standard errors are much smaller than the direct standard errors.

Table 2
Estimates of Small Area Means

Area	$\bar{y}_i$	s.e.	$\hat{\mu}_i^{(\mathrm{RB})}$	s.e.
1	11.08	9.53	10.23	0.81
2	7.91	6.82	9.84	0.85
3	13.48	14.15	13.01	1.08
4	6.53	8.01	10.95	1.11
5	19.52	14.96	17.87	1.57
6	11.21	11.38	10.21	0.93
7	8.72	11.24	9.58	0.97
8	12.81	13.99	10.30	1.19
9	10.18	8.76	11.34	1.01
10	10.01	11.30	9.79	0.87

In order to study the effects of Rao-Blackwellization, we calculated the simulation standard errors of $\hat{\mu}_i^{(E)}$ and $\hat{\mu}_i^{(\mathrm{RB})}$, which are respectively the sample standard errors of $\{\bar{X}_i^T \hat{\beta}_i^{(k)}\}$ and $\{\bar{X}_i^T E[\beta_i | Y, \gamma^{(k)}, \Omega^{(k)}, \tau_e^{(k)}]\}$. Table 3 presents the simulation standard errors. It is clear from Table 3 that the Rao-Blackwellized estimator $\hat{\mu}_i^{(\mathrm{RB})}$ has much smaller simulation standard error than the empirical estimator $\hat{\mu}_i^{(E)}$ for all areas. In all cases the standard error of $\hat{\mu}_i^{(\mathrm{RB})}$ is about 50% to 75% of the standard error of $\hat{\mu}_i^{(E)}$, demonstrating the benefit of Rao-Blackwellization. Thus $\hat{\mu}_i^{(\mathrm{RB})}$ is more stable than $\hat{\mu}_i^{(E)}$ when used to produce point estimates for the posterior means in computational Bayesian analysis. It should be mentioned that the simulation standard error of $\hat{\mu}_i^{(E)}$ is also an estimator of the posterior standard error. Thus the simulation standard error of $\hat{\mu}_i^{(E)}$ in Table 3 is almost identical to the estimated standard error of $\hat{\mu}_i^{(\mathrm{RB})}$ in Table 2.

Table 3
Simulation Standard Errors

Area	$\hat{\mu}_i^{(E)}$	$\hat{\mu}_i^{(\mathrm{RB})}$
1	0.817	0.506
2	0.862	0.498
3	1.090	0.548
4	1.101	0.604
5	1.583	0.878
6	0.930	0.481
7	0.978	0.480
8	1.208	0.842
9	0.997	0.524
10	0.869	0.513

3.4 Sensitivity Analysis

In Model 2, the error variances $\tau_i = \sigma_i^{-2}$ are assumed to be independent with prior distributions $\mathrm{G}(a_i, b_i)$ or σ_i^2 with the inverse gamma $\mathrm{IG}(a_i, b_i)$, where a_i and b_i are known values chosen to reflect our prior knowledge about σ_i^2. In practice, it is always difficult to obtain accurate information about the sampling variances. Also, as the number of small areas m increases, the number of variance components σ_i^2 will increase. We are interested in the possible effects caused by the choice of priors on σ_i^2's; in particular, we would like to evaluate the sensitivity of the posterior means to the choice of priors on the sampling variances σ_i^2. In our data analysis, a_i and b_i were chosen to be 0.001. Thus we used proper priors with very small parameter values for the variance components to reflect our vague knowledge about σ_i^2. In order to test the sensitivity of the posterior estimates to the choice of a_i and b_i under Model 2, we set $a_i = b_i$ at six different values, *i.e.*, 0.0001, 0.001, 0.01, 0.1, 1, and 10. Since

$$\{\tau_i | Y, \beta, \gamma, \Omega] \sim G(a_i + \frac{1}{2} n_i, b_i + \frac{1}{2}(Y_i - X_i\beta_i)^T (Y_i - X_i\beta_i)), \quad (19)$$

the sample effects $n_i/2$ and $(Y_i - X_i\beta_i)^T (Y_i - X_i\beta_i)/2$ dominate the prior information a_i and b_i when a_i and b_i are small. Thus IG(0.0001, 0.0001), IG(0.001, 0.001), and IG(0.01, 0.01) may be viewed as noninformative priors whereas IG(1, 1) and IG(10, 10) may be regarded as informative priors. Table 4 presents posterior means under Model 2 using the different priors on σ_i^2, and Table 5 presents the corresponding posterior variances.

Table 4
Comparison of Estimated Small Area Means

Small Area	$\mathrm{IG}(a_i, b_i)$, $a_i = b_i$					
	0	0	0.01	0.1	1	10
1	10.23	10.23	10.23	10.24	10.25	10.37
2	9.84	9.84	9.84	9.83	9.82	9.62
3	13.00	13.00	13.01	13.01	13.07	13.09
4	10.95	10.95	10.95	10.95	10.94	10.61
5	17.86	17.87	17.85	17.76	17.78	18.27
6	10.21	10.21	10.21	10.21	10.25	10.28
7	9.58	9.58	9.59	9.58	9.63	9.57
8	10.29	10.30	10.30	10.26	10.37	10.86
9	11.34	11.34	11.35	11.32	11.32	11.23
10	9.79	9.79	9.80	9.79	9.82	9.92

It is clear from Table 4 that the small area mean estimates are very stable: they are not sensitive to the choice of a_i and b_i. However, as shown in Table 5, the posterior variances decrease as the priors on σ_i^2 become more informative, and lead to smaller coefficients of variation (CV). This indicates that we can improve estimation results for small areas in terms of CV if we have more prior information on the sampling error variances. In our study, we only considered the case $a_i = b_i$. A more extensive study would involve different combinations of a_i and b_i.

Table 5
Comparison of Estimated Posterior Variances

Small Area	IG(a_i, b_i), $a_i = b_i$					
	0.0001	0.001	0.01	0.1	1	10
1	0.658	0.658	0.658	0.656	0.653	0.499
2	0.724	0.724	0.724	0.711	0.684	0.462
3	1.167	1.167	1.167	1.161	1.152	0.917
4	1.220	1.220	1.218	1.217	1.202	0.919
5	2.455	2.455	2.454	2.462	2.139	1.335
6	0.871	0.870	0.870	0.830	0.826	0.699
7	0.933	0.933	0.931	0.930	0.914	0.779
8	1.418	1.417	1.418	1.375	1.351	1.337
9	1.015	1.014	1.014	1.011	0.975	0.790
10	0.760	0.760	0.760	0.750	0.745	0.613

Table 6 presents the posterior estimates of σ_i^2 using the different priors on σ_i^2. As we can see from Table 6, when a_i and b_i are small (<= 0.01), there is almost no difference among the estimates at all. As a_i and b_i increase, the estimates $\hat{\sigma}_i^{2(\mathrm{RB})}$ become smaller. However, if there is strong prior information on a_i and b_i, for example, $a_i = b_i = 10$, then the posterior estimates of σ_i^2 will be significantly different from the ones under noninformative priors.

Table 6
Comparison of Estimated Sampling Error Variances

Small Area	IG(a_i, b_i), $a_i = b_i$					
	0.0001	0.001	0.01	0.1	1	10
1	40.09	40.10	40.05	39.64	37.14	22.29
2	34.19	34.18	34.17	33.97	31.74	19.05
3	94.48	94.49	94.42	93.76	86.73	50.60
4	52.08	52.08	52.04	51.63	48.21	28.82
5	121.60	121.70	121.60	121.40	113.70	66.75
6	94.03	94.03	93.83	92.96	87.21	52.90
7	102.30	102.30	102.20	101.40	94.85	57.58
8	160.10	160.00	159.90	159.10	147.60	86.61
9	63.46	63.46	63.38	62.99	58.46	34.85
10	65.88	65.87	65.89	65.40	60.76	36.60

4. Concluding Remarks

In this paper, we have presented hierarchical Bayes methods for small area estimation, using multi-level models. Clearly it is not easy to provide a suitable model for all small areas with satisfactory results, even if the Markov Chain Monte Carlo (MCMC) Bayesian methods such as the Gibbs sampling enable us to fit the data using Bayesian models of virtually unlimited complexity. The size and homogeneity of the areas and the availability of good auxiliary information will affect the final results. Models which prove suitable in some situations may be unsuitable in others. The hierarchical Bayes method also has some limitations such as the choice of priors on the model parameters and some sampling issues related to the Gibbs sampling method. Nevertheless, the general hierarchical Bayes methodology is applicable to a wide variety of situations for estimation of small area parameters. Model selection and choice is an important part of the hierarchical Bayes analysis. It is also important to compare the hierarchical Bayes method with other widely used methods in small area estimation, such as empirical Bayes (EB) and empirical best linear unbiased prediction (EBLUP). Work is in progress on extending our work to account for survey design weights, along the lines of You and Rao (1999).

Acknowledgements

We would like to thank two referees and the Editor for their helpful comments and suggestions. We also would like to thank Professor F. Moura of Federal University of Rio de Janeiro in Brazil for providing the data set used in section 3. This work was partially supported by a research grant from the Natural Sciences and Engineering Research Council of Canada.

Appendix

A1:

The Rao-Blackwellized estimator of the posterior variance of β_i is given by:

$$\hat{V}(\beta_i) = \frac{1}{G}\sum_{k=1}^{G} V(\beta_i|Y,\gamma^{(k)},\Omega^{(k)},\tau_e^{(k)})$$

$$+\frac{1}{G}\sum_{k=1}^{G}[E(\beta_i|Y,\gamma^{(k)},\Omega^{(k)},\tau_e^{(k)})]^2$$

$$-\left[\frac{1}{G}\sum_{k=1}^{G}E(\beta_i|Y,\gamma^{(k)},\Omega^{(k)},\tau_e^{(k)})\right]^2$$

$$=\frac{1}{G}\sum_{k=1}^{G}(\tau_e^{(k)}X_i^TX_i+\Omega^{(k)})^{-1}$$

$$+\frac{1}{G}\sum_{k=1}^{G}(\tau_e^{(k)}X_i^TX_i+\Omega^{(k)})^{-1}(\tau_e^{(k)}X_i^TY_i+\Omega^{(k)}Z_i\gamma^{(k)})$$

$$\times(\tau_e^{(k)}X_i^TY_i+\Omega^{(k)}Z_i\gamma^{(k)})^T(\tau_e^{(k)}X_i^TX_i+\Omega^{(k)})^{-1}$$

$$-\frac{1}{G^2}\left[\sum_{k=1}^{G}(\tau_e^{(k)}X_i^TX_i+\Omega^{(k)})^{-1}(\tau_e^{(k)}X_i^TY_i+\Omega^{(k)}Z_i\gamma^{(k)})\right]$$

$$\times\left[\sum_{k=1}^{G}(\tau_e^{(k)}X_i^TX_i+\Omega^{(k)})^{-1}(\tau_e^{(k)}X_i^TY_i+\Omega^{(k)}Z_i\gamma^{(k)})\right]^T.$$

A2:

Lemma: $[\eta|Y, \beta, \tau, \gamma, \Omega, \lambda]$ is a log-concave function of η. Proof: Let $h(\eta) = \log[\eta|Y, \beta, \tau, \gamma, \Omega, \lambda]$. It is enough to show that

$$\frac{\partial^2 h(\eta)}{\partial^2 \eta} \leq 0.$$

Clearly,

$$\frac{\partial h(\eta)}{\partial \eta} = -m\frac{\Gamma'(\eta)}{\Gamma(\eta)} + m\log(\lambda) + \log\left(\prod_{i=1}^{m} \tau_i\right).$$

Let $\psi(\eta) = \Gamma'(\eta)/\Gamma(\eta)$, then we have

$$\frac{\partial^2 h(\eta)}{\partial^2 \eta} = -m\psi'(\eta) \leq 0$$

since $m > 0$ and $\psi'(\eta)$ is positive on $(0, \infty)$ (Temme 1994, 54-55).

References

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

Best, N., Cowles, M.K., and Vines, K. (1996). CODA, *Convergence Diagnosis and Output Analysis Software for Gibbs Sampling Output*, Version 0.30. MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR.

Gelfand, A.E. (1995). Model determination using sampling-based methods. In *Markov Chain Monte Carlo in Practice* (Eds., W.R. Gilks, S. Richardson, and D.J. Spiegelhalter), 145-161. London: Chapman and Hall.

Gelfand, A.E., and Smith, A.F.M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398-409.

Gelfand, A.E., and Smith, A.F.M. (1991). Gibbs sampling for marginal posterior expections. *Communications In Statistics - Theory and Methods*, 20, 1747-1766.

Ghosh, M., and Rao, J.N.K. (1994). Small area estimation: An appraisal (with discussion). *Statistical Science*, 9, 55-93.

Gilks, W.R., Best, N.G., and Tan, K.K.C. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. *Journal of Applied Statistics*, 44, 455-472.

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

Holt, D., and Moura, F. (1993). Small area estimation using multilevel models. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 21-30.

Kleffe, J., and Rao, J.N.K. (1992). Estimation of mean square error of empirical best linear unbiased predictors under a random error variance linear model. *Journal of Multivariate Analysis*, 43, 1-15.

Moura, F., and Holt, D. (1999). Small area estimation using multilevel models. *Survey Methodology*, 25, 73-80.

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

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

Spiegelhalter, D., Thomas, A., Best, N., and Gilks, W. (1996). BUGS 0.5, *Bayesian Inference Using Gibbs Sampling Manual*. MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR.

Temme, N.M. (1994). *Special Functions: An Introduction to the Classical Functions of Mathematical Physics*. New York: John Wiley & Sons, Inc.

You, Y., and Rao, J.N.K. (1999). Pseudo hierarchical Bayes small area estimation using sampling weights. *Proceedings of the Survey Methods Section, Statistical Society of Canada*, 117-122.