Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Robust Bayesian small area estimation

by Malay Ghosh, Jiyoun Myung and Fernando A.S. Moura

Release date: June 21, 2018

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at STATCAN.infostats-infostats.STATCAN@canada.ca

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

• Statistical Information Service	1-800-263-1136
• National telecommunications device for the hearing impaired	1-800-363-7629
• Fax line	1-877-287-4369

Depository Services Program

• Inquiries line	1-800-635-7943
• Fax line	1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard table symbols

The following symbols are used in Statistics Canada publications:

. not available for any reference period
.. not available for a specific reference period
... not applicable
0 true zero or a value rounded to zero
0^s value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{p} preliminary
^{r} revised
x suppressed to meet the confidentiality requirements of the *Statistics Act*
^{E} use with caution
F too unreliable to be published
* significantly different from reference category ($p < 0.05$)

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, June 2018
Vol. 44, No. 1, pp. 101-115
Statistics Canada, Catalogue No. 12-001-X

Robust Bayesian small area estimation

Malay Ghosh, Jiyoun Myung and Fernando A.S. Moura[1]

Abstract

Small area models handling area level data typically assume normality of random effects. This assumption does not always work. The present paper introduces a new small area model with *t* random effects. Along with this, this paper also considers joint modeling of small area means and variances. The present approach is shown to perform better than other methods.

Key Words: Random effects model; Student's *t*-distribution; Non-subjective priors; MCMC; Gibbs sampling; Metropolis-Hastings algorithm.

1 Introduction

The classic paper of Fay and Herriot (1979) has become a cornerstone of research in small area estimation for nearly four decades. The Fay-Herriot model is a random effects model with a normality assumption for both the random effects and the errors. Moreover, the error variances are assumed to be known. The latter is almost imperative due to an identifiability issue. With availability of only the area level direct small area estimates plus nonavailability of microdata, any effective modeling of the error variances is near impossible.

Some valiant remedial attempts were made by W.R. Bell and his colleagues at the US Census Bureau (Bell and Huang, 2006; Bell, 2008) for handling some census data, but questions remain regarding the universal application of their approach. Additionally, nonavailability of microdata for secondary survey users is primarily due to confidentiality reasons, especially from the Federal Agencies. If microdata becomes available, unit level models are more appropriate than area level models. A classic example is the well-cited article of Battese, Harter and Fuller (1988). However, area level models are widely used due to their simplicity of implementation in a complex survey setting when compared to unit level models.

As the field developed and more data started getting analyzed, researchers found the inappropriateness of the assumption of normality as well as that of known error variances. As mentioned in the previous paragraph, the latter is hard to rectify without any extra information. One of the first attempts in this regard is due to Lahiri and Rao (1995) who replaced the normality assumption of random effects by the finiteness of their eighth moments. Datta and Lahiri (1995) considered a general mixture of normal distributions for random effects that includes the $t-$ distribution. There are papers, dispensing fully with the normality, but maintaining linearity of the model, and using ANOVA estimators of the variances. One may refer to Butar and Lahiri (2002) and Jiang, Lahiri and Wan (2002) who calculated the corresponding uncertainty measures either via jackknife or bootstrap. Bell and Huang (2006) used $t-$ distributions for random effects or sampling errors to diminish the effects of outliers.

1. Malay Ghosh, Department of Statistics, University of Florida, 102 Griffin-Floyd Hall, P.O. Box 118545, Gainesville, Florida, U.S.A., 32611. E-mail: ghoshm@stat.ufl.edu; Jiyoun Myung, Department of Statistics and Applied Probability, University of California, Santa Barbara, 5513 South Hall, Santa Barbara, California, U.S.A., 93106. E-mail: myung@pstat.ucsb.edu; Fernando A.S. Moura, Instituto de Matemática, Universidade Federal do Rio de Janeiro, Caixa Postal 68530, CEP: 21941-909, RJ, Brazil. E-mail: fmoura@im.ufrj.br.

The objective of the present article is to address these two important issues in the context of small area estimation. First, we consider small area modeling of both the population means and population variances. This is possible due to the availability of additional data purported to estimate the error variances. Second, in order to induce some robustness of our procedure, we consider $t-$ priors for the random effects.

The data set considered in this paper came from a test demographic census carried out in one municipality in Brazil consisting of 140 enumeration districts, hereafter referred to as small areas. The response variable was the average income of the heads of households for each small area, and the goal was to make predictions for the 140 population means of the heads of household's income. The auxiliary variables were the respective small area population means of the educational attainment of the heads of households, and the respective population means of the number of rooms in the households for each small area. Only area level data was provided to us.

We propose a full non-subjective Bayesian analysis for the general small area problem, where we model both the population means and variances. The initial idea was to use Jeffreys' general rule prior, treating all the parameters including the degrees of freedom of the Student's $t-$ distribution as unknown. However, the resultant prior yielded an improper posterior, which led to a modified Jeffreys' prior resulting in a proper posterior.

The outline of the remaining sections is as follows. Section 2 introduces the model, the Fisher information matrix, Jeffreys' prior and its modification. The impropriety of the posterior under the former, and its propriety under the latter are also included in this section. Section 3 contains a real data analysis as well as a simulation study. Some final remarks are made in Section 4.

The fact that error variances are really random has been recognized for a long time. The work of Otto and Bell (1995), Arora and Lahiri (1997), Wang and Fuller (2003), Rivest and Vandal (2003) and others have tried to account for this in different ways. Slud and Maiti (2006), Dass, Maiti, Ren and Sinha (2012) and Maiti, Ren and Sinha (2014) used an empirical Bayes approach towards this end by estimating the hyper-parameters. Full Bayesian analysis using hierarchical Bayesian methods with normality of area level effects has been considered in You and Chapman (2006) and Sugasawa, Tamae and Kubokawa (2017). We will demonstrate that $t-$ priors for random effects often perform better than the methods of the last two papers via data analysis and simulations.

The use of $t-$ priors for the errors in the standard normal regression models, but not in mixed effects models, was proposed in Lange, Little and Taylor (1989), Fernandez and Steel (1998), Vrontos, Dellaportas and Politis (2000), Jacquier, Polson and Rossi (2004), and Fonseca, Ferreira and Migon (2008) primarily for protection against outliers. However, there are situations where normality of errors is a reasonable assumption, mainly because of the central limit theorem. Also, there are model diagnostic techniques to check this. The normality assumption of random effects, however, does not always work well. For the Brazilian data that we have on hand, joint modeling of both sample means and variances along with $t-$ priors for random effects yields better performance than some of the other area level models. As suggested by referee, we used the data to compute the residuals fitting a regression with the true area means and covariates to investigate the distribution of the random effects for this application. See Section 3.

2 The model

A typical area level model is given by $y_i = \mathbf{x}_i^T \boldsymbol{\beta} + u_i + e_i$, $(i = 1, \ldots, m)$, where m denotes the number of small areas, $\mathbf{x}_1, \ldots, \mathbf{x}_m$ are $p\,(< m)$ dimensional covariates, and $\boldsymbol{\beta}\,(p \times 1)$ is the vector of regression coefficients. The random effects u_i and the sampling errors e_i are assumed to be independently distributed with the $u_i \overset{\text{iid}}{\sim} N(0, \sigma_u^2)$ and the $e_i \overset{\text{ind}}{\sim} N(0, v_i)$. That is, the classic area level model is

$$y_i \,|\, \theta_i \overset{\text{ind}}{\sim} N(\theta_i, v_i),$$

$$\theta_i \,|\, \boldsymbol{\beta}, \sigma_u^2 \overset{\text{ind}}{\sim} N(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma_u^2), \quad i = 1, \ldots, m.$$

The v_i are assumed to be known in order to avoid non-identifiability. The assumption of known v_i almost becomes mandatory for secondary users of survey data who do not have access to any micro data for modeling the v_i. However, in reality they are random, based on sampled data. In situations when one has additional data to model the v_i, the data can be used efficiently for estimating the v_i. Moreover, in such situations, it is possible to have shrinkage estimators of the small area means θ_i as well as of the variances v_i.

We address small area estimation problems where we have additional data to model the v_i. Also, for robustification, we assume $t-$ distribution of the random effects instead of the normal distribution. We state our model as follows,

$$y_i \,|\, \theta_i, v_i \overset{\text{ind}}{\sim} N(\theta_i, v_i),\; s_i^2 \,|\, v_i \overset{\text{ind}}{\sim} G\left(\frac{n_i - 1}{2}, \frac{1}{2v_i}\right)$$

$$\theta_i \,|\, \boldsymbol{\beta}, \sigma_\delta^2, \nu \overset{\text{ind.}}{\sim} t_\nu(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma_\delta), \quad i = 1, \ldots, m, \tag{2.1}$$

where n_i is the sample size in the i^{th} area, $t_\nu(\mu, \sigma)$ denotes the Student's $t-$ distribution with location μ, scale σ and degrees of freedom ν, and $G(c, d)$ denotes the gamma distribution with the kernel density $x^{c-1}\exp(-dx)$ for $x > 0$.

For a full Bayesian analysis, our objective is to find the posterior distribution of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)^T$, given $\mathbf{y} = (y_1, \ldots, y_m)^T$ and $\mathbf{s}^2 = (s_1^2, \ldots, s_m^2)^T$. To this end, we first need to find the prior distributions for all the hyper-parameters, $\boldsymbol{\beta}$, $\mathbf{v} = (v_1, \ldots, v_m)^T$, σ_δ^2, and ν. The usual first try is Jeffreys' prior which is proportional to the positive square root of the determinant of the Fisher information matrix. The Fisher information matrix in our case is

$$\mathbf{I}(\boldsymbol{\beta}, \mathbf{v}, \sigma_\delta^2, \nu) = \begin{bmatrix} \frac{(\nu+1)}{\sigma_\delta^2(\nu+3)} \mathbf{X}^T \mathbf{X} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{m\nu}{2(\sigma_\delta^2)^2(\nu+3)} & \frac{-m}{\sigma_\delta^2(\nu+1)(\nu+3)} \\ \mathbf{0} & \mathbf{0} & \frac{-m}{\sigma_\delta^2(\nu+1)(\nu+3)} & mg(\nu) \end{bmatrix},$$

where $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_m)^T, \mathbf{D} = \text{Diag}\left(\frac{n_1}{2v_1^2}, \ldots, \frac{n_m}{2v_m^2}\right),$ and $g(\nu) = \{\Psi'(\frac{\nu}{2}) - \Psi'(\frac{\nu+1}{2})\}/4 - (\nu+5)/\{2\nu(\nu+1)(\nu+3)\},$ with $\Psi(z) = \Gamma'(z)/\Gamma(z)$ and $\Psi'(z) = d\Psi(z)/dz$ which are the digamma and the trigamma functions. Thus, Jeffreys' prior is

$$\boldsymbol{\pi}_J(\boldsymbol{\beta}, \mathbf{v}, \sigma_\delta^2, \nu) \propto (\sigma_\delta^2)^{-\frac{p}{2}-1} \left|\mathbf{X}^T\mathbf{X}\right|^{\frac{1}{2}} \left|\mathbf{D}\right|^{\frac{1}{2}} \left(\frac{\nu+1}{\nu+3}\right)^{\frac{p}{2}} \left[\frac{\nu g(\nu)}{2(\nu+3)} - \frac{1}{(\nu+1)^2(\nu+3)^2}\right]^{\frac{1}{2}}. \tag{2.2}$$

However, Jeffreys' prior leads to an improper posterior due to the factor of $(\sigma_\delta^2)^{-\frac{p}{2}-1}$ in (2.2).

Theorem 1. Jeffreys' prior (2.2) leads an improper posterior.

Proof. Let $\boldsymbol{\pi}_J \equiv \boldsymbol{\pi}_J(\boldsymbol{\beta}, \mathbf{v}, \sigma_\delta^2, \nu \mid \mathbf{y}, \mathbf{s}^2)$ be the posterior density with Jeffreys' prior (2.2). Considering the terms that contain σ_δ^2 in $\boldsymbol{\pi}_J$ and taking the transformation $w_i = (\theta_i - \mathbf{x}_i^T\boldsymbol{\beta})/\sigma_\delta$, i.e., $\theta_i = \mathbf{x}_i^T\boldsymbol{\beta} + \sigma_\delta w_i$, we have

$$\begin{aligned}
&\int_0^\infty (\sigma_\delta^2)^{-\frac{p}{2}-1} \exp\left[-\frac{1}{2}\sum_{i=1}^m \frac{1}{v_i}(y_i - \mathbf{x}_i^T\boldsymbol{\beta} - \sigma_\delta w_i)^2\right] d\sigma_\delta^2 \\
&\qquad \geq \int_0^\infty (\sigma_\delta^2)^{-\frac{p}{2}-1} \exp\left[-\sum_{i=1}^m \left\{\frac{(y_i - \mathbf{x}_i^T\boldsymbol{\beta})^2}{v_i} + \frac{\sigma_\delta^2 w_i^2}{v_i}\right\}\right] d\sigma_\delta^2 \\
&\qquad \geq \exp\left\{-\sum_{i=1}^m \frac{(y_i - \mathbf{x}_i^T\boldsymbol{\beta})^2}{v_i}\right\} \int_0^k (\sigma_\delta^2)^{-\frac{p}{2}-1} \exp\left(-\sum_{i=1}^m \frac{\sigma_\delta^2 w_i^2}{v_i}\right) d\sigma_\delta^2 \text{ for a constant } k > 0, \\
&\qquad \geq \exp\left[-\sum_{i=1}^m \left\{\frac{(y_i - \mathbf{x}_i^T\boldsymbol{\beta})^2}{v_i} + \frac{k w_i^2}{v_i}\right\}\right] \int_0^k (\sigma_\delta^2)^{-\frac{p}{2}-1} d\sigma_\delta^2 = \infty.
\end{aligned}$$

Therefore, Jeffreys' prior leads to an improper posterior.

However, once the component $(\sigma_\delta^2)^{-\frac{p}{2}-1}$ in (2.2) is replaced by $(\sigma_\delta^2)^{-\frac{p}{2}-1}\exp(-a/2\sigma_\delta^2)$ for some $a > 0$, this modified version of Jeffreys' prior will lead a proper posterior under the condition, $\min(n_1, \ldots, n_m) > p$. Therefore, we suggest a modified Jeffreys' prior for our model as follows:

$$\begin{aligned}
\boldsymbol{\pi}_{\text{MJ}}(\boldsymbol{\beta}, \mathbf{v}, \sigma_\delta^2, \nu) &\propto (\sigma_\delta^2)^{-\frac{p}{2}-1} \exp\left(-\frac{a}{2\sigma_\delta^2}\right)\left(\prod_{i=1}^m \frac{1}{v_i}\right)\left(\frac{\nu+1}{\nu+3}\right)^{\frac{p}{2}} \\
&\quad \times \left[\frac{\nu g(\nu)}{2(\nu+3)} - \frac{1}{(\nu+1)^2(\nu+3)^2}\right]^{\frac{1}{2}} \text{ where } a > 0.
\end{aligned} \tag{2.3}$$

By combining the likelihood of (2.1) and modified Jeffreys' prior (2.3), the full posterior of parameters given the data is

$$\boldsymbol{\pi}_{\text{MJ}}\left(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{v}, \sigma_\delta^2, \nu \,|\, \mathbf{y}, \mathbf{s}^2\right) \propto \left(\sigma_\delta^2\right)^{-\frac{p+m}{2}-1} \exp\left(-\frac{a}{2\sigma_\delta^2}\right)\left(\prod_{i=1}^m v_i^{-\frac{n_i}{2}-1}\right)\exp\left[-\frac{1}{2}\sum_{i=1}^m \frac{1}{v_i}\left(y_i - \theta_i\right)^2\right]$$

$$\times \exp\left(-\frac{1}{2}\sum_{i=1}^m \frac{s_i^2}{v_i}\right)\left[\prod_{i=1}^m \left\{1 + \frac{\left(\theta_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2}{\nu\sigma_\delta^2}\right\}^{-\frac{\nu+1}{2}}\right]\left[\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu}}\right]^m$$

$$\times \left(\frac{\nu+1}{\nu+3}\right)^{\frac{p}{2}}\left[\frac{\nu g(\nu)}{2(\nu+3)} - \frac{1}{(\nu+1)^2(\nu+3)^2}\right]^{\frac{1}{2}}. \qquad (2.4)$$

Theorem 2. Under the model (2.1), the posterior $\boldsymbol{\pi}_{\text{MJ}}\left(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{v}, \sigma_\delta^2, \nu \,|\, \mathbf{y}, \mathbf{s}^2\right)$ in (2.4) is proper, provided $\min\left(n_1, \ldots, n_m\right) > p$.

Proof. See Appendix A.

Theorem 2 shows that modified Jeffreys' prior (2.3) leads to a proper posterior (2.4). The key idea is that we need a prior for σ_δ^2 such that $\int_0^\infty \boldsymbol{\pi}\left(\sigma_\delta^2\right)\left(\sigma_\delta^2\right)^{-\frac{p}{2}-1} d\sigma_\delta^2 < \infty$.

Remark 1. $\boldsymbol{\pi}_{\text{MJ}}\left(\boldsymbol{\beta}, \mathbf{v}, \sigma_\delta^2, \nu\right)$ can be factored into four independent priors for each parameter.

$$\boldsymbol{\pi}_{\text{MJ}}\left(\boldsymbol{\beta}, \mathbf{v}, \sigma_\delta^2, \nu\right) \propto \boldsymbol{\pi}(\boldsymbol{\beta})\,\boldsymbol{\pi}(\mathbf{v})\,\boldsymbol{\pi}\left(\sigma_\delta^2\right)\boldsymbol{\pi}(\nu)$$

where

$$\boldsymbol{\pi}(\boldsymbol{\beta}) \propto 1,\ \boldsymbol{\pi}(v_i) \propto \frac{1}{v_i} \quad \text{for} \quad i = 1, \ldots, m,$$

$$\boldsymbol{\pi}\left(\sigma_\delta^2\right) \sim \text{IG}\left(\frac{p}{2}, \frac{a}{2}\right)$$

and

$$\boldsymbol{\pi}(\nu) \propto \left(\frac{\nu+1}{\nu+3}\right)^{\frac{p}{2}}\left[\frac{\nu g(\nu)}{2(\nu+3)} - \frac{1}{(\nu+1)^2(\nu+3)^2}\right]^{\frac{1}{2}}.$$

Here $\text{IG}(c, d)$ denotes the inverse gamma distribution with the kernel density $x^{-c-1}\exp(-d/x)$ for $x > 0$.

The full conditional distributions to implement the Markov chain Monte Carlo (MCMC) are given in details in Appendix B. To generate samples, we use Gibbs sampling with Metropolis-Hastings algorithm where the conditional distribution of a parameter is known only up to a multiplicative constant. We provide details on how to apply a result of Chib and Greenberg (1995) for the Metropolis-Hastings algorithm to generate samples.

3 Application

3.1 Real data analysis

The data set is selected by a 10% random sampling of households in each area from a test demographic census completed in one municipality in Brazil. The municipality consists of 38,740 households in 140 small areas in total, and the number of households per area in the population ranges from 57 to 588. Thus the area sample sizes in the data set range from 6 to 59. We are interested in estimating the 140 population means of the head of household's income. The response variable y_i denotes the average income of the heads of households in i^{th} area.

This data set includes two centered auxiliary covariates which are the respective small area population means of the educational attainment of the head of households (ordinal scale of $0-5$) and the average number of rooms in households $(1-11+)$. Lastly, the data set contains the respective sampling variances which are calculated in the usual way. Since only area level data were provided to us and the true area means are known, we can compare the 140 small area predictions with the true area means respectively. The analysis suggests that our model performs better than other models where random effects are based on the normal distribution. For comparison, we use three other models.

The first one is the Fay-Herriot model, referred to as FH, with known sampling variances.

$$y_i \,|\, \theta_i, v_i \overset{\text{ind.}}{\sim} N(\theta_i, v_i), \quad \theta_i \,|\, \boldsymbol{\beta}, \sigma_\delta^2 \overset{\text{ind.}}{\sim} N(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma_\delta^2)$$
$$\pi(\boldsymbol{\beta}) \propto 1, \quad \pi(\sigma_\delta^2) \sim \text{IG}(a_0, b_0),$$

where a_0 and b_0 are chosen to be 0.0001 (a small constant) to reflect the vague knowledge of σ_δ^2.

The second model suggested by You and Chapman (2006), referred to as YC, is a hierarchical Bayesian model given by

$$y_i \,|\, \theta_i, v_i \overset{\text{ind.}}{\sim} N(\theta_i, v_i), \quad \theta_i \,|\, \boldsymbol{\beta}, \sigma_\delta^2 \overset{\text{ind.}}{\sim} N(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma_\delta^2)$$
$$s_i^2 \,|\, v_i \overset{\text{ind}}{\sim} G\left(\frac{n_i - 1}{2}, \frac{1}{2v_i}\right)$$
$$\pi(v_i) \sim \text{IG}(a_i, b_i), \quad \pi(\boldsymbol{\beta}) \propto 1, \quad \pi(\sigma_\delta^2) \sim \text{IG}(a_0, b_0),$$

where a_i, b_i, a_0 and b_0 are also chosen to be 0.0001.

The third model is a Bayesian multi-stage small area model proposed by Sugasawa et al. (2017), referred to as STK. The STK model produces shrinkage estimation of both means and variances.

$$y_i \,|\, \theta_i, v_i \overset{\text{ind.}}{\sim} N(\theta_i, v_i), \quad \theta_i \,|\, \boldsymbol{\beta}, \sigma_\delta^2 \overset{\text{ind.}}{\sim} N(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma_\delta^2)$$
$$s_i^2 \,|\, v_i \overset{\text{ind}}{\sim} G\left(\frac{n_i - 1}{2}, \frac{1}{2v_i}\right), \quad v_i \,|\, \gamma \sim \text{IG}(a_i, b_i \gamma),$$
$$\pi(\boldsymbol{\beta}, \sigma_\delta^2, \gamma) = 1,$$

where $a_i = 2$ and $b_i = 1/n_i$ as suggested by authors for a reasonable choice.

We compare the small area means predicted by FH, YC, STK, and our model, hereafter referred to as RTS model. For the MCMC implementation, we generate a chain with a burn-in length of 50,000 and the sampling size of $G = 50{,}000$. The estimates of the θ_i are given by

$$\hat{\theta}_i = \frac{1}{G} \sum_{g=1}^{G} \left(\gamma_i^{(g)} y_i + \left(1 - \gamma_i^{(g)}\right) \mathbf{x}_i^T \boldsymbol{\beta}^{(g)} \right)$$

where

$$\gamma_i^{(g)} = \frac{v_i^{-1(g)}}{v_i^{-1(g)} + \sigma_\delta^{-2(g)} \eta_i^{(g)}}.$$

The comparison criteria are the average squared deviation (ASD), average absolute bias (AAB), average squared relative bias (ASRB), and average relative bias (ARB). They are defined as follows;

$$\text{ASD} = \frac{1}{m} \sum_{i=1}^{m} \left(\hat{\theta}_i - \theta_i \right)^2, \quad \text{AAB} = \frac{1}{m} \sum_{i=1}^{m} \left| \hat{\theta}_i - \theta_i \right|, \quad \text{ASRB} = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{\hat{\theta}_i - \theta_i}{\theta_i} \right)^2,$$

and

$$\text{ARB} = \frac{1}{m} \sum_{i=1}^{m} \left| \frac{\hat{\theta}_i - \theta_i}{\theta_i} \right|,$$

where $\hat{\theta}_i$ and θ_i are the estimated and true values respectively in the i^{th} area. Table 3.1 compares the four models. Recall the prior distribution of σ_δ^2, which is $\boldsymbol{\pi}\left(\sigma_\delta^2\right) \sim \text{IG}\left(\frac{p}{2}, \frac{a}{2}\right)$. With the shape parameter, $\frac{p}{2} = 1$, we consider several values of a. If we choose a to be close to p, RTS model fits better than the rest under all four criteria. When we choose $a = 1$, RTS model performs best. YC model performs worse than the other three models. If we choose very small a, such as 0.01 or 0.001, then RTS model performs the worst.

Table 3.1
Comparison between RTS model, FH model, YC model, and STK model

Model	ASD	AAB	ASRB	ARB
RTS model (a = 0.0001)	57.297	6.152	0.395	0.589
RTS model (a = 0.01)	16.546	2.741	0.090	0.244
RTS model (a = 0.5)	3.244	1.249	0.020	0.118
RTS model (a = 0.2)	4.185	1.439	0.025	0.133
RTS model (a = 1)	2.745	1.164	0.019	0.113
RTS model (a = 2)	3.080	1.231	0.020	0.117
RTS model (a = 3)	3.079	1.229	0.020	0.117
RTS model (a = 5)	2.994	1.213	0.019	0.116
RTS model (a = 10)	3.377	1.278	0.020	0.119
RTS model (a = 50)	2.905	1.180	0.018	0.112
RTS model (a = 100)	2.799	1.154	0.018	0.109
FH model	4.484	1.448	0.026	0.133
YC model	4.983	1.543	0.029	0.141
STK model	3.199	1.257	0.021	0.121

Additionally, as suggested by a referee, we compute the residuals fitting a regression with the true area means and covariates to see the distribution of the random effects for this real data. Figure 3.1 shows that the distribution departs from the normal distribution.

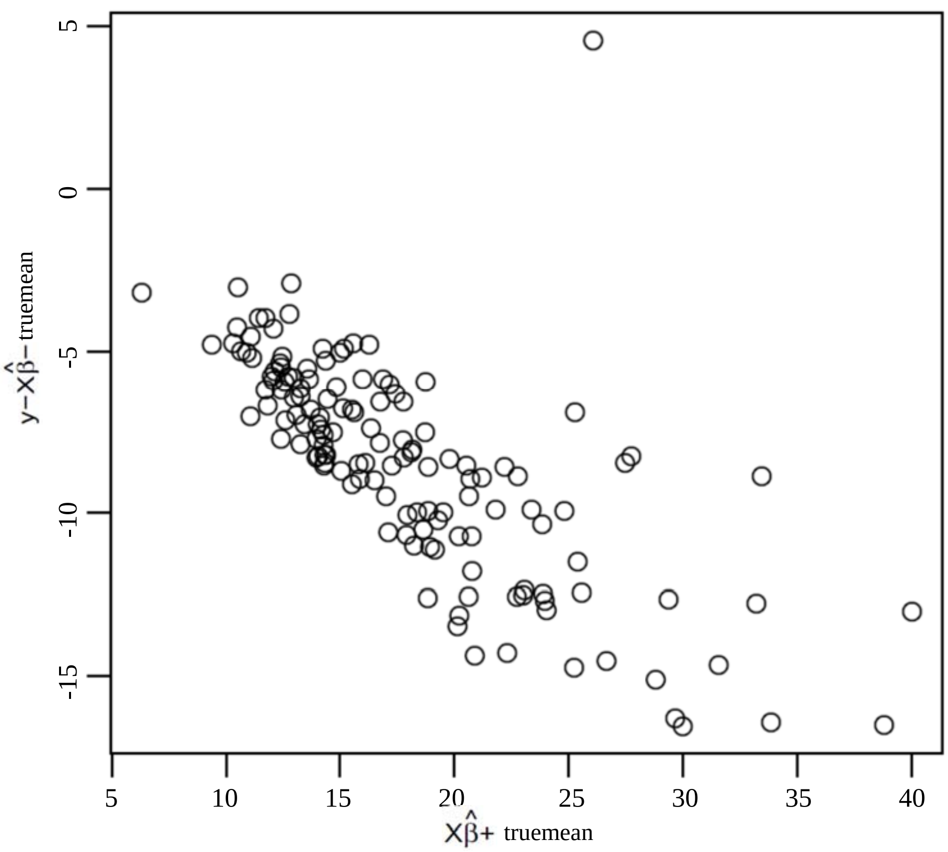

Figure 3.1 Residuals fitting a regression with the true mean and covariates.

3.2 Simulation study

In this section, we set up a simulation close to Maiti et al. (2014) (or Sugasawa et al. (2017)) to compare the accuracy of our estimators to other estimators, specifically those from You and Chapman (2006) and Sugasawa et al. (2017). We generate observations for each small area from the model

$$y_{ij} = \beta_0 + \beta_1 x_i + u_i + e_{ij}, \quad j = 1, \ldots, n_i, \, i = 1, \ldots, m,$$

where $u_i \sim t_\nu (0, \sigma_\delta)$ and $e_{ij} \sim N(0, n_i v_i)$. Then the random effects model for the small area mean is

$$y_i = \beta_0 + \beta_1 x_i + u_i + e_i, \quad i = 1, \ldots, m,$$

where $y_i = \bar{y}_i = n_i^{-1} \sum_{j=1}^{n_i} y_{ij}$ and $e_i = n_i^{-1} \sum_{j=1}^{n_i} e_{ij}$. Hence, $y_i \,|\, \theta_i \sim N(\theta_i, v_i)$ where $\theta_i = \beta_0 + \beta_1 x_i + u_i$; $\theta_i \sim t_\nu (\beta_0 + \beta_1 x_i, \sigma_\delta)$, and $e_i \sim N(0, v_i)$. The interest parameter is the mean θ_i, for the i^{th} small area. Also, the direct estimator of v_i is

$$\frac{1}{n_i\,(n_i - 1)} \sum_{j=1}^{n_i} (y_{ij} - \overline{y}_i)^2.$$

We set $m = 30$ and $n_i = 7$ for all areas, and generate covariates x_i from the uniform distribution on (2, 8). The true parameter values are set as $\beta_0 = 0.5$, $\beta_1 = 0.8$, $\sigma_\delta = 1$, $\nu = 3$ and $v_i \sim \text{IG}(10, 5\exp(0.3x_i))$. Also, we chose $a = 3$ for all simulations.

For the MCMC implementation, we generated 5,000 posterior samples after discarding the first 1,000 for $R = 2{,}000$ simulation runs. Table 3.2 provides comparison among the four models. The comparison criteria are ASD, AAB, and BIAS, the latter being defined as

$$\text{BIAS} = \frac{1}{mR} \sum_{i=1}^{m} \left| \sum_{r=1}^{R} \left(\hat{\theta}_i^{(r)} - \theta_i^{(r)} \right) \right|.$$

Table 3.2
Simulation result for t random effects with $\nu = 3$

Model	Mean			Variance		
	ASD	AAB	BIAS	ASD	AAB	BIAS
RTS	1.393	0.895	0.021	5.048	1.608	1.324
STK	1.821	0.933	0.025	5.042	1.514	1.367
FH	1.540	0.942	0.022			
YC	2.165	0.974	0.030	5.970	1.803	1.689

RTS model performs better than others under ASD, AAB, and BIAS criteria for the mean. While RTS shows small improvements over other models for AAB and BIAS criteria, it shows approximate 23.5%, 10%, and 35.7% improvements over STK, FH and YC models respectively for ASD criteria. For the variance, RTS and STK models perform better than YC model.

The following two tables provide the simulation results when one sets the degrees of freedom as $\nu = 2$ and $\nu = 4$.

Table 3.3
Simulation result for t random effects with $\nu = 2$

Model	Mean			Variance		
	ASD	AAB	BIAS	ASD	AAB	BIAS
RTS	1.617	0.949	0.020	6.569	1.610	1.070
STK	7.566	1.107	0.038	11.144	1.441	0.996
FH	1.921	1.035	0.022			
YC	9.063	1.187	0.038	7.072	1.685	1.340

Table 3.4
Simulation result for t random effects with $\nu = 4$

Model	Mean			Variance		
	ASD	AAB	BIAS	ASD	AAB	BIAS
RTS	1.265	0.862	0.019	4.876	1.619	1.428
STK	1.322	0.874	0.019	5.077	1.577	1.489
FH	1.350	0.894	0.020			
YC	1.509	0.905	0.022	6.201	1.869	1.802

With $\nu = 2$, RTS model performs better than others under ASD, AAB, and BIAS criteria for the mean. In this simulation, the ASD values for STK and YC models are very large compared with RTS and FH models. RTS model shows improvements of about 78.6% over STK model, 82.2% over YC model, and 15.8% over FH model. For AAB and BIAS, the values of RTS model are smaller than those of other models. When considering the variance, ASD for RTS model gives smallest value.

With $\nu = 4$, RTS model also shows better performance over others. Especially, ASD and BIAS values indicate that RTS model improves results when compared with STK and YC model.

The next two tables consider the situation where one assumes normality of the random effects. Here RTS model performs slightly worse than the other models.

Table 3.5
Simulation result for normal random effects with $N(0, 5^2)$

Model	Mean			Variance		
	ASD	AAB	BIAS	ASD	AAB	BIAS
RTS	2.896	1.305	0.038	6.036	1.512	0.514
STK	2.560	1.229	0.051	1.851	0.961	0.114
FH	2.597	1.240	0.036			
YC	2.735	1.259	0.048	3.674	1.305	0.463

Table 3.6
Simulation result for normal random effects with $N(0, 10^2)$

Model	Mean			Variance		
	ASD	AAB	BIAS	ASD	AAB	BIAS
RTS	3.007	1.316	0.032	10.117	1.895	1.202
STK	2.784	1.272	0.031	2.221	1.038	0.155
FH	2.765	1.272	0.048			
YC	2.873	1.285	0.033	9.166	1.798	1.129

4 Final remarks

The paper considers small area models for handling area level data. The new feature of this article is modeling both small area means and variances along with the use of $t-$distribution of random effects. It is shown via both data analysis and simulation that the proposed method performs mostly better than the models of You and Chapman (2006) and Sugasawa et al. (2017) in most situations.

Acknowledgements

Ghosh's research was partially supported in part by NSF Grant SES-1327359.

Appendix A

Proof

Theorem 2. Under the model (2.1) with modified Jeffreys' prior (2.3), the posterior distribution (2.4) is proper, provided $\min(n_1, \ldots, n_m) > p$.

Proof. Recall the posterior distribution (2.4),

$$\boldsymbol{\pi}_{\mathrm{MJ}}\left(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{v}, \sigma_\delta^2, \nu \,|\, \mathbf{y}, \mathbf{s}^2\right) \propto \left(\sigma_\delta^2\right)^{-\frac{p+m}{2}-1} \exp\left(-\frac{a}{2\sigma_\delta^2}\right)\left(\prod_{i=1}^m v_i^{-\frac{n_i}{2}-1}\right)\exp\left[-\frac{1}{2}\sum_{i=1}^m \frac{1}{v_i}\left(y_i-\theta_i\right)^2\right]$$

$$\times \exp\left(-\frac{1}{2}\sum_{i=1}^m \frac{s_i^2}{v_i}\right)\left[\prod_{i=1}^m \left\{1+\frac{\left(\theta_i-\mathbf{x}_i^T\boldsymbol{\beta}\right)^2}{\nu\sigma_\delta^2}\right\}^{-\frac{\nu+1}{2}}\right]\left[\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu}}\right]^m$$

$$\times\left(\frac{\nu+1}{\nu+3}\right)^{\frac{p}{2}}\left[\frac{\nu g(\nu)}{2(\nu+3)}-\frac{1}{(\nu+1)^2(\nu+3)^2}\right]^{\frac{1}{2}}$$

where $g(\nu)=\left\{\Psi'\left(\frac{\nu}{2}\right)-\Psi'\left(\frac{\nu+1}{2}\right)\right\}/4-(\nu+5)/\{2\nu(\nu+1)(\nu+3)\}$.

First of all, since $\log\left[\Gamma\left\{\frac{(\nu+1)}{2}\right\}/\Gamma\left(\frac{\nu}{2}\right)\right] \doteq \{-\log(2e)+\nu\log(\nu+1)-(\nu-1)\log\nu\}/2$ by Stirling's approximation, we have

$$\frac{1}{2}\left[\Psi\left(\frac{\nu+1}{2}\right)-\Psi\left(\frac{\nu}{2}\right)\right] \doteq \frac{d}{d\nu}\log\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}=\frac{1}{2}\left[\log(\nu+1)-\log\nu\right]+\frac{1}{2\nu(\nu+1)}$$

and

$$\frac{1}{4}\left[\Psi'\left(\frac{\nu+1}{2}\right)-\Psi'\left(\frac{\nu}{2}\right)\right] \doteq \frac{1}{2}\left(\frac{1}{\nu+1}-\frac{1}{\nu}\right)-\frac{1}{2\nu^2}+\frac{1}{2(\nu+1)^2}.$$

This leads to

$$g(\nu) \doteq \frac{5\nu+3}{2\nu^2(\nu+1)^2(\nu+3)}$$

and

$$\left[\frac{\nu g(\nu)}{2(\nu+3)}-\frac{1}{(\nu+1)^2(\nu+3)^2}\right] \doteq \frac{1}{4\nu(\nu+1)^2(\nu+3)}.$$

Hence this approximation simplifies the last term in (2.4). The corresponding posterior is

$$\boldsymbol{\pi}_{\mathrm{MJ}}\left(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{v}, \sigma_\delta^2, \nu \,|\, \mathbf{y}, \mathbf{s}^2\right) \propto \left(\sigma_\delta^2\right)^{-\frac{p+m}{2}-1} \exp\left(-\frac{a}{2\sigma_\delta^2}\right)\left(\prod_{i=1}^m v_i^{-\frac{n_i}{2}-1}\right)\exp\left[-\frac{1}{2}\sum_{i=1}^m \frac{1}{v_i}\left(y_i-\theta_i\right)^2\right]$$

$$\times \exp\left(-\frac{1}{2}\sum_{i=1}^m \frac{s_i^2}{v_i}\right)\left[\prod_{i=1}^m \left\{1+\frac{\left(\theta_i-\mathbf{x}_i^T\boldsymbol{\beta}\right)^2}{\nu\sigma_\delta^2}\right\}^{-\frac{\nu+1}{2}}\right]\left[\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu}}\right]^m$$

$$\times\left(\frac{\nu+1}{\nu+3}\right)^{\frac{p}{2}}\left[\frac{1}{\nu(\nu+1)^2(\nu+3)}\right]^{\frac{1}{2}}. \tag{A.1}$$

First, integrating out with respect to $\boldsymbol{\beta}$. By letting $w_i=\left(\theta_i-\mathbf{x}_i^T\boldsymbol{\beta}\right)/\sigma_\delta$, i.e., $\theta_i=\mathbf{x}_i^T\boldsymbol{\beta}+\sigma_\delta w_i$, we have

$$\boldsymbol{\pi}_{\mathrm{MJ}}(\mathbf{w}, \mathbf{v}, \sigma_\delta^2, \nu \mid \mathbf{y}, \mathbf{s}^2) \propto (\sigma_\delta^2)^{-\frac{p}{2}-1} \exp\left(-\frac{a}{2\sigma_\delta^2}\right)\left(\prod_{i=1}^m v_i^{-\frac{n_i}{2}-1}\right)\exp\left(-\frac{1}{2}\sum_{i=1}^m \frac{s_i^2}{v_i}\right)$$

$$\times\left[\prod_{i=1}^m\left(1+\frac{w_i^2}{\nu}\right)^{-\frac{\nu+1}{2}}\right]\left[\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu}}\right]^m\left(\frac{\nu+1}{\nu+3}\right)^{\frac{p}{2}}\left[\frac{1}{\nu(\nu+1)^2(\nu+3)}\right]^{\frac{1}{2}}|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}|^{-\frac{1}{2}}$$

$$\times\exp\left[-\frac{1}{2}(\mathbf{Y}-\sigma_\delta\mathbf{w})^T\left\{\mathbf{V}^{-1}-\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\right\}(\mathbf{Y}-\sigma_\delta\mathbf{w})\right].$$

where $\mathbf{w} = (w_1, \ldots, w_m)^T$. Note that $\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}$ is non-negative definite. Using $K(> 0)$ as a generic constant which depends only on the data $(\mathbf{y}, \mathbf{s}^2)$, we integrate out with respect to $\mathbf{w}$ and σ_δ^2 respectively in order, we have

$$\boldsymbol{\pi}_{\mathrm{MJ}}(\mathbf{v}, \sigma_\delta^2, \nu \mid \mathbf{y}, \mathbf{s}^2) \le K(\sigma_\delta^2)^{-\frac{p}{2}-1}\exp\left(-\frac{a}{2\sigma_\delta^2}\right)\left(\prod_{i=1}^m v_i^{-\frac{n_i}{2}-1}\right)$$

$$\times\exp\left(-\frac{1}{2}\sum_{i=1}^m\frac{s_i^2}{v_i}\right)\left[\frac{1}{\nu(\nu+1)^2(\nu+3)}\right]^{\frac{1}{2}}|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}|^{-\frac{1}{2}},$$

and then

$$\boldsymbol{\pi}_{\mathrm{MJ}}(\mathbf{v}, \nu \mid \mathbf{y}, \mathbf{s}^2) \le K\left(\prod_{i=1}^m v_i^{-\frac{n_i}{2}-1}\right)\exp\left(-\frac{1}{2}\sum_{i=1}^m\frac{s_i^2}{v_i}\right)\times\left[\frac{1}{\nu(\nu+1)^2(\nu+3)}\right]^{\frac{1}{2}}|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}|^{-\frac{1}{2}}.$$

Note that

$$\int_0^\infty \nu^{-\frac{1}{2}}(\nu+1)^{-1}(\nu+3)^{-\frac{1}{2}}d\nu = \int_0^c \nu^{-\frac{1}{2}}(\nu+1)^{-1}(\nu+3)^{-\frac{1}{2}}d\nu + \int_c^\infty \nu^{-\frac{1}{2}}(\nu+1)^{-1}(\nu+3)^{-\frac{1}{2}}d\nu$$

$$\le \int_0^c \nu^{-\frac{1}{2}}d\nu + \int_c^\infty \nu^{-2}d\nu = 2c^{\frac{1}{2}} + c^{-1} < \infty \text{ for any } c > 0.$$

After integrating out with respect to ν, we have

$$\boldsymbol{\pi}_{\mathrm{MJ}}(\mathbf{v} \mid \mathbf{y}, \mathbf{s}^2) \le K\left(\prod_{i=1}^m v_i^{-\frac{n_i}{2}-1}\right)\exp\left(-\frac{1}{2}\sum_{i=1}^m\frac{s_i^2}{v_i}\right)|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}|^{-\frac{1}{2}}.$$

Finally, since $|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}|^{-\frac{1}{2}} \le (v_{\max})^{\frac{p}{2}}|\mathbf{X}^T\mathbf{X}|^{-\frac{1}{2}}$, where $v_{\max} = \max(v_1, \ldots, v_m)$, if $\min(n_1, \ldots, n_m) > p$, modified Jeffrey's prior leads to a proper posterior.

Appendix B

Full conditional distributions

The full posterior of the parameters given the data is specified in (A.1). For the MCMC implementation, it is convenient to use the latent parameters η_i $(i = 1, \ldots, m)$ such that

$$\theta_i \mid \eta_i \overset{\text{ind}}{\sim} \mathrm{N}\left(\mathbf{x}_i^T \boldsymbol{\beta}, \eta_i^{-1}\sigma_\delta^2\right)$$
$$\eta_i \overset{\text{iid}}{\sim} \mathrm{G}\left(\frac{\nu}{2}, \frac{\nu}{2}\right).$$

Let $\boldsymbol{\Lambda} = \mathrm{Diag}(\eta_1, \ldots, \eta_m)$. The full conditional distributions are

I. $\left[\sigma_\delta^2 \mid \boldsymbol{\theta}, \boldsymbol{\Lambda}, \boldsymbol{\beta}, \mathbf{v}, \nu, \mathbf{y}, \mathbf{s}^2\right] \sim \mathrm{IG}\left[\frac{p+m}{2}, \frac{1}{2}\left\{a + \sum_{i=1}^m \eta_i \left(\theta_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2\right\}\right];$

II. $\left[v_i \mid \boldsymbol{\theta}, \boldsymbol{\Lambda}, \boldsymbol{\beta}, \sigma_\delta^2, \nu, \mathbf{y}, \mathbf{s}^2\right] \overset{\text{ind}}{\sim} \mathrm{IG}\left[\frac{n_i}{2}, \frac{1}{2}\left\{\left(y_i - \theta_i\right)^2 + s_i^2\right\}\right];$

III. $\left[\boldsymbol{\beta} \mid \boldsymbol{\theta}, \boldsymbol{\Lambda}, \mathbf{v}, \sigma_\delta^2, \nu, \mathbf{y}, \mathbf{s}^2\right] \sim \mathrm{N}\left[\left(\mathbf{X}^T \boldsymbol{\Lambda} \mathbf{X}\right)^{-1} \mathbf{X}^T \boldsymbol{\Lambda} \boldsymbol{\theta}, \sigma_\delta^2 \left(\mathbf{X}^T \boldsymbol{\Lambda} \mathbf{X}\right)^{-1}\right];$

IV. $\left[\eta_i \mid \boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{v}, \sigma_\delta^2, \nu, \mathbf{y}, \mathbf{s}^2\right] \overset{\text{ind}}{\sim} \mathrm{G}\left[\frac{\nu+1}{2}, \frac{1}{2}\left\{\nu + \frac{\left(\theta_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2}{\sigma_\delta^2}\right\}\right];$

V. $\left[\theta_i \mid \boldsymbol{\Lambda}, \boldsymbol{\beta}, \mathbf{v}, \sigma_\delta^2, \nu, \mathbf{y}, \mathbf{s}^2\right] \overset{\text{ind}}{\sim} \mathrm{N}\left[\left(v_i^{-1} + \sigma_\delta^{-2}\eta_i\right)^{-1}\left(v_i^{-1} y_i + \sigma_\delta^{-2}\eta_i \mathbf{x}_i^T \boldsymbol{\beta}\right), \left(v_i^{-1} + \sigma_\delta^{-2}\eta_i\right)^{-1}\right];$

and

$$\text{IV.} \quad \left[\nu \mid \boldsymbol{\theta}, \boldsymbol{\Lambda}, \boldsymbol{\beta}, \mathbf{v}, \sigma_\delta^2, \mathbf{y}, \mathbf{s}^2\right] \propto \nu^{\frac{m-1}{2}} \exp\left\{-\frac{\nu}{2}\sum_{i=1}^m \left(\eta_i - \log \eta_i - 1\right)\right\}$$
$$\times \left\{\frac{\nu^{\frac{m\nu - m}{2}} \exp\left(-\frac{m\nu}{2}\right)}{2^{\frac{m\nu}{2}} \Gamma^m\left(\frac{\nu}{2}\right)} \left(\frac{\nu+1}{\nu+3}\right)^{\frac{p}{2}} (\nu+1)^{-1} (\nu+3)^{-\frac{1}{2}}\right\}. \qquad \text{(B.1)}$$

All but (VI) requires the generation of samples from standard distributions. While we use the Gibbs sampling method for (I)-(V), we use the Metropolis-Hastings algorithm for generating samples from (VI) as given in Chib and Greenberg (1995).

How to apply the result of Chib and Greenberg (1995) to (IV) in (B.1).

If the target density $\pi(t)$ can be written as $\pi(t) \propto \psi(t) h(t)$, where $h(t)$ is a density that can be sampled and $\psi(t)$ is uniformly bounded, then one can set $h(t)$ as a candidate density to draw samples and use $\psi(t)$ in $\alpha(x, y) = \min\{\psi(y)/\psi(x), 1\}$ which is the probability of move.

Recall that the full conditional distribution of ν is

$$\boldsymbol{\pi}(\nu \mid \cdot) \propto \nu^{\frac{m-1}{2}} \exp\left\{-\frac{\nu}{2}\sum_{i=1}^m \left(\eta_i - \log \eta_i - 1\right)\right\} \times \left\{\frac{\nu^{\frac{m\nu - m}{2}} \exp\left(-\frac{m\nu}{2}\right)}{2^{\frac{m\nu}{2}} \Gamma^m\left(\frac{\nu}{2}\right)} \left(\frac{\nu+1}{\nu+3}\right)^{\frac{p}{2}} (\nu+1)^{-1} (\nu+3)^{-\frac{1}{2}}\right\}.$$

Since $\Gamma\left(\frac{\nu}{2}\right) \geq \sqrt{2\pi} \exp\left(-\frac{\nu}{2}\right)\left(\frac{\nu}{2}\right)^{\frac{\nu-1}{2}}$, the second $\{\cdot\}$ term above is bounded by $\left(\sqrt{2\pi}\right)^{-m} (\nu+1)^{\frac{p}{2}-1} (\nu+3)^{-\frac{p}{2}-\frac{1}{2}} \leq \left(\sqrt{2\pi}\right)^{-m} \big/ \sqrt{3}$. Hence, we can apply third method in the Section 5 of Chib and Greenberg (1995) with

$$h(\nu) \sim \mathrm{G}\left(\frac{m+1}{2}, \frac{1}{2}\sum_{i=1}^m \left(\eta_i - \log \eta_i - 1\right)\right)$$

and

$$\psi(\nu) = \frac{\nu^{\frac{m\nu - m}{2}} \exp\left(-\frac{m\nu}{2}\right)}{2^{\frac{m\nu}{2}} \Gamma^m\left(\frac{\nu}{2}\right)} \left(\frac{\nu+1}{\nu+3}\right)^{\frac{p}{2}} (\nu+1)^{-1} (\nu+3)^{-\frac{1}{2}}.$$

Here $h(v)$ is a candidate-generating density, and $\psi(v)$ is uniformly bounded.

References

Arora, V., and Lahiri, P. (1997). On the superiority of the Bayesian method over the BLUP in small area estimation problems. *Stastica Sinica*, 7, 1053-1063.

Battese, G.E., Harter, R.M. and Fuller, W.A. (1988). An error component model for prediction of mean crop areas using survey and satellite data. *Journal of the American Statistical Association*, 95, 28-36.

Bell, W.R. (2008). Examining sensitivity of small area inferences to uncertainty about sampling error variances. In *Proceedings of the Survey Research Methods Section*, American Statistical Association, 327-333.

Bell, W.R., and Huang, E.T. (2006). Using the $t-$distribution to deal with outliers in small area estimation. Proceedings: *Symposium 2006, Methodological Issues in Measuring Population Health*, Statistics Canada.

Butar, F., and Lahiri, P. (2002). On the measures of uncertainty of empirical Bayes small-area estimators. *Journal of Statistical Planning and Inference*, 112, 63-76.

Chib, S., and Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, 49, 327-335.

Dass, S.C., Maiti, T., Ren, H. and Sinha, S. (2012). Confidence interval estimation of small area parameters shrinking both means and variances. *Survey Methodology*, 38, 2, 173-187. Paper available at https://www150.statcan.gc.ca/n1/pub/12-001-x/2012002/article/11756-eng.pdf.

Datta, G.S., and Lahiri, P. (1995). Robust hierarchical Bayes estimation of small area characteristics in the presence of outliers. *Journal of Multivariate Analysis*, 54, 310-328.

Fay, R.E., and Herriot, R.A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74, 269-277.

Fernandez, C., and Steel, M.F.J. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93, 359-371.

Fonseca, T.C.O., Ferreira, M.A.R. and Migon, H.S. (2008). Objective Bayesian analysis for the student $-t$ regression model. *Biometrika*, 95, 325-333.

Jacquier, E., Polson, N.G. and Rossi, P.E. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics*, 122, 185-212.

Jiang, J., Lahiri, P. and Wan, S. (2002). A unified jackknife theory for empirical best prediction with M-estimation. *Annals of Statistics*, 30, 1782-1810.

Lahiri, P., and Rao, J.N.K. (1995). Robust estimation of mean square error of small area estimators. *Journal of the American Statistical Association*, 90, 758-766.

Lange, K.L., Little, R.J.A. and Taylor, J.M.G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, 84, 881-896.

Maiti, T., Ren, H. and Sinha, A. (2014). Prediction error of small area predictors shrinking both means and variances. *Scandinavian Journal of Statistics*, 41, 775-790.

Otto, M.C., and Bell, W.R. (1995). Sampling error modelling of poverty and income statistics for states. In *Proceedings of the Section on Government Statistics*, American Statistical Association, 160-165.

Rivest, L.-P., and Vandal, N. (2003). Mean squared error estimation for small areas when the small area variances are estimated. In *Proceedings of the International Conference on Recent Advances in Survey Sampling*.

Slud, E.V., and Maiti, T. (2006). Mean-squared error estimation in transformed Fay-Herriot models. *Journal of Royal Statistical Society, Series B*, 68, 239-257.

Sugasawa, S., Tamae, H. and Kubokawa, T. (2017). Bayesian estimators for small area models shrinking both means and variances. *Scandinavian Journal of Statistics*, 44, 150-167.

Vrontos, I.D., Dellaportas, P. and Politis, D.N. (2000). Full Bayesian inference for GARCH and EGARCH models. *Journal of Business and Economic Statistics*, 18, 187-198.

Wang, J., and Fuller, W. (2003). The mean squared error of small area predictors constructed with estimated error variances. *Journal of the American Statistical Association*, 98, 716-723.

You, Y., and Chapman, B. (2006). Small area estimation using area level models and estimated sampling variances. *Survey Methodology*, 32, 1, 97-103. Paper available at https://www150.statcan.gc.ca/n1/pub/12-001-x/2006001/article/9263-eng.pdf.