Journal of the Korean Data & Information Science Society
2016, **27**(5), 1423–1431

http://dx.doi.org/10.7465/jkdi.2016.27.5.1423
한국데이터정보과학회지

Simultaneous modeling of mean and variance in small area estimation†

Myungjin Kim[1] · Dal Ho Kim[2]

[1]Department of Statistics, Iowa State University
[2]Department of Statistics, Kyungpook National University
Received 10 July 2016, revised 28 July 2016, accepted 8 September 2016

Abstract

When the sample size in a certain domain is too small to produce adequate information, small area model with random effects is usually used. Also, if we do not consider an inherent pattern which data possess, it considerably affects inference. In this paper, we mainly focus on modeling to handle increased variation of the Current Population Survey (CPS) median income as the Internal Revenue Service (IRS) mean income increases. In a hierarchical Bayesian framework, most estimations are carried out through the Gibbs sampler while the grid method is used to generate parameters from non-standard form. Numerical study indicates that the performance of proposed model is better than that of CPS method in terms of four comparison measurements.

Keywords: Gibbs sampler, grid method, small area model, variance structure.

1. Introduction

The size of sample influences the amount of information. Sometimes, the domain-specific sample size may be too small to guarantee relevant information. In such a case, small area models which include small area random effects are usually considered. The available auxiliary information and the small area effects are connected to the small area parameters in additive way. In this way, the usual small area models produce shrinkage estimators that "borrow strength" from other areas. For detail, one can refer to Ghosh and Rao (1994), and Ghosh, Nangia and Kim (1996) and Rao (2003).

In this paper, we adjust the inference strategy to account for increased variation of the Current Population Survey (CPS) median income in data from the Annual Social and Economic Supplement (ASEC) for the period 1995-1999. To explain this point, we consider simultaneous small area modeling both mean and variance. Model fitting and parameter estimations are carried out in a hierarchical Bayesian framework. A hierarchical Bayesian model for median household income of four-person families has recently been considered in Bhadra, Ghosh and Kim (2012), Goo and Kim (2013) and Lee and Kim (2013). To overcome calculation problems, we perform the Gibbs sampler and the grid method to generate some parameters from non-standard posterior distributions. Also, we use comparison measurements to evaluate the performance of proposed model.

† This paper is based on part of the first author's master thesis.
[1] Ph.D student, Department of Statistics, Iowa State University, IA 50011, U.S.
[2] Corresponding author: Professor, Department of Statistics, Kyungpook National University, Daegu 41566, Korea. E-mail: dalkim@knu.ac.kr

2. Hierarchical Bayesian inference

A direct estimation is derived using only data from one source for the state and time period. There are several direct estimators such as the Horvitz-Thompson (H-T) estimator, generalised regression estimator, modified direct estimator or survey regression estimator etc. The Current Population Survey (CPS) provides direct estimates for median household income of four-person families. We will use the available CPS direct estimates to compare to the estimates of proposed model.

Let Y_{ij} and x_{ij} denote the CPS median household income and the IRS mean income recorded for the i^{th} state at the j^{th} time. Let $\boldsymbol{X}_{ij} = (1, x_{ij}, \cdots, x_{ij}^p)'$. We consider the small area model which consists of two parts. The first part of our model can be generally expressed as

$$\begin{aligned} Y_{ij} &= \beta_0 + \beta_1 x_{ij} + \cdots + \beta_p x_{ij}^p + b_i + e_{ij} \\ &= \boldsymbol{X}_{ij}'\boldsymbol{\beta} + b_i + e_{ij} \\ &= \theta_{ij} + e_{ij}, \end{aligned} \tag{2.1}$$

where $\theta_{ij} = \boldsymbol{X}_{ij}'\boldsymbol{\beta} + b_i$ is our target of interest. Here, b_i is state-specific random effect and e_{ij} is white noise errors for the i^{th} state at the j^{th} time ($i = 1, \cdots, m; j = 1, \cdots, t$). We assume $b_i \sim N\left(0, \sigma_b^2\right)$ and $e_{ij} \sim N\left(0, \sigma_{ij}^2/n_{ij}\right)$ where n_{ij} is the sample size corresponding to the i^{th} state at the j^{th} time.

To allow the covariates to influence the within-state variation, we consider the following log-linear representation to characterize the error variance,

$$log(\sigma_{ij}^2) = \boldsymbol{X}_{ij}'\boldsymbol{\tau}. \tag{2.2}$$

The exponential function ensures a positive multiplicative factor for any vector $\boldsymbol{\tau}$. Here, $\boldsymbol{\tau} = (\tau_0, \tau_1, ..., \tau_p)'$. In order to allow the within-state variance to vary across states, we can extend the above formulation as $\log(\sigma_{ij}^2) = \boldsymbol{X}_{ij}'\boldsymbol{\tau} + v_i$ where the state-specific effects v_i can be assumed to have a normal distribution with mean 0 and variance σ_v^2. The skewed, non-negative nature of the log-normal distribution makes it a reasonable choice for representing variances as has been previous done in diverse research areas. Thus, the second part of our model can be generally expressed as

$$\begin{aligned} log(\sigma_{ij}^2) &= \tau_0 + \tau_1 x_{ij} + \cdots + \tau_p x_{ij}^p + v_i \\ &= \boldsymbol{X}_{ij}'\boldsymbol{\tau} + v_i, \end{aligned} \tag{2.3}$$

where $\boldsymbol{\tau} = (\tau_0, \tau_1, \cdots, \tau_p)'$. In this setup, we perform the hierarchical Bayesian analysis.

Let $\boldsymbol{Y}_i = (Y_{i1}, \cdots, Y_{it})'$ be the response and $\boldsymbol{X}_i = (\boldsymbol{X}_{i1}, \cdots, \boldsymbol{X}_{it})'$ be the covariate for the i^{th} state at the j^{th} time ($i = 1, \cdots, m$; $j = 1, \cdots, t$). Let $\boldsymbol{\Omega}_i = (\boldsymbol{\theta}_i, \boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_b^2, \sigma_v^2, \boldsymbol{\sigma}_i^2)$ be the parameter space corresponding to the i^{th} state where $\boldsymbol{\theta}_i = (\theta_{i1}, \cdots, \theta_{it})'$ and $\boldsymbol{\sigma}_i^2 = (\sigma_{i1}^2, \cdots, \sigma_{it}^2)'$. The full parameter space is $\boldsymbol{\Omega} = \boldsymbol{\Omega}_1 \times \cdots \times \boldsymbol{\Omega}_m$. For the i^{th} state, the likelihood function can be written as

$$\begin{aligned} L(\boldsymbol{Y}_i, \boldsymbol{X}_i | \boldsymbol{\Omega}_i) &\propto L(\boldsymbol{Y}_i | \boldsymbol{\theta}_i, \boldsymbol{\sigma}_i^2) L(\boldsymbol{\theta}_i | \boldsymbol{\beta}, b_i) L(log(\boldsymbol{\sigma}_i^2) | \boldsymbol{\tau}, v_i) \\ &\propto \prod_{j=1}^{t} L(Y_{ij} | \theta_{ij}, \sigma_{ij}^2) L(\theta_{ij} | \boldsymbol{X}_{ij}'\boldsymbol{\beta}, \sigma_b^2) L(log(\sigma_{ij}^2) | \boldsymbol{X}_{ij}'\boldsymbol{\tau}, \sigma_v^2). \end{aligned} \tag{2.4}$$

Here $L(Z|\mu, \sigma^2)$ denotes a normal density with data Z, mean μ and variance σ^2.

To accomplish the Bayesian specification of our model, we need to assign prior distributions to the unknown parameters. We assume noninformative improper uniform prior for the polynomial coefficients $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$. Also, we suppose proper conjugate gamma priors on the inverse of the variance components (σ_b^2, σ_v^2). The prior distributions are assumed to be mutually independent. We choose small values (10^{-3}) for the gamma shape and rate parameters to make the priors diffuse in nature so that inference is mainly controlled by the data distribution. Thus, we have the following priors: $\boldsymbol{\beta} \sim uniform(R^{p+1}), \boldsymbol{\tau} \sim uniform(R^{p+1}), (\sigma_b^2)^{-1} \sim Gamma(a, b)$ and $(\sigma_v^2)^{-1} \sim Gamma(c, d)$.

The full posterior of the parameters given the data is obtained by combining the likelihood and the prior distribution as follows

$$p(\boldsymbol{\Omega}|\boldsymbol{Y}, \boldsymbol{X}) \propto \prod_{i=1}^{m} L(\boldsymbol{Y}_i, \boldsymbol{X}_i|\boldsymbol{\Omega}_i)\pi(\boldsymbol{\beta})\pi(\boldsymbol{\tau})\pi(\sigma_b^2)\pi(\sigma_v^2). \tag{2.5}$$

Our target of inference is $\theta_{ij}, i = 1, \cdots, m; j = 1, \cdots, t$, the true median household income of all the states. Because the marginal posterior distribution of θ_{ij} is analytically intractable, high dimensional integration needs to be carried out in a theoretical framework. However, this task can be easily accomplished in an MCMC framework by using Gibbs sampler to sample from the full conditionals of θ_{ij} and other relevant parameters. To diminish the effects of the starting distributions, the first d iterations of chain are discarded and posterior summaries are calculated based on the rest of the d iterates.

Since we use improper prior to accomplish the Bayesian specification of our model, we need to check posterior propriety for our model. Posterior propriety holds in our model and the proof is outlined below.

Theorem 2.1 Posterior propriety holds in our model.

Proof: Integrating first w.r.t. $\boldsymbol{\beta}$, we have

$$\begin{aligned} I_{\boldsymbol{\beta}} &\propto exp\left\{-\frac{1}{2\sigma_b^2}\sum_i \boldsymbol{\theta}_i'\boldsymbol{\theta}_i + \frac{1}{2\sigma_b^2}\left(\sum_i \boldsymbol{\theta}_i'\boldsymbol{X}_i\right)\left(\sum_i \boldsymbol{X}_i'\boldsymbol{X}_i\right)^{-1}\left(\sum_i \boldsymbol{X}_i'\boldsymbol{\theta}_i\right)\right\}(\sigma_b^2)^{1/2} \\ &\propto exp\left\{-\frac{1}{2}\left\{S'(I - T(T'T)^{-1}T')S\right\}\right\}(\sigma_b^2)^{1/2}. \end{aligned} \tag{2.6}$$

Let $Q = -\frac{1}{2}\left\{S'(I - T(T'T)^{-1}T')S\right\}$. Since $(I - T(T'T)^{-1}T')$ is idempotent, $S'[I - T(T'T)^{-1}T']S$ is non-negative, implying $exp(Q) \leq 1$. Next, we consider integration w.r.t. σ_b^2.

$$I_{\sigma_b^2} = \int (\sigma_b^2)^{-(\frac{tm-1}{2}+a)-1} exp\left\{-\frac{b}{\sigma_b^2}\right\} d\sigma_b^2 \leq M1 \tag{2.7}$$

where M1 is some positive constant. Let $\boldsymbol{\eta}_i = (log\sigma_{i1}^2, log\sigma_{i2}^2, \cdots, log\sigma_{iJ}^2)$. Integrating first w.r.t. $\boldsymbol{\tau}$, we have

$$I_{\boldsymbol{T}} \propto exp\left\{-\frac{1}{2\sigma_v^2}\sum_i \boldsymbol{\eta}_i'\boldsymbol{\eta}_i + \frac{1}{2\sigma_v^2}\left(\sum_i \boldsymbol{\eta}_i'\boldsymbol{X}_i\right)\left(\sum_i \boldsymbol{X}_i'\boldsymbol{X}_i\right)^{-1}\left(\sum_i \boldsymbol{X}_i'\boldsymbol{\eta}_i\right)\right\}(\sigma_v^2)^{1/2}$$
$$\propto exp\left\{-\frac{1}{2}\left\{K'(I-B(B'B)^{-1}B')K\right\}\right\}(\sigma_v^2)^{1/2}. \quad (2.8)$$

Let $W = -\frac{1}{2}\left\{K'(I-B(B'B)^{-1}B')K\right\}$. Since $(I-B(B'B)^{-1}B')$ is idempotent, $K'[I-B(B'B)^{-1}B']K$ is non-negative, implying $exp(W) \leq 1$. Again, we consider itegration w.r.t. σ_v^2.

$$I_{\sigma_v^2} = \int (\sigma_v^2)^{-(\frac{tm-1}{2}+c)-1} exp\left\{-\frac{d}{\sigma_v^2}\right\} d\sigma_v^2 \leq M2 \quad (2.9)$$

where M2 is some positive constant. Since all the components of the integrand have proper distributions, the above integral is finite. □

3. Numerical studies

We use income data from ASEC for the period 1995-1999 to estimate median household income for all the U.S states and District of Columbia for 1999. Our response is the CPS median household income while the covariate is the IRS mean household income for the U.S states for 1995-1999. Figure 3.1 demonstrates the CPS median income against IRS mean income for all the states for the period 1995-1999. Clearly, heteroscedasticity seems to be an issue with regard to Figure 3.1. Variation of CPS median income seems to increase with IRS mean income. To clarify this point, we plot log of the state-specific variance of the CPS median income values against state-specific means of the IRS mean income values. Note that correlation coefficients between IRS mean and log of variance of CPS is 0.4047. There is positive correlation between log of within-state variance of CPS income and state-specific IRS mean income. Based on the above information, we have restricted ourselves to a linear coefficient in mean and variance structure.

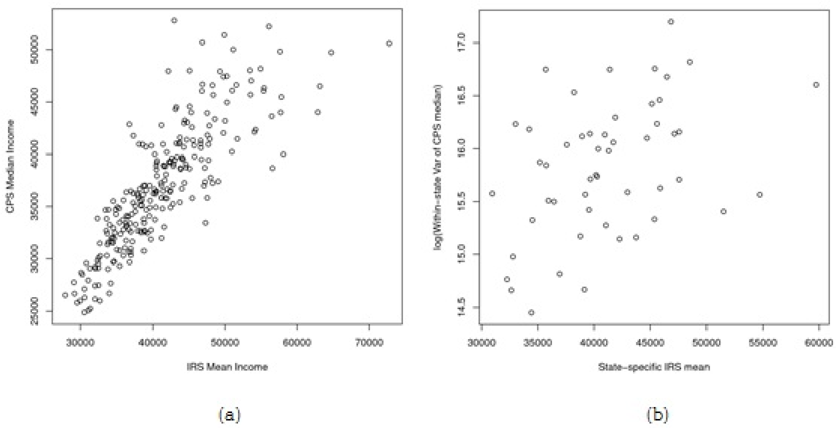

Figure 3.1 Scatter plot

The general structure of our model would remain the same as Section 2.1. Based on the observation, we only consider linear coefficient (p=1) in mean and variance structure and formulate the following small area model for our data:

$$Y_{ij} = x_{ij}\beta + b_i + e_{ij} \tag{3.1}$$

and

$$log(\sigma_{ij}^2) = x_{ij}\tau + v_i \tag{3.2}$$

where Y_{ij} is the CPS median income while x_{ij} is the IRS mean income. We assume $b_i \sim N\left(0, \sigma_b^2\right)$, $v_i \sim N\left(0, \sigma_v^2\right)$ and $e_{ij} \sim N\left(0, \sigma_{ij}^2/n_{ij}\right)$ where n_{ij} is the sample size corresponding to the i^{th} state at the j^{th} time.

We can infer all parameters by using Gibbs sampler to generate from the full conditional relevant parameters except for σ_{ij}^2. To perform Gibbs sampling, we need to calculate the full conditional distributions. The full conditional distributions of θ_{ij}, β, τ, σ_b^2, σ_v^2 and σ_{ij}^2 are given by

$$\theta_{ij}|Y_{ij}, x_{ij}, \beta, \sigma_b^2, \sigma_v^2, \sigma_{ij}^2 \sim N\left(\frac{\sigma_b^2 n_{ij} Y_{ij} + x_{ij}\beta\sigma_{ij}^2}{n_{ij}\sigma_b^2 + \sigma_{ij}^2}, \frac{\sigma_{ij}^2\sigma_b^2}{n_{ij}\sigma_b^2 + \sigma_{ij}^2}\right) \tag{3.3}$$

$$\beta|\theta_{ij}, x_{ij}, \sigma_b^2 \sim N\left(\left(\sum_i\sum_j x_{ij}\theta_{ij}\right)\left(\sum_i\sum_j x_{ij}^2\right)^{-1}, \left(\sum_i\sum_j x_{ij}^2\right)^{-1}\sigma_b^2\right) \tag{3.4}$$

$$\tau|x_{ij}, \sigma_{ij}^2, \sigma_v^2 \sim N\left(\sum_i\sum_j x_{ij}log(\sigma_{ij}^2)\left(\sum_i\sum_j x_{ij}^2\right)^{-1}, \left(\sum_i\sum_j x_{ij}^2\right)^{-1}\sigma_v^2\right) \tag{3.5}$$

$$(\sigma_b^2)^{-1}|\theta_{ij}, x_{ij}, \beta \sim G\left(a + \frac{tm}{2}, \frac{\sum_i\sum_j\left(\theta_{ij} - x_{ij}\beta\right)^2}{2} + b\right) \tag{3.6}$$

$$(\sigma_v^2)^{-1}|x_{ij}, \tau, \sigma_{ij}^2 \sim G\left(c + \frac{mt}{2}, \frac{\sum_i\sum_j(log(\sigma_{ij}^2) - x_{ij}\tau)^2}{2} + d\right) \tag{3.7}$$

$$\sigma_{ij}^2|others \sim \left(\sigma_{ij}^2\right)^{-\frac{3}{2}} exp\left(-\frac{n_{ij}(Y_{ij} - \theta_{ij})^2}{2\sigma_{ij}^2}\right) exp\left(-\frac{(log(\sigma_{ij}^2) - x_{ij}\tau)^2}{2\sigma_v^2}\right) \tag{3.8}$$

Since the full conditionals of σ_{ij}^2 are non-standard forms of distribution, we cannot use the Gibbs sampler to generate σ_{ij}^2. Thus, an alternative approach to generate σ_{ij}^2 is required. The grid method is known as one of approaches to sample from an unknown form. The grid method to generate σ_{ij}^2 can be described by the following steps.

Grid method:

Step 1. For $i = 1, \cdots, m, j = 1, \cdots, t$, set $\eta_{ij} = \frac{\sigma_{ij}^2}{1+\sigma_{ij}^2}$.

Step 2. Divide full interval (0,1) into 100 subintervals.

Step 3. Calculate mid-points for each intervals $(M_k, k = 1, \cdots, 100)$.

Step 4. Input the mid-points to the posterior density of η_{ij} $(a_k, k = 1, \cdots, 100)$.

Step 5. Calculate $b_k = \frac{\sum_{m=1}^{k} a_m}{\sum_{m=1}^{100} a_m}, k = 1, \cdots, 100$.

Step 6. Generate $u \sim U(0, 1)$ and find k^* such that $u \in [b_{k^*}, b_{k^*+1}]$.

Step 7. Generate s from $U(M_{k^*}, M_{k^*+1})$.

Then, s/(1-s) is sample from $\pi(\sigma_{ij}^2 | others)$.

In order to evaluate the performance of our estimates, we need criteria for comparison. Once every 10 years, the U.S. Bureau of the Census has conducted decennial census of population. We can use the census figures for 1999 to compare to the corresponding estimates of median household incomes for 1999. Thus, the decennial census values are regarded as "gold standard" against which all other estimates are compared using the following four criteria:

- Average Relative Bias (ARB) $= (51)^{-1} \sum_{i=1}^{51} \frac{|c_i - e_i|}{c_i}$
- Average Squared Relative Bias (ASRB) $= (51)^{-1} \sum_{i=1}^{51} \frac{|c_i - e_i|^2}{c_i^2}$
- Average Absolute Bias (AAB) $= (51)^{-1} \sum_{i=1}^{51} |c_i - e_i|$
- Average Squared Deviation (ASD) $= (51)^{-1} \sum_{i=1}^{51} (c_i - e_i)^2$

Here, c_i and e_i respectively denote the census and model based estimates for the i^{th} state $(i = 1, \cdots, 51)$.

Convergence of the algorithm refers to whether the algorithm has reached its equilibrium distribution. Hence, monitoring the convergence of the algorithm is essential for producing results from the posterior distribution of interest. Convergence of the Gibbs sampler with grid method was monitored by visually checking the dynamic trace plots and the acf plots. We need to check the autocorrelations of the generated values since the MCMC generated sample may not be independent. Figure 3.2 plots the acf with thin interval for (a) β, (b) τ, (c) σ_b^2 and (d) σ_v^2. It indicates that the generated samples are not independent. So, we can produce independent samples by keeping the first generated values in every batch of 5 iterations. The trace plots for observations after discarding a burning period of 1000 iterations are provided in Figure 3.3. Generated observations of Figure 3.3 are convincing in terms of convergence, with all generated values within a parallel zone and no obvious tendencies or periodicities. From now, we will use these finally generated observations to estimate parameters for our model.

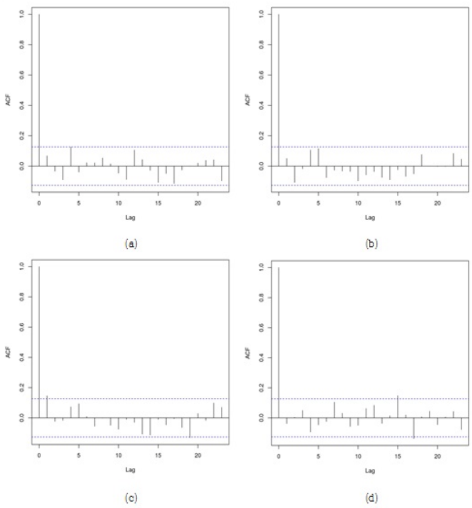

Figure 3.2 ACF for (a) β, (b) τ, (c) σ_b^2 and (d) σ_v^2

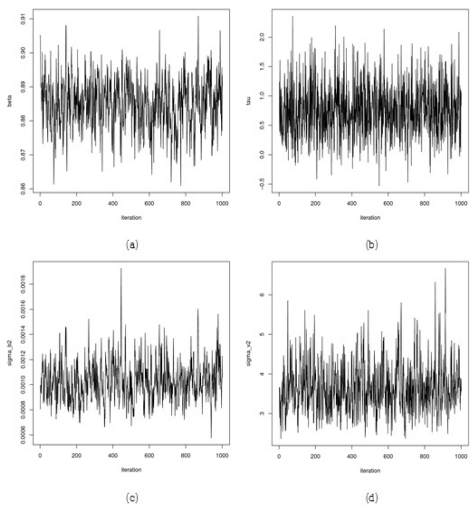

Figure 3.3 Trace plots with thin interval for (a) β, (b) τ, (c) σ_b^2 and (d) σ_v^2

Table 3.1 reports parameter estimates of small area model with variance structure (model 1). As mentioned in Section 3.1, it is of interest to note that β and τ are significant for our model. Furthermore, it is seen that τ demonstrates positive correlation between log of CPS income and IRS mean income.

Table 3.1 Parameter estimates of small area model with variance structure.

Parameter	Mean	Median	90% CI
β	0.8850	0.8854	(0.8723, 0.8966)
τ	0.7578	0.7269	(0.0510, 1.5630)
σ_b^2	0.0010	0.0010	(0.0008, 0.0013)
σ_v^2	3.6470	3.5854	(2.7532, 4.7687)

Parameter estimates in small area model without variance structure (model 2) are given in Table 3.2. In this table, it is seen that there is little difference in the estimates.

Table 3.2 Parameter estimates of small area model without variance structure.

Parameter	Mean	Median	90% CI
β	0.8820	0.8820	(0.8707, 0.8930)
σ_b^2	0.00007	0.00007	(0.00005, 0.0009)

Table 3.3 shows the comparison measurements for the CPS estimates and the small area model estimates while Table 3.4 depicts the percentage improvement of the small area model estimates over the CPS estimates. It is clear that the comparison measures for the small area model with variance structure are lower than those corresponding to the CPS estimates as well as those of model 2.

Table 3.3 Comparison measurements

Estimate	ARB	ASRB	AAB	ASD
CPS	0.0415	0.0027	1753.33	5,300,023
Model 1	0.0332	0.0022	1430.86	4,610,690
Model 2	0.0354	0.0027	1528.43	5,890,131

Table 3.4 Percentage improvements of small area model with variance structure

Estimate	ARB	ASRB	AAB	ASD
Model 1	20.00%	18.52%	18.39%	13.01%

4. Concluding remarks

In this paper, we have obviously seen that the IRS mean income has positive correlation not only with CPS median income but also with log of the state-specific variance of the CPS median income. Considering this point, we have proposed the small area model including variance structure part to explain the inherent patterns. This plays an important role in handling heterogeneous variance, which yields adequate small area estimations. It is seen that estimations from the proposed model is slightly better than CPS method in terms of four criteria which we have mentioned. The proposed model including variance structure makes some improvements by capturing the underlying pattern in data. Thus, when overdispersion is present in small area estimation, this approach could be an alternative to overdispersed mean modeling. Also, we can extend our model by allowing two state-specific random effects in the mean and variance of a state to be correlated to account for the association with each other.

References

Bhadra, D., Ghosh, M. and Kim, D. (2012). Estimation of median household income for small areas: A Bayesian semiparametric approach. *Calcutta Statistical Association Bulletin*, **64**, 115-142.

Ghosh, M., Nangia, N., and Kim, D. H. (1996). Estimation of median income of four-person families : A Bayesian time series approach. *Journal of the American Statistical Association*, **91**, 1423-1431.

Ghosh, M. and Rao, J. N. K. (1994). Small area estimation: An appraisal. *Statistical Science*, **9**, 55-76.

Goo, Y. M. and Kim, D. H. (2013). Bayesian small area estimations with measurement errors. *Journal of the Korean Data & Information Science Society*, **24**, 885-893.

Lee, J. and Kim, D. H. (2013). Bayesian estimation of median household income for small areas with some longitudinal pattern. *Journal of the Korean Data & Information Science Society*, **26**, 755-762.

Rao, J. N. K. (2003). *Small area estimation*, 2nd Ed., Wiley, New York.