Statistics and Applications {ISSN 2454-7395 (online)}
Special Issue in Memory of Prof. C R Rao
Volume 22, No. 3, 2024 (New Series), pp 71–94
http://www.ssca.org.in/journal

Hierarchical Bayesian Probit Models for Sub-Areas and Ordinal Data

Lu Chen[1] **and Balgobin Nandram**[2]
[1]*National Institute of Statistical Sciences, Washington D.C., US*
[2]*Department of Mathematical Sciences*
Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA, US

Received: 12 April 2024; Revised: 16 May 2024; Accepted: 28 May 2024

Abstract

Many population-based surveys have polychotomous responses from a number of individuals in each household within small areas. An example is the second Nepal Living Standards Survey (NLSS II), in which health categorical data for each individual from the sampled households (sub-areas) are available in sampled wards (small areas). When the survey responses are ordinal, the sub-area hierarchical Bayesian probit models are considered to make inference about finite population proportions of individuals with different health statuses within the small areas. A standard assumption is that the ordered categorical responses are determined by an unobservable continuous variable. We discuss how to fit the model to avoid poor mixing problems in Markov chain Monte Carlo methods when simulating samples from the joint posterior distribution. The application is on health status data in the NLSS II, and the sub-area and the small area models are compared. The results show that the sub-area models are preferred over the small area models that ignore households (sub-areas) within the wards (areas). Our theoretical and methodological work can help provide small area official statistics for numerous surveys worldwide.

Key words: Bayesian Inference; Hierarchical Bayesian model; Metropolis-Hastings algorithm; Ordinal Variables; Small Area Estimation

1. Introduction

Most sample surveys are designed to provide reliable estimates of totals, means and other parameters of interest for large areas or domains (e.g., state level, national level). Such estimates are usually called "direct" estimates if they are only based on the domain-specific sample data. However, direct estimates are not reliable for the areas or domains for which only small samples or no samples are available. In recent years, more and more policymakers demand small area estimates. In fact, many new programs, such as fund allocation for needed areas, new educational or health programs, rely heavily on these estimates. Taking the cost and operation issues into consideration, it is not practical to conduct surveys with large

Corresponding Author: Lu Chen
Email: lchen@niss.org

enough sample sizes within the areas. In particular, small area estimation (SAE) deals with the problem of how to produce reliable "indirect" estimates of characteristics of interest for the small areas or domains.

Small area models are generally classified into two broad types. The basic area level model was introduced originally for SAE by Fay and Herriot (1979). The area level model is applied when individual auxiliary information is not available. Unit level model was first proposed for SAE by Battese, Harter and Fuller (1988). The generalized linear mixed model (GLMM) is one extension of the basic unit-level models. It was considered for SAE by MacGibbon and Tomberlin (1989). GLMM is useful in the case that the small area quantities of interest are finite population proportions.

In this paper, we are particularly interested in the small area models that can capture hierarchical structures, such as the Nepal Living Standard Survey II (NLSS II) data. The sampling scheme of NLSS II is a two-stage stratified sampling design. Nepal is stratified into primary sample units (wards) and within each ward, twelve households (sub-area) are systematically selected and all individuals from the selected households are interviewed. Although the above basic models are very popular and in common use in producing reliable estimates, the hierarchical structure of the data and the consistency between the estimates for different levels may not hold. Therefore, we focus on two-fold models, an important extension of basic small area models.

Hierarchical Bayesian methods are very popular in the two-fold models. Yan and Sedransk (2007) studied the case that the data follow a normal model with a two-stage (three-stage) hierarchical structure while the fitted model has a one-stage (two-stage) hierarchical structure by using posterior predictive p-values. Yan and Sedransk (2010) discussed the ability to detect a three-stage model when a two-stage model is actually fitted. Nandram (2016) and Chen and Nandram (2022) showed that it is important to consider the sample design within each area and proposed a two-fold small-area Beta-Binomial model. Lee *et al.* (2017) use a Bayesian method to infer about a finite population proportion when binary data are collected using a two-fold sample design from small areas. Erciulescu *et al.* (2018), Chen *et al.* (2022), and Nandram *et al.* (2023) illustrated hierarchical Bayesian approaches to provide estimates for the sub-area models with and without constraints. Chen and Nandram (2023) proposed a hierarchical Bayesian logistic regression model for binary data in small area estimation. This model is a unit level model with the sub-area effect. The results show that two-fold models can capture the heterogeneity between samples within not only small areas but also sub-areas.

Many population-based surveys have polychotomous responses from a number of individuals in each household within small areas, and many responses are ordered. For example, in the NLSS II, the answers to the question on health status range from 1 to 4, four options (excellent, good, fair, poor). There are few studies for ordinal response variables in SAE. Early papers on regression models for ordinal data include McKelvey and Zavoina (1975), McCullagh (1980), and Winship and Mare (1984). Nandram (1989) discussed the discrimination between the log-log link and logit link models for ordinal data. The textbook of Agresti (2010) gives a thorough treatment of ordinal data, while O'Connel (2006) provides applied researchers in the social sciences with accessible and comprehensive coverage of analysis for ordinal outcomes.

Albert and Chib (1993) discussed the algorithm to fit the Bayesian ordinal regression model with probit link. They introduced an underlying continuous variable, Z with a standard normal cumulative distribution function Φ. The ordinal response variable, Y_i is then observed in category t if Z_i comes from Normal$(\underset{\sim}{x}_i^T\underset{\sim}{\beta}, 1)$ between the cutpoints $\theta_{t-1} < Z_i \leq \theta_t$ and $\underset{\sim}{x}$ are the covariates. To capture the ordinal nature of the observed data, the cut-points are constrained to be monotonically increasing, $-\infty = \theta_0 < \theta_1 < \ldots < \theta_{T-1} < \theta_T = +\infty$. In addition, they assume that Z_i follows scale mixture of normal distributions, that is, Normal$(\underset{\sim}{x}_i^T\underset{\sim}{\beta}, \lambda_i^{-1})$. They assume that the underlying continuous variables follow the normal distribution without subgroup random effects. In this paper, we focus on the heterogeneous variances among the small areas and the subareas and conduct the subgroup analysis. We start with their models and build additional models with the small area and sub-area random effects.

For the probit analysis, Holmes and Held (2006) discussed Albert and Chib (1993) algorithm and showed that it gives a poorly mixing Gibbs sampler. They showed how to solve this mixing problem by adding latent variables and using the block Gibbs sampler (*i.e.*, some variables are drawn simultaneously). In this paper, we discuss how to fit the heterogeneous model to avoid poor mixing problems in Markov chain Monte Carlo methods when simulating samples from the joint posterior distribution.

In Section 2, a full description of the area and sub-area hierarchical Bayesian ordered probit models is given. In Section 3, we apply the models to the NLSS II data to predict the four health conditions of the household proportions of members for both sampled and nonsampled households. The comparisons between the small area models and the subarea models are presented. Finally, in Section 4, we make concluding remarks and discuss the future work. Technical details are given in the appendices.

2. Bayesian ordered probit models with covariates

In this section, we discuss two hierarchical Bayesian ordered probit models with covariates: the heterogeneous small area model and the heterogeneous sub-area model. We explain in detail about how to draw samples from the joint posterior distributions of heterogeneous models to avoid poor mixing problems in MCMC algorithm.

Suppose that the Y_i are categorical responses, falling in $t = 1, \ldots, T$, categories. Then Y_i follows a multinomial distribution with parameter $\underset{\sim}{p}$ where p_{it} denotes the probability that the i^{th} observation falls in the response category t. The cumulative probabilities are

$$\gamma_{it} = P(Y_i \leq t) = p_{i1} + \cdots + p_{it}.$$

Let $g(\cdot)$ denote a link function mapping probabilities to the real line, $g(\gamma_{it}) = \theta_t + \underset{\sim}{x}_i^T\underset{\sim}{\beta}$, where $\underset{\sim}{x}_i^T$ is a vector of explanatory variables for the i^{th} observation and $\underset{\sim}{\beta}$ is the corresponding set of regression parameters. The θ_t parameters are constant representing the baseline value for category t. Notice that the predictors do not include a column of ones for the intercept term since the constants are written explicitly. In this paper, we primary discuss the model with probit link function within the Bayesian paradigm, that is, $g(\cdot) = \Phi^{-1}(\cdot)$.

2.1. Heterogeneous small area model

In this section, we focus on the model with heterogeneous variances among small areas. In the small area models, we ignore the differences among households. Assume that there are ℓ areas, within the i^{th} area there are M_i individuals. For sampling, m_i individuals are selected from the M_i units available. Suppose that the independent response $y_{ij},\ i=1,\ldots,\ell,\ j=1,\ldots,m_i$ are observed and y_{ij} takes one of T ordered categories, *i.e.*, $y_{ij}\in\{1,2,...,T\}$.

The interest is to provide indirect estimates of the finite population proportions of the small areas in each category which are

$$\bar{P}^a_{it}=\frac{1}{M_i}\sum_{j=1}^{M_i} I(y_{ij}=t)=f^a_i\bar{I}^a_{sti}+(1-f^a_i)\bar{I}^a_{nsti}\ ,\ i=1,\ldots,\ell,\ t=1,\ldots,T,$$

where a denotes the small area estimates, and $\bar{I}^a_{sti}=\sum_{j=1}^{m_i} I(y_{ij}=t)/m_i$, $\bar{I}^a_{nsti}=\sum_{j=m_i+1}^{M_i} I(y_{ij}=t)/(M_i-m_i)$, and $f^a_i=m_i/M_i$ are sampled proportions, non-sampled proportions and sample fraction respectively in the small area model. Bayesian predictive inference is required for non-sample proportions.

Define the underlying continuous variable z_{ij}, where the z_{ij} follows Normal$(\underset{\sim}{x}^T_{ij}\underset{\sim}{\beta}+\nu_i,\lambda_i^{-1})$ with small area random effect ν_i. If $\theta_{t-1}<z_{ij}\le\theta_t$, then $y_{ij}=t$. Define $\theta_0=-\infty$ and $\theta_T=\infty$. θ_t is a constant representing the baseline value for category t. Since the variances of latent variable z_{ij} vary within areas, we called this ordered probit model heterogeneous small area model. Therefore, the small area Bayesian ordinal probit model with heterogeneous variances is

$$z_{ij}|\underset{\sim}{\nu},\underset{\sim}{\beta},\lambda_i,\underset{\sim}{x},\underset{\sim}{y}\overset{ind}{\sim}\text{Normal}(\underset{\sim}{x}^T_{ij}\underset{\sim}{\beta}+\nu_i,\lambda_i^{-1}),\tag{1}$$

where $\theta_{t-1}<z_{ij}\le\theta_t$ if $y_{ij}=t$ and the priors are

$$\nu_i|\delta^2\overset{iid}{\sim}\text{Normal}(0,\delta^2),\ i=1,\ldots,\ell,$$

$$\underset{\sim}{\beta}\sim\text{MN}\left(\underset{\sim}{\beta}_0,1000\Sigma_0\right),$$

$$\lambda_i|a\overset{iid}{\sim}\text{Gamma}(a,a),\ i=1,\ldots,\ell,$$

$$\pi(a,\delta^2)=\frac{1}{(1+a)^2}\frac{1}{(1+\delta^2)^2},$$

$$\pi(\theta_t)=(n-1)!I(\theta_1<\ldots<\theta_{T-1}),\ t=1,\ldots,T-1.$$

A diffuse prior is placed on the coefficient $\underset{\sim}{\beta}$. The prior of λ_i is gamma distribution, which makes the latent variable z_{ij} follows a student's t distribution. We placed the shrinkage priors on both a and δ^2 so that they are proper but with heavy tail. The detail of how to obtain a sample from joint posterior density is shown in Appendix A.

If the variances among the small areas are the same, no λ_i but 1, we call that model

a homogeneous small area model. That is,

$$z_{ij}|\underset{\sim}{\nu},\underset{\sim}{\beta},\underset{\sim}{x},\underset{\sim}{y} \overset{ind}{\sim} \text{Normal}(\underset{\sim}{x}_{ij}^T\underset{\sim}{\beta}+\nu_i,1), \tag{2}$$

where $\theta_{t-1} < z_{ij} \leq \theta_t$ if $y_{ij} = t$ and the priors are

$$\nu_i|\delta^2 \overset{iid}{\sim} \text{Normal}(0,\delta^2),\ i=1,\ldots,\ell,$$

$$\underset{\sim}{\beta} \sim \text{MN}\left(\underset{\sim}{\beta}_0, 1000\Sigma_0\right),$$

$$\pi(a,\delta^2) = \frac{1}{(1+a)^2}\frac{1}{(1+\delta^2)^2},$$

$$\pi(\theta_t) = (n-1)!I(\theta_1 < \ldots < \theta_{T-1}),\ t=1,\ldots,T-1.$$

Since the heterogeneous model is more general than the homogeneous model, the methods can be easily applied to the homogeneous model and the computations are simpler.

2.2. Heterogeneous sub-area model

We focus on the model with heterogeneous variances among sub-areas and discuss how to fit it. In sub-area models, we assume that there are ℓ areas, within the i^{th} area there are N_i sub-areas (households) and within the j^{th} sub-areas, there are M_{ij} individuals. For sampling, n_i sub-areas are sampled from the N_i sub-areas and all individuals are selected in sampled sub-areas, that is, $m_{ij} = M_{ij}$. Let $y_{ijk},\ k=1,\ldots,m_{ij}, j=1,\ldots,n_i, i=1,\ldots\ldots,\ell$, denote the categorical response and y_{ijk} takes one of T ordered categories, *i.e.*, $y_{ijk} \in \{1,2,...,T\}$.

The interest is also to provide estimates of the finite population proportions of small areas in each category which are

$$\bar{P}_{it}^s = \frac{1}{\sum_{j=1}^{N_i} M_{ij}} \sum_{j=1}^{N_i}\sum_{k=1}^{M_{ij}} I(y_{ijk}=t) = f_i^s \bar{I}_{sti}^s + (1-f_i^s)\bar{I}_{nsti}^s\ ,\ i=1,\ldots,\ell,\ t=1,\ldots,T,$$

where s denotes the estimates considering sub areas, and $\bar{I}_{sti}^s = \sum_{j=1}^{n_i}\sum_{k=1}^{m_{ij}} I(y_{ijk}=t)/\sum_{j=1}^{n_i} m_{ij}$, $\bar{I}_{nsti}^s = \sum_{j=n_i+1}^{N_i}\sum_{k=1}^{M_{ij}} I(y_{ijk}=t)/\sum_{j=n_i+1}^{N_i} M_{ij}$, and $f_i^s = \sum_{j=1}^{n_i} M_{ij}/\sum_{j=1}^{N_i} M_{ij}$ are sampled proportions, non-sampled proportions and sample fraction in sub-area models respectively. Bayesian predictive inference is required for non-sample proportions.

Let the z_{ijk} follow Normal$(\underset{\sim}{x}_{ijk}^T\beta+\nu_i+\mu_{ij},\lambda_i^{-1})$ distribution with the small area random effects ν_i and sub-area random effects μ_{ij}. If $\theta_{t-1} < z_{ijk} \leq \theta_t$, then $y_{ijk} = t$. Since the variance of latent variable z_{ijk} are different among small areas, we call this ordered probit model a heterogeneous sub-area model.

Our sub-area Bayesian ordered probit model as

$$z_{ijk}|\underset{\sim}{\nu},\underset{\sim}{\beta},\lambda_i,\underset{\sim}{x},\underset{\sim}{y} \overset{ind}{\sim} \text{Normal}(\underset{\sim}{x}_{ijk}^T\underset{\sim}{\beta}+\nu_i+\mu_{ij},\lambda_i^{-1}), \tag{3}$$

where $\theta_{t-1} < z_{ijk} \leq \theta_t$ if $y_{ijk} = t$ and the priors are

$$\mu_{ij}|\sigma^2 \overset{iid}{\sim} \text{Normal}(0, \sigma^2),\ j = 1, \ldots, n_i,$$

$$\nu_i|\delta^2 \overset{iid}{\sim} \text{Normal}(0, \delta^2),\ i = 1, \ldots, \ell,$$

$$\underset{\sim}{\beta} \sim \text{MN}\left(\underset{\sim}{\beta}_0, 1000\Sigma_0\right),$$

$$\lambda_i|a \overset{iid}{\sim} \text{Gamma}(a, a),\ i = 1, \ldots, \ell,$$

$$\pi(a, \sigma^2, \delta^2) = \frac{1}{(1+a)^2}\frac{1}{(1+\delta^2)^2}\frac{1}{(1+\sigma^2)^2},$$

$$\pi(\theta_t) = (T-1)!I(\theta_1 < \ldots < \theta_{T-1}),\ t = 1, \ldots, T-1.$$

Similarly, the detail of how to obtain a sample from this joint posterior density is shown in Appendix B.

If the variances among areas are the same, no λ_i but 1, we call that model as homogeneous sub-area model, that is

$$z_{ijk}|\underset{\sim}{\nu}, \underset{\sim}{\beta}, \underset{\sim}{x}, \underset{\sim}{y} \overset{ind}{\sim} \text{Normal}(\underset{\sim}{x}_{ijk}^T\underset{\sim}{\beta} + \nu_i + \mu_{ij}, 1), \tag{4}$$

where $\theta_{t-1} < z_{ijk} \leq \theta_t$ if $y_{ijk} = t$ and the priors are

$$\mu_{ij}|\sigma^2 \overset{iid}{\sim} \text{Normal}(0, \sigma^2),\ j = 1, \ldots, n_i,$$

$$\nu_i|\delta^2 \overset{iid}{\sim} \text{Normal}(0, \delta^2),\ i = 1, \ldots, \ell,$$

$$\underset{\sim}{\beta} \sim \text{MN}\left(\underset{\sim}{\beta}_0, 1000\Sigma_0\right),$$

$$\pi(a, \sigma^2, \delta^2) = \frac{1}{(1+a)^2}\frac{1}{(1+\delta^2)^2}\frac{1}{(1+\sigma^2)^2},$$

$$\pi(\theta_t) = (T-1)!I(\theta_1 < \ldots < \theta_{T-1}),\ t = 1, \ldots, T-1.$$

Since the heterogeneous model is more general than the homogeneous model, the methods can be easily applied to the homogeneous model and the computations are simpler.

2.3. Prediction

In this paper, our interest is to predict the finite population proportions of the 102 sampled wards in both sampled and non-sampled households. The covariates of individuals in non-sampled households and the size of non-sampled households are unknown. Bayesian bootstrap (Rubin 1981) is used to draw them. The bootstrapping is done within sampled wards. The detail of the Bayesian bootstrap procedure is shown in Appendix C. Bayesian predictive inference for the individuals in the non-sampled sub-areas within the sampled small areas can be made once the set of samples are obtained from the posterior distribution.

For the small area models, we can draw samples of the non-sampled underlying vari-

able, $z_{ij}^{(h)},\ h=1,\ldots,M,\ j=m_i+1,\ldots,M_i,\ i=1,\ldots,\ell$, based on the likelihood functions in the models, where h denote the h^{th} samples drawn from the predictive distribution and we draw M samples in total. Then given the set of samples of $\underset{\sim}{\theta}$, the non-sampled responses, y_{ij}, can be predicted based on the criteria:

$$\theta_{t-1}^{(h)} < z_{ij} \leq \theta_t^{(h)},\ \text{then} \quad y_{ij}=t,\ t=1,\ldots,T.$$

For the sub-area models, we can draw samples of the non-sampled underlying variable, $z_{ijk}^{(h)},\ \ h=1,\ldots,M,\ k=1,\ldots,M_{ij},\ j=n_i+1,\ldots,N_i,\ i=1,\ldots,\ell$, based on the likelihood functions in the models. Then given the set of samples of $\underset{\sim}{\theta}$, the non-sampled responses, y_{ijk}, can be predicted based on the criteria:

$$\theta_{t-1}^{(h)} < z_{ijk} \leq \theta_t^{(h)},\ \text{then} \quad y_{ijk}=t,\ t=1,\ldots,T.$$

3. Application

3.1. Nepal living standards survey II

In this section, we describe the second Nepal Living Standards Survey (NLSS II) and the responses and the covariates. The performance of our method is studied using NLSS II, conducted in the years 2003-2004. NLSS is a national household survey in Nepal, actually population based (*i.e.*, interviews are done for all individual household members). Sometimes the head of the household answers the questions. NLSS follows the World Bank Living Standards Measurement Survey methodology with a two-stage stratified sampling scheme. It is an integrated survey which covers samples from the whole country. The main objective of the NLSS is to collect data from Nepalese households and provide information to monitor progress in national living standards. We study the polychotomous variable, health status, from the health section of the questionnaire.

The sampling design of NLSS II is two-stage stratified sampling. One selects the primary units (small areas) in the first stage and then some of the units (sub-areas) are selected from the secondary stage. Figure 1 shows that the area level of NLSS II is wards (circle) and the sub-area level is all selected households (house shape). That is, Nepal is stratified into primary sample units (wards) and within each ward, twelve households (sub-areas) are systematically selected. All household members in the sample were interviewed. Note that any analysis is done for each stratum.

According to the 2001 census data, only about 0.091% of households and only 0.904% of wards were sampled. NLSS II was designed to provide reliable estimates only at stratum level or even larger areas than stratum. It cannot give reliable estimates in small areas (ward or household level) since the sample sizes are too small. Therefore, we need to use statistical models to fit the available data and find reliable estimates in small areas.

3.2. Response variables and covariates

NLSS II has sparse counts of household members within the wards for four health status groups: excellent, good, fair and poor, denoted by 1 to 4. The distribution of all

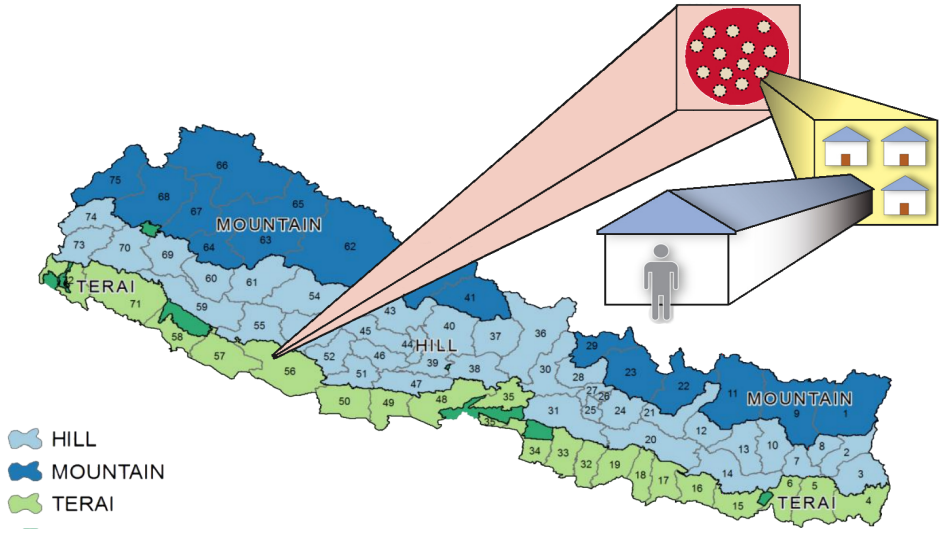

Figure 1: Illastration of NLSS II two-stage sampling design

responses of the health conditions in each stratum is shown in Table 1. Notice that the counts in the fair and poor cells are very sparse. There are six strata in the NLSS II. In this paper, we study the Rural Terai, the largest stratum in Nepal. It has 102 wards with 7,034 individuals in the sample of 12,239 wards in the population with 9,744,810 people. The number of people in the sample is 7,034 with 3,950 in the "excellent" cell, 2,926 in the "good" cell, 153 in the "fair" cell and 5 in the "poor" cell with percentages 56.1%, 41.6%, 2.1% and 0.02%. Notice that the counts in the last cell are mostly zeros.

Table 1: Distributions of wards and households in the sample and the distributions of the responses in each health cell

Stratum	Ward	Household	Individual	Excellent	Good	Fair	Poor
Mountains	32	384	1,949	1,262	658	24	4
Kathemandu	34	408	1,954	1,494	459	1	0
Urban Hills	28	336	1,467	820	626	20	1
Rural Hills	96	1,152	5,755	3,028	2,613	110	4
Urban Terai	34	408	2,104	1,239	811	52	2
Rural Terai	102	1,224	7,034	3,950	2,926	153	5

We choose four relevant covariates which can influence health status from the NLSS II survey for our sub-area logistic model and ordered probit models. They are age, nativity, sex and religion. We created binary variables: nativity (Indigenous = 1, Non-indigenous = 0), religion ((Hindu = 1, Non-Hindu = 0), and sex (Male = 1, Female = 0).

Table 2 shows some details of these 4 covariates. In the model fitting, we standardize age covariate. Elder age and children's age are more vulnerable than younger age. Indigenous people can have different health status from migrated people.

In NLSS II, the ordinal response variable of health status has 4 categories, from 1 to 4, where 1 means excellent health condition and 4 means poor health condition, respectively. When respondents answer this question, there is an underlying order among 1 to 4. Then the baseline values are $-\infty = \theta_0^* < \theta_1^* < \theta_2^* < \theta_3^* < \theta_4^* = +\infty$. In order to make the computation simpler, we subtract θ_2^* in each side and then $-\infty < \theta_1 < 0 < \theta_2 < +\infty$, where $\theta_1 = \theta_1^* - \theta_2^*$ and $\theta_2 = \theta_3^* - \theta_2^*$.

Table 2: Summaries of the four covariates: age, gender, nativity, and religion

Covariates		Frequency	Percentage
Age	0-14	7,765	38.32
	15-59	10,951	54.04
	60+	1,547	7.64
Gender	Male	9,763	48.18
	Female	10,500	51.82
Nativity	Indigenous	11,903	41.25
	Non-Indigenous	8,360	58.75
Religion	Hindu	16,378	80.83
	Non-Hindu	3,385	19.17

3.3. Numerical results

In this section, we show the numerical results and comparisons among the four models: homogeneous and heterogeneous wards models (small area models); homogeneous and heterogeneous household models (sub-area models).

3.3.1. MCMC diagnostics

For each of four models, we run 12,000 MCMC iterations, burn in 2,000 and thin every 10th to obtain 1,000 converged posterior samples. Table 3 and Table 5 give the p-values of the Geweke test and the effective sample sizes for the parameters $\underset{\sim}{\beta}$, θ_1, θ_2 and δ^2 of the homogeneous models. Table 4 gives the p-values of the Geweke test and the effective sample sizes for the parameters $\underset{\sim}{\beta}$, θ_1, θ_2, a and δ^2 of the heterogeneous area model. Table 6 gives the p-values of the Geweke test and the effective sample sizes for the parameters $\underset{\sim}{\beta}$, θ_1, θ_2, a, δ^2 and σ^2 for the heterogeneous household model. The p-values are all large, so we do not reject the null hypothesis test which is that the Markov chain is in the stationary distribution. The effective sample sizes are not too far away from 1,000. These model diagnostic summaries indicate that the MCMC chains converge and strongly mixing.

3.3.2. Model comparisons

For evaluating and comparing these models, the Bayesian posterior predictive p-value (Meng,1994), the deviance information criterion (DIC) and the logarithm of the pseudo marginal likelihood (LPML) are computed.

In the ordered probit models, denote $\Omega = (\underset{\sim}{\nu}, \underset{\sim}{\mu}, \underset{\sim}{\theta}, \underset{\sim}{\beta}, \underset{\sim}{\lambda})$. Since the responses y_{ijk} follow

Table 3: Summary of MCMC diagnostics: posterior mean, posterior standard deviation, the p-values of the Geweke test and the effective sample sizes for the homogeneous wards model

Model	Homogeneous Wards Model			
	Mean	SD	Geweke pval	Effective Size
β_1	0.08817	0.02238	0.34	1000
β_2	0.00038	0.02818	0.65	900
β_3	-0.02079	0.02555	0.57	1000
β_4	-0.37098	0.02379	0.71	1123
θ_1	-0.50001	0.00021	0.11	1000
θ_2	0.59635	0.60952	0.54	1000
δ^2	0.59320	0.11752	0.35	1092

Table 4: Summary of MCMC diagnostics: posterior mean, posterior standard deviation, the p-values of the Geweke test and the effective sample sizes for the heterogeneous wards model

Model	Heterogeneous Wards Model			
	Mean	SD	Geweke pval	Effective Size
β_1	0.1711	0.0148	0.78	1000
β_2	-0.0404	0.0164	0.48	910
β_3	-0.0074	0.0140	0.46	1000
β_4	-0.3467	0.0103	0.17	1000
θ_1	-0.5028	0.0044	0.28	908
θ_2	0.5963	0.6706	0.69	1000
a	0.8645	0.9327	0.51	1000
δ^2	1.4927	0.0396	0.12	875

Table 5: Summary of MCMC diagnostics: posterior mean, posterior standard deviation, the p-values of the Geweke test and the effective sample sizes for the homogeneous household model

Model	Homogeneous Household Model			
	Mean	SD	Geweke pval	Effective Size
β_1	0.10234	0.02344	0.88	1000
β_2	-0.00755	0.02477	0.71	1006
β_3	-0.02113	0.02562	0.93	888
β_4	-0.05413	0.02141	0.93	598
θ_1	-0.50000	0.00020	0.67	1000
θ_2	0.58684	0.10882	0.66	1000
δ^2	0.55568	0.13674	0.78	901
σ^2	0.03291	0.01905	0.30	855

Table 6: Summary of MCMC diagnostics: posterior mean, posterior standard deviation, the p-values of the Geweke test and the effective sample sizes for the heterogeneous household model

Model	Heterogeneous Household Model			
	Mean	SD	Geweke pval	Effective Size
β_1	0.1752	0.0149	0.78	1000
β_2	-0.0132	0.0177	0.39	1000
β_3	-0.0403	0.0159	0.69	1000
β_4	0.0089	0.0192	0.24	1000
θ_1	-0.4843	0.0044	0.51	888
θ_2	0.5584	0.0622	0.65	1000
a	0.8498	0.0997	0.52	1000
δ^2	1.6472	0.9333	0.16	1000
σ^2	0.3367	0.1957	0.68	1000

multinomial distributions, we consider a measure of form

$$T(\underset{\sim}{y},\Omega)=\sum_{t=1}^{T}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\sum_{k=1}^{m_{ij}}\frac{(I(y_{ijk}=t)-p_{ijkt})^2}{n_t p_{ijkt}(1-p_{ijkt})},$$

where $n_t=\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\sum_{k=1}^{m_{ij}}I(y_{ijk}=t)$ is the total number of y_{ijk} in t category and $p_{ijkt}=\Phi(\theta_t-\underset{\sim}{x}_{ijk}^T\underset{\sim}{\beta}-\nu_i-\mu_{ij})$. We calculate $T(\underset{\sim}{y}^{rep},\Omega)$ for each of 1,000 samples, and then seeing what percent are above single calculated $T(\underset{\sim}{y}^{obs},\Omega)$. The Bayesian posterior predictive p-value (BPP) is used in order to check the discrepancy between data and the posited model. The BPPs of all models shown in Table 7 are not in the extreme range (close to 0 or 1). Therefore, they are appropriate and adequate to make inference for the finite population proportions of interest. Note that the BPP cannot be used for ranking the models, but for checking if the model is good or not.

In addition, we calculated their DICs and LPMLs. The deviance information criterion (DIC) (Spiegelhalter *et al.* 2002) is a Bayesian measure of goodness-of-fit,

$$\text{DIC}=2\left\{\frac{1}{M}\sum_{h=1}^{M}D(\underset{\sim}{y},\Omega^{(h)})\right\}-D(\underset{\sim}{y},\hat{\Omega}),$$

where $\hat{\Omega}$ is a point estimate for Ω such as the mean of the posterior simulations, $\Omega^{(h)}$ are posterior simulations and $D(\underset{\sim}{y},\hat{\Omega})=-\log f(y|\Omega)$. DIC has been suggested as a criterion of model fit when the goal is to pick a model with best out-of-sample predictive power. A smaller value of DIC indicates a better fit and it provides reasonable assessments of model fit while considering the model complexity.

Similar to the DIC, LPML is also based on the same cross-validation (leave-one-out) procedure. A summary statistic of the conditional predictive ordinate (CPO) values is LPML. CPO is defined as the predictive density of observation i given all the other observa-

tions, that is, $\text{CPO}_i = p(y_i|y_{(i)}) = \int p(y_i|\Omega)p(\Omega|y_{(i)})d\Omega$, where $y_{(i)}$ is the data y without i^{th} observation. If observations are conditionally independent, a harmonic mean approximation of CPO is $\widehat{\text{CPO}_i} = \left\{\frac{1}{M}\sum_{h=1}^{M}\frac{1}{p(y_i|\Omega^{(h)})}\right\}^{-1}$, where $\Omega^{(h)}, h = 1, \ldots, M$ are samples from the posterior distribution. Then,

$$\text{LPML} = \sum_i \log(\widehat{\text{CPO}_i}).$$

Larger values of LPML indicate better fitting models (Geisser and Eddy 1979).

The DICs of the heterogeneous and homogeneous wards models are 1,852.58 and 4,039.93 respectively. The LPML of heterogeneous and homogeneous wards models are -1,096.38 and -1,838.45 respectively. So the heterogeneous ward model is better than the homogeneous one. The DICs of heterogeneous and homogeneous household model are 1,329.35 and 1,927.68 respectively. The LPML of the heterogeneous and homogeneous area models are -1,056.01 and -1,272.85 respectively. So the heterogeneous household model is better than the homogeneous one. Overall, based on the DIC and LPML, the heterogeneous household model has the smallest DIC and the largest LPML. The household models have relatively small DIC and large LPML. The household models are better when fitting the NLSS II health data.

Table 7: Comparison of BPP and DIC among four models: heterogeneous household model (HES), heterogeneous wards model (HEA), homogeneous household model (HOS), homogeneous wards model (HOA) for NLSS II data

Model	BPP	DIC	LPML
HEA	0.415	1852.58	-1096.38
HES	0.475	1329.35	-1056.01
HOA	0.155	4039.93	-1838.45
HOS	0.280	1927.68	-1272.85

We are interested in the finite population proportions of four health conditions in the small areas. We use all four ordered probit models to predict the nonsampled households in the 102 sampled wards. Bayesian bootstraps are used to generate unknown household sizes and nonsampled covariates within sampled wards and the bootstrapping is done within wards. The 2001 Census could potentially provide these two pieces of information, but there is a mismatch between the households in the census and the NLSS II (a record linkage can be performed). We note, however, that there is linkage between the wards, but this information is not useful to household estimates. In this application, we know the total number of households and individuals in each sampled wards, and we have sampled household information. Therefore, we decide to use these information in the Bayesian bootstrap approach to generate nonsampled household sizes and corresponding nonsampled covariates within sampled wards.

Based on 1,000 samples of parameters from the joint posterior distribution, we get 1,000 values of $\bar{P}_{it}$; order these values and pick the 95% prediction interval to be $(\bar{P}_{it}^{(25)}, \bar{P}_{it}^{(975)})$, $t = 1, 2, 3, 4$, where the values are arranged in increasing order.

The health status proportions of the 102 sampled wards based on both sampled and non-sampled households ($\bar{P}_t,\ t = 1, 2, 3, 4$) under all four models are shown in Figure 2. The proportions in excellent health condition are similar among all four models. The estimates from the heterogeneous household model are slightly less than those from the other models. The proportions in good health condition from both household models are more than those from the small area models. The proportions of fair condition and poor condition from the area models are relatively similar. The proportions of fair condition from the household models are larger than the proportions of poor conditions, which is consistent with the observed data. The error bars are the 95% credible intervals of $\bar{P}_t$. We can notice that the 95% credible intervals of the estimates in the homogeneous wards model are widest among all four models. The 95% credible intervals in the heterogeneous model have relatively the narrowest among all four models.

We examine plots to further compare the predictive inference of the finite population proportions of the four health conditions between the heterogeneous ward model and the heterogeneous household model. Figure 3 shows the comparison of the finite population proportions of four health conditions in each household within the sampled wards respectively between two models. One of our interest is to provide estimates for sampled wards. We can get the finite population proportions of health status in each sampled wards by taking the average on those estimates for households in each ward. Figure 4 shows the comparison of the finite population proportions of the four health conditions in sampled wards respectively between the two models. We can see that the points do not fall reasonably well on the 45^o line, which indicates that everything being equal, the model with sub-area random effects can capture more information, the heterogeneity of different households (sub-areas).

4. Concluding remarks and future works

In this paper, we study several hierarchical Bayesian ordered probit models for polychotomous responses. The sub-area models can capture the heterogeneity among the sub-areas (households) within the small areas (wards) and borrow strength from the sub-areas to obtain more efficient estimators. A full Bayesian analysis is provided for each model and predictive inference of the finite population proportions of the small areas is conducted. We have demonstrated our application to health status data from NLSS II.

We discussed one posterior computation algorithm to avoid poor mixing problems that the Gibbs sampler may cause. NLSS II health data were used in order to examine the performance of two models. We have performed a Bayesian predictive inference for the finite population proportion of each health status in the sampled wards based on the sampled and non-sampled households. BPP and DIC are used to assess and compare our ordinal probit models. The sub-area models perform better than the small area models.

In the paper, we assume the samples are self-weighted. However, if the sample unit cannot represent the target population, survey weights should be used to adjust selection bias. In the future, incorporating survey weights into the models can be explored. The observed biased samples actually followed a weighted distribution instead of the original distribution that the random samples follow. In order to predict and make inference about the finite population, the surrogate sampling approach by Nandram (2007) can be used to predict the finite population proportions.

We focus on parametric statistical models in this paper. Nonparametric Bayesian models using the stick-breaking priors can be considered to robustify the inference by embedding parametric models in nonparametric models. Ishwaran and James (2001) discussed the Gibbs samplers that can be used to fit posteriors of Bayesian hierarchical models based on stick-breaking priors. They are more flexible and better than the stick breaking prior of the Dirichlet process.

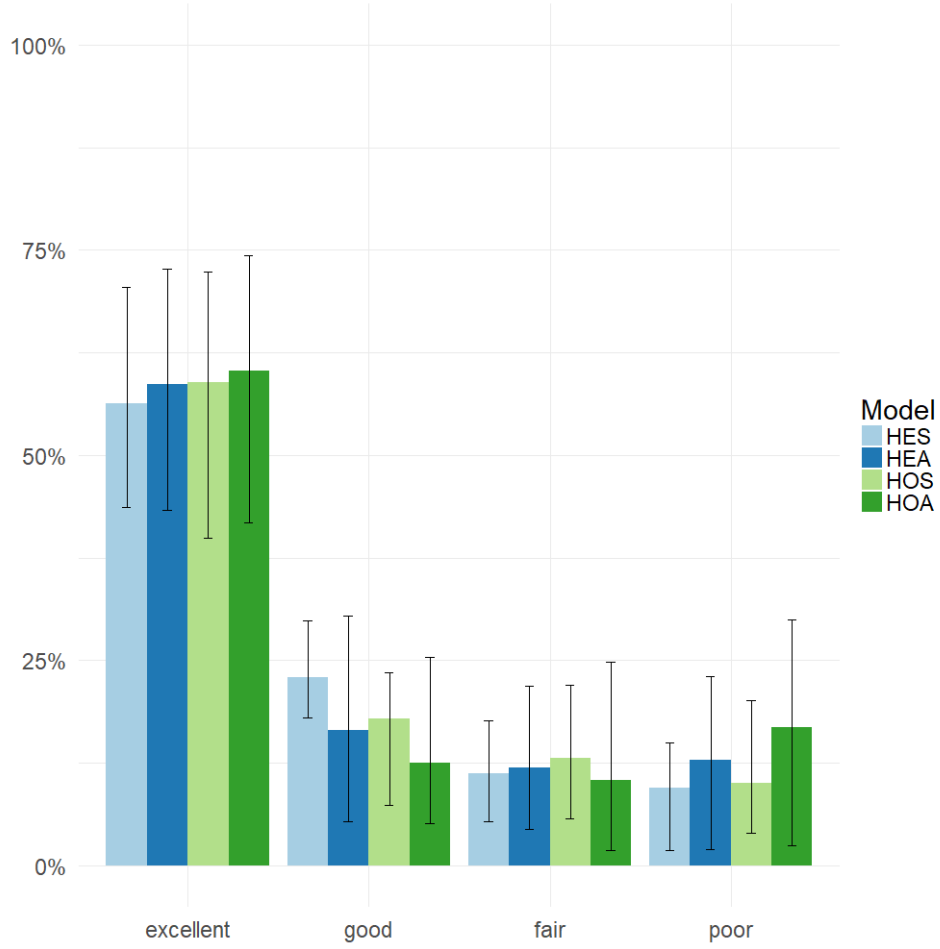

Figure 2: Comparison of finite population proportions of each health condition cell among four models

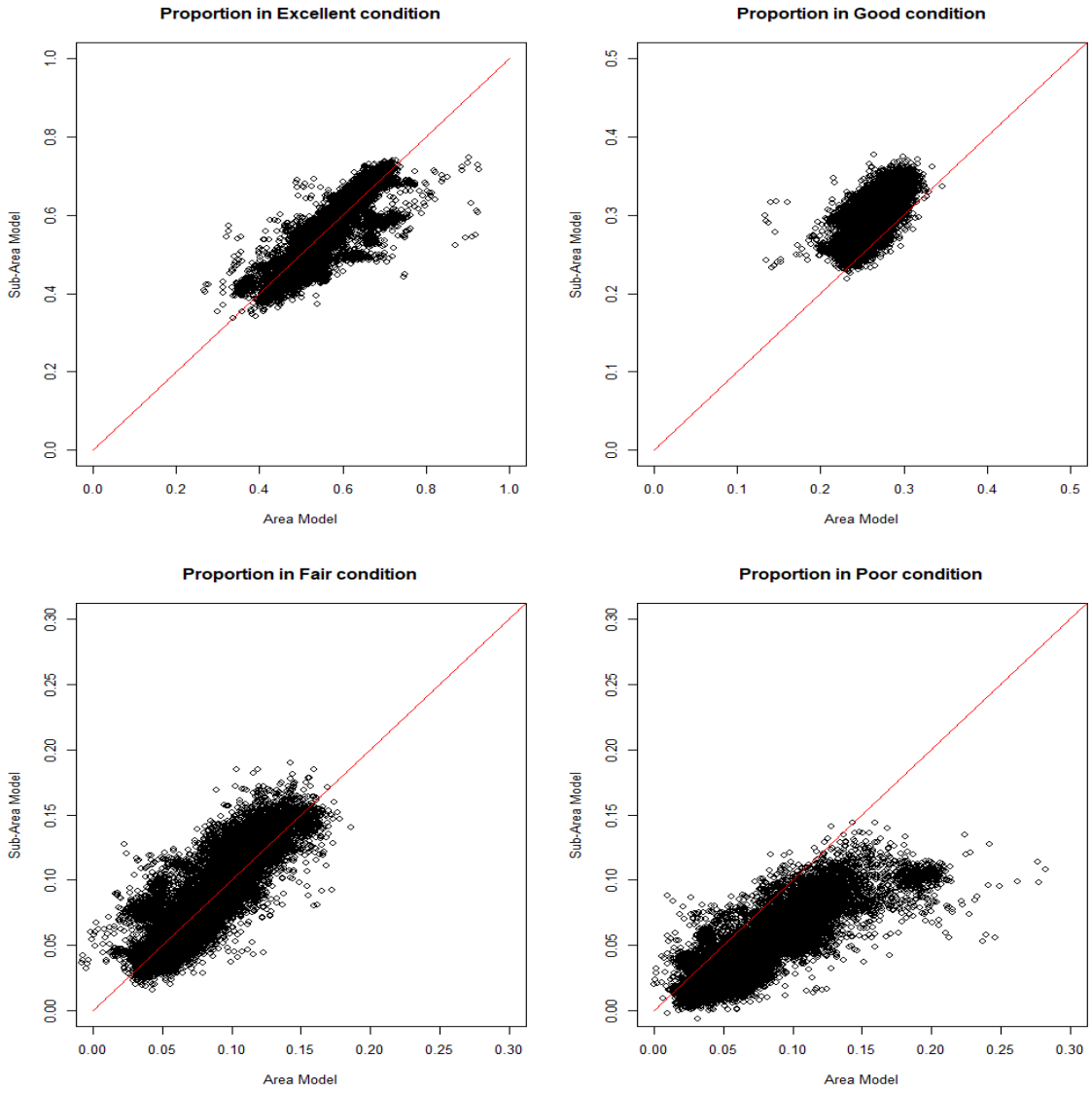

Figure 3: Comparison of the wards model and household models for prediction of the finite population proportions of 4 different health conditions of household

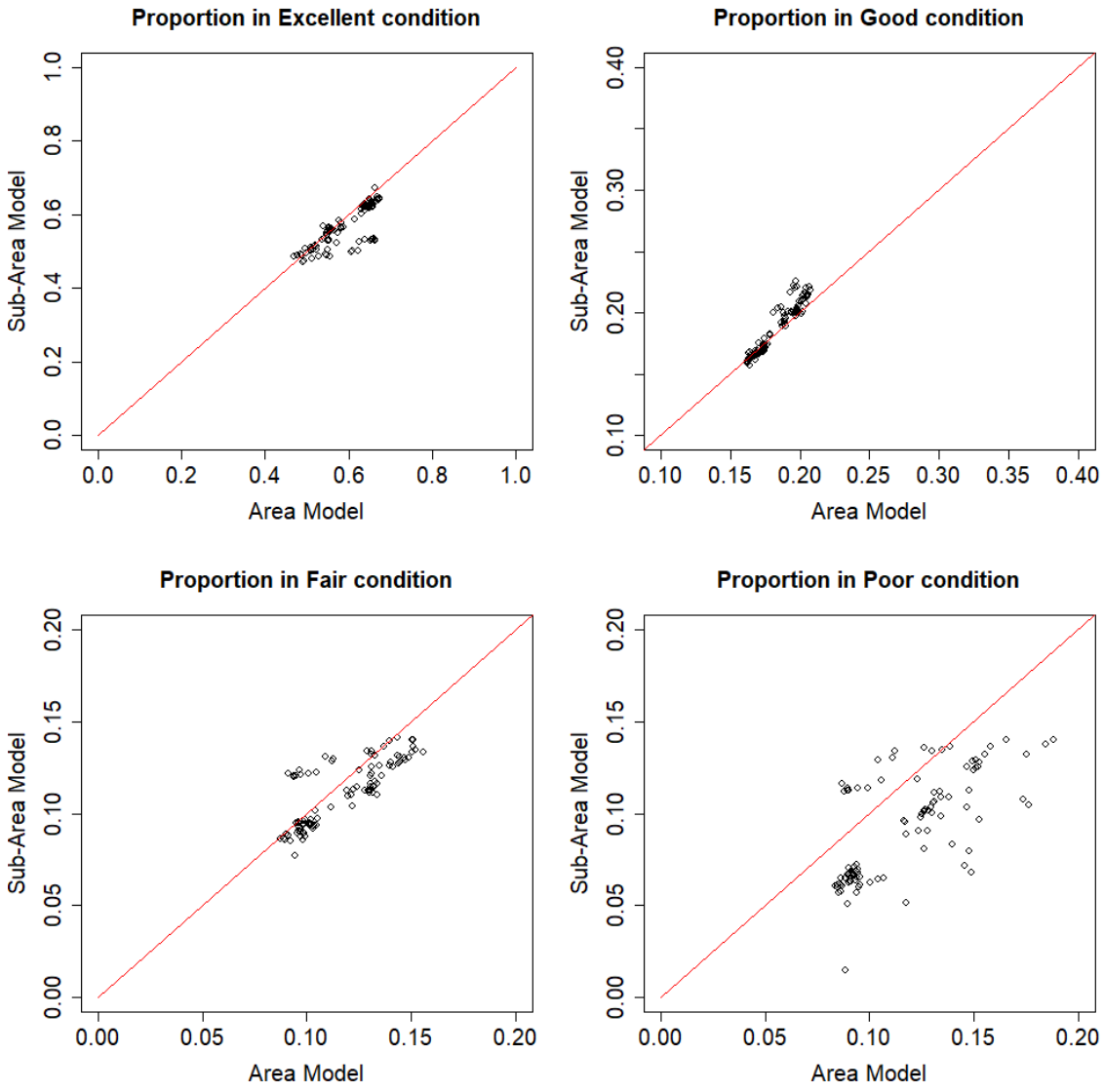

Figure 4: Comparison of the wards model and household models for prediction of the finite population proportions of 4 different health conditions of wards

Acknowledgements

We are very pleased and honored to be invited to write a paper for the special issue in memory of Professor C.R. Rao. We thank the reviewer for these informative comments.

References

Agresti, A. (2010). *Analysis of Ordinal Categorical Data.* John Wiley & Sons, Vol. **656**. DOI:10.1002/9780470594001.

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, **88**, 669-679. DOI: 10.2307/2290350.

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

Central Bureau of Statistics Thapathali. Nepal Living Standards Survey 2003/04. *Statistical Report*, Kathmandu, Nepal.

Chen, L. and Nandram, B. (2023). Bayesian logistic regression model for sub-areas. *Stats*, **6**, 209-231. DOI: 10.3390/stats6010013.

Chen, L. and Nandram, B. (2020). A hierarchical Bayesian Beta-Binomial model for sub-areas. In *Annual Conference of the Society of Statistics, Computer and Applications*, 23-40.

Chen, L., Nandram, B., and Cruze, N. B. (2022). Hierarchical Bayesian model with inequality constraints for US county estimates. *Journal of Official Statistics*, **38**, 709-732. DOI: 10.2478/jos-2022-0032.

Erciulescu, A. L., Cruze, N. B., and Nandram, B. (2019). Model-based county level crop estimates incorporating auxiliary sources of information. *Journal of the Royal Statistical Society Series A (Statistics in Society)*, **182**, 283-303. DOI: 10.1111/rssa.12390.

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. DOI: 10.1080/01621459.1979.10482505.

Fuller, W. A. and Goyeneche, J. J. (1998). Estimation of the state variance component. *unpublished*.

Geisser, S. (1980). Discussion on sampling and Bayes' inference in scientific modeling and robustness. *Journal of the Royal Statistical Society. Series A (General)*, 383-430.

Geisser, S. and Eddy, W. F. (1977). Predictive sample reuse approaches to model selection. *University of Minnesota*.

Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, **96**, 161-173. DOI: 10.1198/01621450 1750332758.

Holmes, C. C. and Held, L. (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis*, **1**, 145-168. DOI: 10.1214/06-BA105.

Lee, D., Nandram, B., and Kim, D. (2017). Bayesian predictive inference of a proportion under a two-fold small area model with heterogeneous correlations. *Survey Methodology*, **43**, 69-93.

Malec, D., Davis, W., and Cao, X. (1999). Model-based small area estimates of overweight prevalence using sample selection adjustment. *Statistics in Medicine*, **18**, 3189-3200. DOI: 10.1002/(sici)1097-0258(19991215)18:23<3189::aid-sim309>3.0.co;2-c.

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

McKelvey R.D. and Zavoina W. (1975). A statistical model for the analysis of ordinal level dependent variables. *The Journal of Mathematical Sociology*, **4**, 103-120. DOI: 10.1080/0022250X.1975.9989847.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological)*, **42**, 109-127. DOI: 10.1111/j.2517-6161.1980.tb01109.x.

Meng, X.L. (1994). Posterior predictive P-values. *The Annals of Statistics*, **22**, 1142-1160. DOI:10.1214/AOS/1176325622.

Nandram, B. (1989). Discrimination between complementary log-log and logistic model for ordinal data. *Communications in Statistics, Theory and Methods*, **18**, 2155-2164.

Nandram, B. (1998). A Bayesian analysis of the three-stage hierarchical multinomial model. *Journal of Statistical Computation and Simulation*, **61**, 97-126.

Nandram, B. (2000). Bayesian generalized linear models for inference about small areas. In *Generalized linear models: A Bayesian Perspective, Eds. D. K. Dey, S. K. Ghosh and B. K. Mallick, Marcel Dekker, Chapter 6*, 91-114.

Nandram, B. (2007). Bayesian predictive inference under information sampling via surrogate samples. *Bayesian Statistics and Its Applications*, 357-374.

Nandram, B. (2016). Bayesian predictive inference of a proportion under a two-fold small area model. *Journal of Official Statistics*, **32**, 187-207. DOI: 10.1515/jos-2016-0009.

Nandram, B., Chen, L., Fu, S-T., and Manandhar, B. (2018). Bayesian logistic regression for small areas with numerous households. *Statistics and Application*, **1**, 171-205. arXiv:1806.00446 [stat.ME].

Nandram, B., Chen, L., and Manandhar, B. (2018). Bayesian analysis of multinomial counts from small areas and sub-areas. *In JSM proceedings: Section on Survey Research Methods*, 1140–1162. Vancouver.

Nandram, B. and Erhardt, E. (2005). Fitting Bayesian two-stage generalized linear models using random samples via the SIR algorithm. *Sankhya*, **66**, 733-755. DOI: 10.2307/2505 3398.

Nandram, B., Cruze, N. B., and Erciulescu, A. L. (2023). Bayesian small area models under inequality constraints with benchmarking and double shrinkage. *Survey Methodology*, **49**, 517-546. http://www.statcan.gc.ca/pub/12-001-x/2023002/article/00004-eng.htm.

Nandram, B. and Sedransk, J. (1993). Bayesian predictive inference for a finite population proportion: two-stage cluster sampling. *Journal of the Royal Statistical Society. Series B (Methodological)*, 399-408. DOI:10.1111/J.2517-6161.1993.TB01910.X.

O'Connell, A. A. (2006). Logistic regression models for ordinal response variables. *Sage*, **146**, 140-153. DOI:10.4135/9781412984812.

Rubin, D. B. (1981). The Bayesian bootstrap. *The Annals of Statistics*, **9**, 130-134. DOI:10.1214/AOS/1176345338.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**, 583–639. DOI:10.1111/1467-9868.00353.

Winship, C. and Mare, R. D. (1984). Regression models with ordinal variables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 109-142. DOI:10.2307/2095465.

Yan, G. and Sedransk, J. (2007). Bayesian diagnostic techniques for detecting hierarchical structure. *Bayesian Analysis*, **2**, 735-760. DOI:10.1214/07-BA230.

Yan, G. and Sedransk, J. (2010). A note on Bayesian residuals as a hierarchical model diagnostic technique. *Statistical Papers*, **51**, 1-10. DOI:10.1007/S00362-007-0111-2.

APPENDIX

A. Computation method for the heterogeneous small area model

Using Bayes' theorem, the joint posterior distribution of the heterogeneous small area model in Section 2.1 is

$$\begin{aligned}\pi(\underset{\sim}{z},\nu,\underset{\sim}{\beta},\underset{\sim}{\lambda},\underset{\sim}{\theta},a,\delta^2|\underset{\sim}{y}) \propto & \prod_{i=1}^{\ell}\prod_{j=1}^{m_i}\left\{\sqrt{\lambda_i}e^{-\frac{\lambda_i}{2}(z_{ij}-\underset{\sim}{x}_{ij}^T\underset{\sim}{\beta}-\nu_i)^2}\sum_{t=1}^{T}[I(y_{ij}=t,\theta_{t-1}<z_{ij}\le\theta_t)]\right\}\\ & \times\left(\frac{1}{\delta^2}\right)^{\frac{\ell}{2}}\prod_{i=1}^{\ell}\left\{e^{-\frac{1}{2\delta^2}\nu_i^2}\right\}\times\exp\left\{-(\underset{\sim}{\beta}-\beta_0)^T(1000\Sigma_0)^{-1}(\underset{\sim}{\beta}-\beta_0)\right\}\\ & \times\left\{\prod_{i=1}^{\ell}\frac{a^a\lambda_i^{a-1}e^{-a\lambda_i}}{\Gamma(a)}\right\}\frac{1}{(1+a)^2}\frac{1}{(1+\delta^2)^2}.\end{aligned}$$

In order to jointly draw samples of $\underset{\sim}{z}$ and $\underset{\sim}{\lambda}$, we integrate out $\underset{\sim}{z}$ from the joint posterior distribution $\pi(\underset{\sim}{z},\underset{\sim}{\lambda}|\underset{\sim}{\nu},\underset{\sim}{x},\underset{\sim}{\beta},\underset{\sim}{\theta},a,\underset{\sim}{y})$. That is,

$$\begin{aligned}\pi(\lambda_i|\underset{\sim}{\nu},\underset{\sim}{x},\underset{\sim}{\beta},\underset{\sim}{\theta},a,\underset{\sim}{y}) &= \int\pi(z,\lambda_i|\underset{\sim}{\nu},\underset{\sim}{x},\underset{\sim}{\beta},\underset{\sim}{\theta},a,\underset{\sim}{y})d\underset{\sim}{z}\\ &\propto\prod_{j=1}^{m_i}\left\{\int\sqrt{\lambda_i}e^{-\frac{\lambda_i}{2}(z_{ij}-\underset{\sim}{x}_{ij}^T\underset{\sim}{\beta}-\nu_i)^2}\sum_{t=1}^{T}[I(y_{ij}=t,\theta_{t-1}<z_{ij}\le\theta_t)]\,d\underset{\sim}{z}\right\}\frac{a^a\lambda_i^{a-1}e^{-a\lambda_i}}{\Gamma(a)}\\ &=\prod_{j=1}^{m_i}\left\{\sum_{t=1}^{T}\int_{\theta_{t-1}}^{\theta_t}\left[\sqrt{\lambda_i}e^{-\frac{\lambda_i}{2}(z_{ij}-\underset{\sim}{x}_{ij}^T\underset{\sim}{\beta}-\nu_i)^2}\right]I(y_{ij}=t)dz_{ij}\right\}\frac{a^a\lambda_i^{a-1}e^{-a\lambda_i}}{\Gamma(a)}\\ &=\prod_{j=1}^{m_i}\left\{\sum_{t=1}^{T}\left[\Phi\left(\sqrt{\lambda_i}(\theta_t-\underset{\sim}{x}_{ij}^T\underset{\sim}{\beta}-\nu_i)\right)-\Phi\left(\sqrt{\lambda_i}(\theta_{t-1}-\underset{\sim}{x}_{ij}^T\underset{\sim}{\beta}-\nu_i)\right)\right]I(y_{ij}=t)\right\}\\ &\quad\times\frac{a^a\lambda_i^{a-1}e^{-a\lambda_i}}{\Gamma(a)}.\end{aligned}$$

Then Metropolis-Hastings algorithm is used to draw samples of λ_i from the marginal conditional distribution. Given λ_i and samples of $\underset{\sim}{\beta},\nu_i$ and $\underset{\sim}{\theta}$, we draw z_{ij} from truncated normal $\mathrm{N}(\underset{\sim}{x}_{ij}^T\underset{\sim}{\beta}+\nu_i,\lambda_i^{-1})$, where $y_{ij}=t$ if $\theta_{t-1}<z_{ij}\le\theta_t$.

To implement the Gibbs sampler once we get a sample of $\underset{\sim}{\lambda}$ and $\underset{\sim}{z}$, we need to draw samples from the full conditional posterior distributions of $\underset{\sim}{\nu}$, $\underset{\sim}{\beta}$, a, δ^2 and $\underset{\sim}{\theta}$.

First, the conditional distribution of $\underset{\sim}{\nu}$ is

$$\nu_i|\lambda_i,\underset{\sim}{z},\underset{\sim}{\beta},\delta^2,\underset{\sim}{y}\overset{ind}{\sim}\mathrm{Normal}\left(\frac{\lambda_i\sum_{j=1}^{n_i}(z_{ij}-\underset{\sim}{x}_{ij}^T\underset{\sim}{\beta})}{\frac{1}{\delta^2}+n_i\lambda_i},(\frac{1}{\delta^2}+n_i\lambda_i)^{-1}\right).$$

Second, the conditional distribution of $\underset{\sim}{\beta}$ is $\underset{\sim}{\beta}|\underset{\sim}{\nu},\underset{\sim}{\lambda},\underset{\sim}{z},\underset{\sim}{x},\underset{\sim}{y} \sim \text{MN}\left(\underset{\sim}{\hat{\beta}},\Sigma_{\underset{\sim}{\hat{\beta}}}\right)$, where

$$\underset{\sim}{\hat{\beta}} = \Sigma_{\underset{\sim}{\hat{\beta}}}\left(\sum_{i=1}^{\ell}\sum_{j=1}^{m_j}\lambda_i(z_{ij}-\nu_i)\underset{\sim}{x}_{ij} + (1000\Sigma_0)^{-1}\underset{\sim}{\beta}_0\right),$$

$$\Sigma_{\underset{\sim}{\hat{\beta}}} = \left(\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\lambda_i z_{ij}\underset{\sim}{x}_{ij}\underset{\sim}{x}_{ij}^T + (1000\Sigma_0)^{-1}\right)^{-1}.$$

Third, the fully conditional distribution of θ_t, given $\underset{\sim}{z}$, $\underset{\sim}{\theta}_{(t)} = \{\theta_s, s \neq t\}$ and data, is given by

$$\pi(\theta_t|\underset{\sim}{z},\underset{\sim}{\theta}_{(t)},\underset{\sim}{y}) \propto \prod_{i=1}^{\ell}\prod_{j=1}^{m_i}\left[I(y_{ij}=t,\theta_{t-1}<z_{ij}\leq\theta_t) + I(y_{ij}=t+1,\theta_t<z_{ij}\leq\theta_{t+1})\right].$$

Notice that this conditional density is uniform density on the interval. That is

$$\theta_t|\underset{\sim}{z},\underset{\sim}{\theta}_{(t)},\underset{\sim}{y} \sim \text{Uniform}\Big(\max\{\max\{z_{ij},y_{ij}=t\},\theta_{t-1}\},\min\{\min\{z_{ij},y_{ij}=t+1\},\theta_{t+1}\}\Big).$$

Fourth, given the sample of $\underset{\sim}{\lambda}$, we can use grid method to draw a. Transform a to $\phi_1 = \frac{a}{1+a}$, which is in $(0,1)$. The conditional posterior distribution of ϕ_1 is

$$\pi(\phi_1|\underset{\sim}{\lambda}) \propto \left(\prod_{i=1}^{\ell}\frac{a^a\lambda_i^{a-1}e^{-a\lambda_i}}{\Gamma(a)}\right)_{\phi_1=\frac{a}{1+a}}.$$

Fifth, to draw δ^2 we also use the grid method. Transform δ^2 to $\phi_2 = \frac{\delta^2}{1+\delta^2}$, which is in $(0,1)$. The conditional posterior distribution of ϕ_2 is

$$\pi(\phi_2|\underset{\sim}{\nu}) \propto \left\{\left(\frac{1}{\delta^2}\right)^{\frac{\ell}{2}}\exp\left(-\frac{1}{2\delta^2}\sum_{i=1}^{\ell}\nu_i^2\right)\right\}|_{\phi_2=\frac{\delta^2}{1+\delta^2}}.$$

To implement the algorithm, we chose starting points $\underset{\sim}{\beta}^{(0)},\underset{\sim}{\theta}^{(0)}$ equal to the maximum likelihood estimators (MLE) based on the previous paper by Chen and Nandram (2023), $\lambda_i^{(0)} = 1$ and $\nu_i^{(0)} = 1$. We first draw a and δ^2 using grid method, and then jointly draw a sample of $\underset{\sim}{\lambda}$ and $\underset{\sim}{z}$, and simulate from the conditional distribution of ν_i, $\underset{\sim}{\beta}$ and θ_t.

B. Computation method for the heterogeneous sub-area model

Using Bayes' theorem, the joint posterior distribution of the Heterogeneous Sub-Area Model in Section 2.2 is

$$\begin{aligned}
&\pi(\underset{\sim}{z},\underset{\sim}{\nu},\underset{\sim}{\mu},\underset{\sim}{\beta},\underset{\sim}{\lambda},a,\sigma^2,\delta^2|\underset{\sim}{y}) \\
&\quad\propto \prod_{i=1}^{\ell}\prod_{j=1}^{n_i}\prod_{k=1}^{m_{ij}}\left\{\sqrt{\lambda_i}e^{-\frac{\lambda_i}{2}(z_{ijk}-\underset{\sim}{x}_{ijk}^T\underset{\sim}{\beta}-\nu_i-\mu_{ij})^2}\sum_{t=1}^{T}\left[I(y_{ijk}=t,\theta_{t-1}<z_{ijk}\le\theta_t)\right]\right\} \\
&\quad\times\left(\frac{1}{\sigma^2}\right)^{\frac{\ell}{2}}\prod_{i=1}^{\ell}\prod_{j=1}^{n_i}\left\{e^{-\frac{1}{2\sigma^2}\mu_{ij}^2}\right\}\left(\frac{1}{\delta^2}\right)^{\frac{\ell}{2}}\prod_{i=1}^{\ell}\left\{e^{-\frac{1}{2\delta^2}\nu_i^2}\right\} \\
&\quad\times\exp\left\{-(\underset{\sim}{\beta}-\beta_0)^T(1000\Sigma_0)^{-1}(\underset{\sim}{\beta}-\beta_0)\right\}\left\{\prod_{i=1}^{\ell}\frac{a^a\lambda_i^{a-1}e^{-a\lambda_i}}{\Gamma(a)}\right\} \\
&\quad\times\frac{1}{(1+a)^2}\frac{1}{(1+\delta^2)^2}\frac{1}{(1+\sigma^2)^2}.
\end{aligned}$$

The method to fit the sub-area probit model is discussed in the following steps. In order to jointly draw samples of $\underset{\sim}{z}$ and $\underset{\sim}{\lambda}$, we integrate out $\underset{\sim}{z}$ from the joint posterior distribution $\pi(\underset{\sim}{z},\underset{\sim}{\lambda}|\underset{\sim}{\mu},\underset{\sim}{\nu},\underset{\sim}{x}.\underset{\sim}{\beta},a,\underset{\sim}{y})$. That is,

$$\begin{aligned}
&\pi(\lambda_i|\underset{\sim}{\nu},\underset{\sim}{x}.\underset{\sim}{\beta},a,\underset{\sim}{y})=\int\pi(\lambda_i|\underset{\sim}{\mu},\underset{\sim}{\nu},\underset{\sim}{x},\underset{\sim}{\beta},a,\underset{\sim}{y})d\underset{\sim}{z} \\
&\propto\prod_{j=1}^{n_i}\prod_{k=1}^{m_{ij}}\left\{\int\sqrt{\lambda_i}e^{-\frac{\lambda_i}{2}(z_{ijk}-\underset{\sim}{x}_{ijk}^T\underset{\sim}{\beta}-\nu_i-\mu_{ij})^2}\sum_{t=1}^{T}[I(y_{ijk}=t,\theta_{t-1}<z_{ijk}\le\theta_t)]\,d\underset{\sim}{z}\right\}\frac{a^a\lambda_i^{a-1}e^{-a\lambda_i}}{\Gamma(a)} \\
&=\prod_{j=1}^{n_i}\prod_{k=1}^{m_{ij}}\left\{\sum_{t=1}^{T}\int_{\theta_{t-1}}^{\theta_t}\left[\sqrt{\lambda_i}e^{-\frac{\lambda_i}{2}(z_{ijk}-\underset{\sim}{x}_{ijk}^T\underset{\sim}{\beta}-\nu_i-\mu_{ij})^2}\right]I(y_{ijk}=t)dz_{ijk}\right\}\frac{a^a\lambda_i^{a-1}e^{-a\lambda_i}}{\Gamma(a)} \\
&=\prod_{j=1}^{n_i}\prod_{k=1}^{m_{ij}}\left\{\sum_{t=1}^{T}\left[\Phi\left(\sqrt{\lambda_i}(\theta_t-\underset{\sim}{x}_{ijk}^T\underset{\sim}{\beta}-\nu_i-\mu_{ij})\right)-\Phi\left(\sqrt{\lambda_i}(\theta_{t-1}-\underset{\sim}{x}_{ijk}^T\underset{\sim}{\beta}-\nu_i-\mu_{ij})\right)\right]I(y_{ijk}=t)\right\} \\
&\qquad\times\frac{a^a\lambda_i^{a-1}e^{-a\lambda_i}}{\Gamma(a)}.
\end{aligned}$$

Then we can use accept-reject algorithm to draw samples of $\lambda_i,\ i=1,\ldots,\ell$. Once we get the sample, we can draw z_{ijk}. Similarly, we first draw a sample $\underset{\sim}{\beta}^*$ from prior MN$\left(\underset{\sim}{\beta}_0,1000\Sigma_0\right)$, draw a sample ν_i^* from prior Normal$(0,\delta^2)$ and draw a sample μ_{ij}^* from prior Normal$(0,\sigma^2)$ and given $\underset{\sim}{\beta}^*,\underset{\sim}{\mu}^*,\underset{\sim}{\nu}^*,\underset{\sim}{\lambda}$ and data, we can draw sample z_{ijk} from truncated Normal$(\underset{\sim}{x}_{ijk}^T\underset{\sim}{\beta}^*+\nu_i^*+\mu_{ij}^*,\lambda_i^{-1})$, where $\theta_{t-1}<z_{ijk}\le\theta_t$ if $y_{ijk}=t,\ t=1,\ldots,T,\ i=1,\ldots,\ell,\ j=1,\ldots,n_i,\ k=1,\ldots,m_{ij}$.

To implement the Gibbs sampler once we get a sample of $\underset{\sim}{\lambda}$ and $\underset{\sim}{z}$, we need to draw samples from the full conditional posterior distributions of $\underset{\sim}{\mu}$, $\underset{\sim}{\nu}$, $\underset{\sim}{\beta}$, a, σ^2, δ^2 and $\underset{\sim}{\theta}$.

First, the conditional distribution of $\underset{\sim}{\nu}$ is

$$\mu_{ij}|\nu_i,\lambda_i,\underset{\sim}{z},\underset{\sim}{\beta},\sigma^2,\underset{\sim}{y}\overset{ind}{\sim}\text{Normal}\left(\frac{\lambda_i\sum_{k=1}^{m_{ij}}(z_{ijk}-\underset{\sim}{x}_{ijk}^T\underset{\sim}{\beta}-\nu_i)}{\frac{1}{\sigma^2}+m_{ij}\lambda_i},(\frac{1}{\sigma^2}+m_{ij}\lambda_i)^{-1}\right).$$

Second, the conditional distribution of $\underset{\sim}{\nu}$ is

$$\nu_i|\underset{\sim}{\mu},\lambda_i,\underset{\sim}{z},\underset{\sim}{\beta},\delta^2,\underset{\sim}{y} \overset{ind}{\sim} \text{Normal}\left(\frac{\lambda_i\sum_{j=1}^{n_i}\sum_{k=1}^{m_{ij}}(z_{ijk}-\underset{\sim}{x}_{ijk}^T\underset{\sim}{\beta}-\mu_{ij})}{\frac{1}{\delta^2}+\sum_{j=1}^{n_i}m_{ij}\lambda_i},(\frac{1}{\delta^2}+\sum_{j=1}^{n_i}m_{ij}\lambda_i)^{-1}\right).$$

Third, the conditional distribution of $\underset{\sim}{\beta}$ is $\underset{\sim}{\beta}|\underset{\sim}{\mu},\underset{\sim}{\nu},\underset{\sim}{\lambda},\underset{\sim}{z},\underset{\sim}{x},\underset{\sim}{y} \sim \text{MN}\left(\underset{\sim}{\hat{\beta}},\Sigma_{\underset{\sim}{\hat{\beta}}}\right)$, where

$$\underset{\sim}{\hat{\beta}} = \Sigma_{\underset{\sim}{\hat{\beta}}}\left(\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\sum_{k=1}^{m_{ij}}\lambda_i(z_{ijk}-\nu_i-\mu_{ij})\underset{\sim}{x}_{ijk}+(1000\Sigma_0)^{-1}\underset{\sim}{\beta}_0\right),$$

$$\Sigma_{\underset{\sim}{\hat{\beta}}} = \left(\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\sum_{k=1}^{m_{ij}}\lambda_i\underset{\sim}{x}_{ijk}\underset{\sim}{x}_{ijk}^T+(1000\Sigma_0)^{-1}\right)^{-1}.$$

Fourth, the fully conditional distribution of θ_t given $\underset{\sim}{z}$, $\underset{\sim}{\theta}_{(t)}=\{\theta_s, s\neq t\}$ and data is given by

$$\pi(\theta_t|\underset{\sim}{z},\underset{\sim}{\theta}_{(t)},\underset{\sim}{y}) \propto \prod_{i=1}^{\ell}\prod_{j=1}^{m_i}\left[I(y_{ij}=t,\theta_{t-1}<z_{ij}\leq\theta_t)+I(y_{ij}=t+1,\theta_t<z_{ij}\leq\theta_{t+1})\right].$$

Notice that this conditional density is uniform density on the interval. That is

$$\theta_t|\underset{\sim}{z},\underset{\sim}{\theta}_{(t)},\underset{\sim}{y} \sim \text{Uniform}\left[\max\left\{\max\left\{z_{ij},y_{ij}=t\right\},\theta_{t-1}\right\},\min\left\{\min\left\{z_{ij},y_{ij}=t+1\right\},\theta_{t+1}\right\}\right].$$

Fifth, given the sample of $\underset{\sim}{\lambda}$, we can use grid method to draw a. Transform a to $\phi_1=\frac{a}{1+a}$, which is in $(0,1)$. The conditional posterior distribution of ϕ_1 is

$$\pi(\phi_1|\underset{\sim}{\lambda}) \propto \left(\prod_{i=1}^{\ell}\frac{a^a\lambda_i^{a-1}e^{-a\lambda_i}}{\Gamma(a)}\right)_{\phi_1=\frac{a}{1+a}}.$$

Sixth, to draw δ^2 we also use grid method. Transform δ^2 to $\phi_2=\frac{\delta^2}{1+\delta^2}$, which is in $(0,1)$. The conditional posterior distribution of ϕ_2 is

$$\pi(\phi_2|\underset{\sim}{\nu}) \propto \left\{\left(\frac{1}{\delta^2}\right)^{\frac{\ell}{2}}\exp\left(-\frac{1}{2\delta^2}\sum_{i=1}^{\ell}\nu_i^2\right)\right\}_{\phi_2=\frac{\delta^2}{1+\delta^2}}.$$

Seventh, to draw σ^2 we also use grid method. Transform σ^2 to $\phi_3=\frac{\sigma^2}{1+\sigma^2}$, which is in $(0,1)$. The conditional posterior distribution of ϕ_2 is

$$\pi(\phi_2|\underset{\sim}{\mu}) \propto \left\{\left(\frac{1}{\sigma^2}\right)^{\frac{\sum_{i=1}^{\ell}n_i}{2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\mu_{ij}^2\right)\right\}_{\phi_3=\frac{\sigma^2}{1+\sigma^2}}.$$

To implement the algorithm, we chose start points $\underset{\sim}{\beta}^{(0)},\underset{\sim}{\theta}^{(0)}$ equal to the MLE based on the

previous paper by Chen and Nandram (2023), $\lambda_i^{(0)} = 1,\ \mu^0(0)_{ij} = 1,\ \nu^0(0)_i = 1$. We first draw a, δ^2, and σ^2 using the grid method, and then jointly draw a sample of $\underset{\sim}{\lambda}$ and $\underset{\sim}{z}$, and simulate from the conditional distribution of $\mu_{ij},\ \nu_i\ ,\underset{\sim}{\beta}$ and θ_t.

C. Bayesian bootstrap

Our interest is to predict the finite population proportions of 102 sampled wards for all households. The covariates of individuals in non-sampled households and the size of non-sampled households are unknown. We know the total number of households and individuals in each sampled ward and we have all information about the sampled households. Therefore, we decide to use these information in the Bayesian bootstrap approach to generate the non-sampled household sizes and corresponding non-sampled covariates within sampled wards. The Bayesian bootstrap (Rubin 1981) method is used sample the sampled households to impute the non-sampled households. There are $n = 12$ sampled households in the sampled wards and everyone is sampled from the sampled households. We know the sizes and covariates of all sampled households, and we simple need to have the sample sizes and the covariates for all the non-sampled households in any sampled ward to do Bayesian predictive inference in each sampled ward; the procedure is done independently for each sampled ward.

Let N denote the number of households in one of the sampled wards. We simply need to fill in the sizes of the households and their covariates. This procedure is equivalent to simply sampling the households. Denote the labels of the sampled households by $1,\ldots,n$ to provide the information (sizes and covariates) of the non-sampled households with labels, $n+1,\ldots,N$. Denote the sampled indicators of each household by $I_i,\ i = 1,\ldots,n$. After the bootstrap is executed, because it is based on sampling with replacement, there will be N_i^* non-sampled households corresponding to the i^{th} sampled household, and $\sum_{i=1}^n N_i^* = N - n$.

The Bayesian bootstrap assumes that

$$\underset{\sim}{I} \mid \underset{\sim}{p} \sim \text{Multinomial}(n, \underset{\sim}{p}),$$

where we actually observed $I_i = 1, i = 1,\ldots,n$,

$$\underset{\sim}{p} \sim \text{Dirichlet}(Q),$$

Haldane's improper prior, where Q is a vector of zeros. Then, a posterior

$$\underset{\sim}{p} \mid \underset{\sim}{I} \sim \text{Dirichlet}(\underset{\sim}{j}), \tag{5}$$

where $\underset{\sim}{j}$ is a vector of ones. Therefore,

$$\underset{\sim}{N}^* \mid \underset{\sim}{p}, \underset{\sim}{I} \sim \text{Multinomial}(N^*, \underset{\sim}{p}). \tag{6}$$

To execute the bootstrap, simply draw $\underset{\sim}{p}$ from (5) and $\underset{\sim}{N}^*$ from (6).