STATISTICS IN TRANSITION new series, August 2020
Special Issue, pp. 249–265, DOI 10.21307/stattrans-2020-040
Submitted – 31.01.2020; paper accepted for published – 30.06.2020

A general Bayesian approach to meet different inferential goals in poverty research for small areas

Partha Lahiri[1], **Jiraphan Suntornchost**[2]

ABSTRACT

Poverty mapping that displays spatial distribution of various poverty indices is most useful to policymakers and researchers when they are disaggregated into small geographic units, such as cities, municipalities or other administrative partitions of a country. Typically, national household surveys that contain welfare variables such as income and expenditures provide limited or no data for small areas. It is well-known that while direct survey-weighted estimates are quite reliable for national or large geographical areas they are unreliable for small geographic areas. If the objective is to find areas with extreme poverty, these direct estimates will often select small areas due to the high variability in the estimates. Empirical best prediction and Bayesian methods have been proposed to improve on the direct point estimates. These estimates are, however, not appropriate for different inferential purposes. For example, for identifying areas with extreme poverty, these estimates would often select areas with large sample sizes. In this paper, using real life data, we illustrate how appropriate Bayesian methodology can be developed to address different inferential problems.

Key words: Bayesian model, cross-validation, hierarchical models, Monte Carlo simulations

1. Introduction

Eradication of poverty, one of the greatest challenges facing humanity, has been the central tool to guide various public policy efforts in many countries. According to the United Nations [3]: "Extreme poverty rates have fallen by more than half since 1990. While this is a remarkable achievement, one-in-five people in developing regions still live on less than $1.90 a day. Millions more make little more than this daily amount and are at risk of slipping back into extreme". On September 25, 2015, the United Nations adopted the 2030 Agenda for Sustainable Development with 17 new Sustainable Development Goals (SDGs), beginning with a historical pledge to end poverty in all forms and dimensions by 2030 everywhere permanently.[4] In order to achieve these goals, basic resources and services need to be more accessible to people living in vulnerable situations. Moreover, support for communities affected by conflict and climate related disasters needs to be raised.

[1]Department of Mathematics & Joint Program in Survey Methodology, University of Maryland, College Park, USA. E-mail: plahiri@umd.edu. ORCID: https://orcid.org/0000-0002-7103-545X.
[2]Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University, Thailand. E-mail: jiraphan.s@chula.ac.th. ORCID: https://orcid.org/0000-0001-5410-9659.
[3]see https://sdg-tracker.org/no-poverty
[4]see https://www.un.org/sustainabledevelopment/

National estimate of an indicator usually hides important differences among different regions or areas with respect to that indicator. In almost all countries, these differences exist and can often be substantial. The smaller the geographic regions for which indicators are available, the greater the effectiveness of interventions. Indeed this allows to reduce transfers to the non-poor and minimizes the risk that a poor person will be missed by the program. Ravallion (1994) found that Indian and Indonesian states or provinces are too heterogeneous for targeting to be effective. This underlines the need for production of estimates of indicators for small areas that are relatively homogenous.

It is now widely accepted that direct estimates of poverty based on household survey data are unreliable. There are now several papers available in the literature that attempt to improve on the direct estimates by borrowing strength from multiple relevant databases. Hierarchical models that combine information from different databases are commonly used to achieve the goal because such models not only provide improved point estimates but also incorporate different sources of variabilities. These models can be implemented using a synthetic approach (e.g., Elbers et al. 2003), empirical best prediction approach (Fay and Herriot 1979; Franco and Bell 2015; Bell et al. 2016; Molina and Rao 2010; Casas-Cordero et al. 2016), and Bayesian approach (Molina et al. 2014). See Jiang and Lahiri (2006), Pfeffermann (2013) and Rao and Molina (2015) for a review of different small area estimation techniques.

Empirical best prediction and hierarchical Bayesian methods (also shrinkage methods) have been employed in numerous settings, including studies of cancer incidence in Scotland (Clayton and Kaldor 1987), cancer mortality in France (Mollie and Richardson 1991), stomach and bladder cancer mortality in Missouri cities (Tsutukawa et al. 1985), toxoplasmosis incidence in El Salvador (Efron and Morris 1975), infant mortality in New Zealand (Marshall 1991), mortality rates for chronic obstructive pulmonary disease (Nandram et al. 2000), poverty research (Molina and Rao 2010; Molina et al. 2014; Bell et al. 2016; Casas-Cordero et al. 2016). The basic approach in all these applications is the same: a prior distribution of rates is posited and is combined with the observed rates to calculate the posterior, or stabilized, rates.

All the papers cited in the previous paragraph deal with the point estimation and the associated measure of uncertainty. However, in many cases, there could be different inferential goals where point estimates, whether empirical best prediction estimates or posterior means, though they can provide a solution, are not efficient. For example, one inferential goal could be to flag a geographical area (e.g., municipality) for which the true poverty measure of interest exceeds a pre-specified standard. The point estimates can certainly flag such areas but do not provide any reasonable uncertainty measure to assess the quality of such action. It is not clear how to propose such a measure for a method based on direct estimates. For the method based on posterior mean, one can perhaps propose a normal approximation using the posterior mean and posterior standard deviation to approximate the posterior probability of the true poverty measure exceeding the pre-specified standard. In some cases, quality of such approximation could be questionable. One can have a more complex inferential goal. For example, we may be interested in identifying the worst geographical area with respect to the poverty measure. In this case, the use of direct estimates for identification can be misleading when the

sample sizes vary across the geographic units. The regions with small sample sizes will tend to have both high and low poverty indices merely because they have the largest variability. The method based on posterior means is not good either since it tends to identify areas with more samples; see Gelman and Price (1999). Moreover, in either case there does not seem to be a clear way to produce a reasonable quality measure.

Gelman and Price (1999), Morris and Christiansen (1996), Langford and Lewis (1998), Jones and Spigelhalter (2011) discussed various inferential problems other than the point estimation. For example, Morris and Christiansen (1996) outlined inferential procedures for identifying areas with extreme indicator. Our approach is similar to theirs but applied for different complex parameters and used much more complex hierarchical model that is appropriate for survey data.

We would like to stress that our proposed approach for solving different inferential problems is fundamentally different from the constrained empirical hierarchical Bayesian (Louis 1984; Ghosh 1992; Lahiri 1990) and the triple-goal (Shen and Louis, 1998) approaches where the goal is to produce one set of estimates for different purposes. In contrast, we propose to use the same synthetic data matrix generated from the posterior predictive distribution for different inferential purposes. Other than this fundamental difference, our approach has a straightforward natural way to produce appropriate quality measures. In poverty research, Molina et al. (2014) suggested an interesting approach for estimating different poverty indices by generating a synthetic population from a posterior predictive density. However, they restricted themselves to point estimates and their associated uncertainty measures and did not discuss how one would solve a variety of statistical inferences.

We outline a general approach to deal with different inferential goals and illustrate our methodology using the data used by the Chilean government for their small area poverty mapping system. Section 2 describes the Chilean data, the hierarchical model, a general poverty index, inferential approach for achieving various goals and data analysis. In Section 3, we provide some concluding remarks and a direction for future research.

2. Illustration of the proposed methodology using Chilean poverty data

There has been a consistent downward trend in the official poverty rate estimates, which are the usual national survey-weighted direct estimates, in Chile since the early 90's. While this national trend is encouraging, there is an erratic time series trend in the direct estimates for small comunas (municipalities) - the smallest territorial entity in Chile. Moreover, for a handful of extremely small comunas, survey estimates of poverty rates are unavailable for some or all time points simply because the survey design, which traditionally focuses on obtaining precise estimates for the nation and large geographical areas, excludes these comunas for some or all of the time points. In any case, direct survey estimates of poverty rates typically do not meet the desired precision for small comunas and thus the assessment of implemented policies is not straightforward at the comuna level. In order to successfully monitor trends, identify influential factors,

develop effective public policies and eradicate poverty at the comuna level, there is a growing need to improve on the methodology for estimating poverty rates at this level of geography. In this section, we use the Chilean case to illustrate our Bayesian approach to answer a variety of research questions.

2.1. Data used in the Analysis

To illustrate our Bayesian approach, we use a household survey data as the primary source of information and comuna level summary statistics obtained from different administrative data sources as supplementary sources of information. We now provide a brief description of the primary and supplementary databases. Further details can be obtained from Casas-Cordero et al. (2016).

2.1.1 The Primary Data Source: The CASEN 2009 Data

The Ministry of Social Development estimates the official poverty rates using the National Socioeconomic Characterization Survey, commonly referred to as the CASEN. The Ministry has been conducting CASEN since 1987 every two or three years. The CASEN is a household survey collecting a variety of information of Chilean households and persons, including information about income, work, health, subsidies, housing and others. The Ministry calculates poverty rate estimates at national, regional and municipality (comuna) levels. The Ministry is the authority specified by the Chilean law to deliver poverty estimates for all the 345 comunas in Chile. These estimates are used, along with other variables, to allocate public funding to municipalities.

In a joint effort by the Ministry and United Nations Development Programme (UNDP), a Small Area Estimation (SAE) official system was developed for estimating poverty rates at comuna level using the CASEN 2009 survey. The Chilean method is based on an empirical Bayesian method using an area level Fay-Herriot Model (Fay and Herriot 1979) to combine the CASEN survey data with a number of administrative databases. The SAE system provides point estimates and parametric bootstrap confidence intervals (see, e.g., Chatterjee et al. 2008; Li and Lahiri 2010) for the Chilean comunas.

The CASEN 2009 used a stratified multistage complex sample of approximately $75,000$ housing units from $4,156$ sample areas. The entire Chile was divided into a large number of sections (Primary Stage Units or PSUs). The PSUs were then grouped into strata on the basis of two geographic characteristics: comuna and urban/rural classification. Overall, there were 602 strata in the CASEN 2009 survey and multiple PSUs were sampled per stratum. The probability of selection for each PSU in a stratum was proportional to the number of housing units in the (most recently updated) 2002 Census file.

Prior to the second stage of sampling, listers were sent to the sampled PSUs to update the count of housing units. This procedure was implemented in both urban and rural areas. In the second stage of sampling, a sample of housing units was selected within the sampled PSUs. The probability of selection for each housing unit (Secondary Stage Unit or SSU) is the same within each PSU. On the average, 16-22 housing units

were selected within each PSU by implementing a procedure that used a random start and a systematic interval to select the units to be included in the sample.

2.1.2 Administrative Data at the Comuna Level

Casas-Cordero et al. (2016) carried out an extensive task to identify a set of auxiliary variables derived from different administrative records of different agencies. In this paper, we use the same set of comuna level auxiliary variables for illustrating our approach. For completeness, we list them below:

(1) Average wage of workers who are not self-employed,

(2) Average of the poverty rates from CASEN 2000, 2003, and 2006,

(3) Percentage of population in rural areas,

(4) Percentage of illiterate population,

(5) Percentage of population attending school.

Like in Casas-Cordero et al. (2016), we also use arcsine square-root transformation for all the auxiliary variables except the first one, for which we use logarithmic transformation. We note that our approach is general and can use a different set of auxiliary variables that may be deemed appropriate in the future.

2.2. The Foster-Greer-Thorbecke (FGT) poverty indices

Currently, the Chilean government publishes headcount ratios or poverty rates for the nation and its comunas. To present our approach in a general setting, we consider a general class of poverty indices commonly referred to as the FGT indices, after the names of the three authors of the paper by Foster et al. (1984). To describe the FGT index, we first introduce the following notations:

N_c: total number of households in comuna c,

U_c: number of urbanicity statuses for comuna c; since for urbanicity status, we use urban and rural statuses only, U_c is either 1 or 2 for a given comuna,

k_u: fixed poverty line for urban-rural classification u ($u = 1$ and $u = 2$ for urban and rural, respectively),

M_{cu}: total number of PSUs in the universe for urban-rural classification u of comuna c,

N_{cup}: total number of households in the universe of the PSU p belonging to the urban-rural classification u of comuna c,

y_{cuph}: per-capita income of household h (that is, total income of the household divided by the number of household members) in PSU p, urban-rural classification u, and comuna c.

In our context, the class of FGT indices for comuna c is given by

$$Q_{c;\alpha} = \frac{1}{N_c} \sum_{u=1}^{U_c} \sum_{p=1}^{M_{cu}} \sum_{h=1}^{N_{cup}} g_\alpha(y_{cuph}),$$

where

$g_\alpha(y_{cuph}) = \left(\frac{k_u - y_{cuph}}{k_u} \right)^\alpha I(y_{cuph} < k_u)$, α is a "sensitivity" parameter ($\alpha = 0, 1, 2$ corresponding to poverty ratio, poverty gap, and poverty severity, respectively).

2.3. Hierarchical Model

A hierarchical model could be effective in capturing different salient features of the CASEN survey data and in linking comuna level auxiliary variables derived from different administrative records. We consider the following working hierarchical model to illustrate our general approach for inference. We call the model a working model because we recognize that it is possible to improve on it in the future. But this model will suffice to illustrate the central theme of the paper, i.e., how to carry out a particular inferential procedure given a hierarchical model.

Let $T_{cuph} = T(y_{cuph})$ be a given transformation on the study variable y_{cuph}. For the application of this paper, we consider $T(y_{cuph}) = \ln(y_{cuph} + 1)$. We consider the following hierarchical model for the sampled units:

$$\begin{aligned}
\text{Level 1:} \quad & T_{cuph} | \theta_{cup}, \sigma_T \overset{ind}{\sim} N\left(\theta_{cup}, \sigma_T^2\right), \\
\text{Level 2:} \quad & \theta_{cup} | \mu_{cu}, \sigma_\theta \overset{ind}{\sim} N\left(\mu_{cu}, \sigma_\theta^2\right), \\
\text{Level 3:} \quad & \mu_{cu} | \beta_u, \sigma_\mu \overset{ind}{\sim} N\left(\mathbf{x}_c^T \beta_u, \sigma_\mu^2\right),
\end{aligned}$$

where $\mathbf{x}_c$ is a vector of comuna level known fixed auxiliary variables; θ_{cup} and μ_{cu} are random effects; $\beta_u, \sigma_T^2, \sigma_\theta^2$ and σ_μ^2 are unknown hyperparameters.

We follow the recommendation of Gelman (2015) in assuming weakly informative priors for the hyperparameters. For example, we assume independent $N(0,1)$ prior for all regression coefficients and independent half normal prior for the standard deviations.

2.4. Inferential Approach

We first note that the inference on $Q_{c,\alpha}$ is equivalent to that of

$$Q_{c;\alpha} = \frac{1}{N_c} \sum_{u=1}^{U_c} \sum_{p=1}^{M_{cu}} \sum_{h=1}^{N_{cup}} g_\alpha \left(T^{-1}(T_{cuph}) \right),$$

where T is a monotonic function (e.g., logarithm). Under full specification of the model for the finite population, one can make inferences about $Q_{c;\alpha}$ in a standard way. However, full specification of model for the unobserved units of the finite population seems to be a challenging task. To this end, appealing to the law of large numbers, we first approximate

$Q_{c;\alpha}$ by $\tilde{Q}^P_{c;\alpha}$, where

$$\tilde{Q}^P_{c;\alpha} = \frac{1}{N_c}\sum_{u=1}^{U_c}\sum_{p=1}^{M_{cu}}\sum_{h=1}^{N_{cup}} \mathbb{E}\left\{g_\alpha\left(T^{-1}(T_{cuph})\right)|\theta_{cup},\sigma_T\right\}.$$

This is reasonable under Level 1 of the hierarchical model (even without the normality assumption) since N_c is typically large. We then propose the following approximation to $\tilde{Q}^P_{c;\alpha}$.

$$\tilde{Q}_{c;\alpha} \equiv \tilde{Q}_{c;\alpha}(\theta_c,\sigma_T) = \sum_{u=1}^{U_c}\sum_{p=1}^{m_{cu}}\sum_{h=1}^{n_{cup}} w_{cuph}\mathbb{E}\left\{g_\alpha\left(T^{-1}(T_{cuph})\right)|\theta_{cup},\sigma_T\right\}, \tag{1}$$

where

w_{cuph} is the survey weight for the household h in the PSU p within urbanicity u of comuna c,

$\theta_c = \text{col}_{u,p}\theta_{cup}$; a $\sum_{u=1}^{U_c} m_{cu}\times 1$ column vector (we follow the notation of Prasad and Rao 1990),

$$g_\alpha\left(T^{-1}(T_{cuph})\right) = \left\{\frac{k_u - \left(T^{-1}(T_{cuph})\right)}{k_u}\right\}^\alpha I\left(T_{cuph}\le l_u\right),$$

$l_u = \ln(k_u+1)$, the poverty line of the urbanicity u in the transformed scale.

The weights are scaled within each comuna so that the sum of the weights for all households equals 1. In the last approximation, we assume that the scaled survey weight w_{cuph} represents proportion of units in the finite population (including the unit $cuph$) of comuna c that are similar to the unit $cuph$.

The calculations of (1) under the model described in Section 2.3 can be done through the following formula:

For $\alpha = 0$,

$$\begin{aligned}\mathbb{E}\left\{g_0(\left(T^{-1}(T_{cuph})\right)|\theta_{cup},\sigma_T^2\right\} &= \int_{-\frac{\theta_{cup}}{\sigma_T}}^{\frac{l_u-\theta_{cup}}{\sigma_T}} \phi(z|\theta_{cup},\sigma_T^2)dz \\ &= \Phi\Big(\frac{l_u-\theta_{cup}}{\sigma_T}\Big) - \Phi\Big(-\frac{\theta_{cup}}{\sigma_T}\Big),\end{aligned}$$

where ϕ and Φ are the density function and the distribution function of the standard normal distribution, respectively.

For $\alpha = 1$,

$$\begin{aligned}
&\mathbb{E}\{g_1(T^{-1}(T_{cuph}))|\theta_{cup},\sigma_T^2\} \\
&\quad = \mathbb{E}\left\{\left(\frac{\exp(l_u)-\exp(T_{cuph})}{k_u}\right) I\big(T_{cuph}\le l_u\big)\right\} \\
&\quad = \frac{1}{k_u}\int_{-\frac{\theta_{cup}}{\sigma_T}}^{\frac{l_u-\theta_{cup}}{\sigma_T}} (\exp(l_u)-\exp(\sigma_T z+\theta_{cup}))\,\phi(z|\theta_{cup},\sigma_T^2)dz \\
&\quad = \frac{\exp(l_u)}{k_u}\Big[\Phi\Big(\frac{l_u-\theta_{cup}}{\sigma_T}\Big)-\Phi\Big(-\frac{\theta_{cup}}{\sigma_T}\Big)\Big] \\
&\qquad - \frac{\exp\left(\theta_{cup}+\frac{\sigma_T^2}{2}\right)}{k_u}\left[\Phi\Big(\frac{l_u-\theta_{cup}-\sigma_T^2}{\sigma_T}\Big)-\Phi\Big(-\frac{\theta_{cup}+\sigma_T^2}{\sigma_T}\Big)\right],
\end{aligned}$$

where we use the fact that $\int_a^b \exp(\sigma z)\,\phi(z)dz = \exp\left(\frac{\sigma^2}{2}\right)\Big[\Phi(b-\sigma)-\Phi(a-\sigma)\Big]$ to obtain the last equation.

For $\alpha = 2$,

$$\begin{aligned}
&\mathbb{E}\{g_2(T^{-1}(T_{cuph}))|\theta_{cup},\sigma_T^2\} \\
&\quad = \mathbb{E}\left\{\left(\frac{\exp(l_u)-\exp(T_{cuph})}{k_u}\right)^2 I\big(T_{cuph}\le l_u\big)\right\} \\
&\quad = \frac{1}{k_u^2}\int_{-\frac{\theta_{cup}}{\sigma_T}}^{\frac{l_u-\theta_{cup}}{\sigma_T}} (\exp(l_u)-\exp(\sigma_T z+\theta_{cup}))^2\,\phi(z|\theta_{cup},\sigma_T^2)dz \\
&\quad = \frac{\exp(2l_u)}{k_u^2}\Big[\Phi\Big(\frac{l_u-\theta_{cup}}{\sigma_T}\Big)-\Phi\Big(-\frac{\theta_{cup}}{\sigma_T}\Big)\Big] \\
&\qquad + \frac{\exp(2\theta_{cup}+2\sigma_T^2)}{k_u^2}\left[\Phi\Big(\frac{l_u-\theta_{cup}-2\sigma_T^2}{\sigma_T}\Big)-\Phi\Big(-\frac{\theta_{cup}+2\sigma_T^2}{\sigma_T}\Big)\right] \\
&\qquad - 2\frac{\exp(l_u+\theta+\frac{\sigma_T^2}{2})}{k_u^2}\left[\Phi\Big(\frac{l_u-\theta_{cup}-\sigma_T^2}{\sigma_T}\Big)-\Phi\Big(-\frac{\theta_{cup}+\sigma_T^2}{\sigma_T}\Big)\right].
\end{aligned}$$

In order to carry out a variety of inferential problems about $\tilde{Q}_{c;\alpha}$ for a given α, we use the Monte Carlo Markov Chain (MCMC). The procedures are described below.

Let C be the number of comunas covered by the model and R be the number of MCMC samples after burn-in. Let $\theta_{c;r}$ and $\sigma_{T;r}$ denote the rth MCMC draw of θ_c and σ_T, respectively $(r = 1,\ldots,R)$. We define the $C\times R$, matrix $\tilde{Q}^s_\alpha = (\tilde{Q}^s_{(c,r);\alpha})$, where the (c,r) entry is defined as

$$\tilde{Q}^s_{(c,r);\alpha} \equiv \tilde{Q}^s_{c;\alpha}(\theta_{c;r},\sigma_{T;r}).$$

This matrix $\tilde{Q}^s_\alpha$ provides samples generated from the posterior distribution of $\{\tilde{Q}_{c,\alpha},\, c = 1,\ldots,C\}$ and so is adequate for solving a variety of inferential problems in a Bayesian way. We now elaborate on the following three different inferential problems:

(1) Point estimation of an indicator of interest and the associated measure of uncertainty: This is the focus of current poverty mapping research in both classical and Bayesian approaches. Under the squared error loss function, the Bayes estimate of $Q_{c;\alpha}$ for comuna c and the associated measure of uncertainty are the posterior mean and posterior standard deviation of $\tilde{Q}_{c;\alpha} \equiv \tilde{Q}_{c;\alpha}(\theta_c, \sigma_T)$, respectively. These can be approximated by the average and standard deviation over the columns of $\tilde{Q}^s_\alpha$, respectively, for the row c, which corresponds to comuna c.

(2) Identification of comunas that are not in conformity with a given standard of a poverty indicator: In this inferential problem, the goal is to flag a comuna for which the true poverty indicator (e.g., poverty rate) exceeds a pre-specified standard, say a. In this case, point estimates, whether direct estimates or posterior means, do not give any idea about the quality of flagging a comuna not meeting the given standard. A reasonable Bayesian solution for this inferential problem is to flag comuna c for not meeting the given standard if the posterior probability $P(\tilde{Q}_{c;\alpha} > a|\text{data})$ is greater than a specified cutoff, say 0.5. This posterior probability for comuna c can be easily approximated by the proportion of columns of $\tilde{Q}^s_{c;\alpha}$ exceeding the threshold for row c. If the posterior distribution of $\tilde{Q}_{c;\alpha}$ is approximately normal, then one can alternatively use the posterior mean and posterior standard deviation to approximate the posterior probability. However, such an approximation may not perform well in many situations.

(3) Identification of the worst (best) comuna, i.e., the comuna with the maximum (minimum) value of the poverty indicator: A common solution is to identify the comuna with the maximum (minimum) point estimate of the indicator. Evidently, the use of direct point estimates would be quite misleading since such a method may identify a small comuna as being the worst (best) in terms of the indicator, even though it is not, simply because of high variability in the direct estimates. The Bayesian point estimates (posterior means) are definitely better than the direct estimates as they have generally less variability. However, the use of posterior means alone does not provide any quality measure associated with the identification of the worst (best) comuna. Even the use of posterior means along with posterior standard deviations does not help either as posterior standard deviations relate to the individual areas. A reasonable Bayesian solution in this case would be to compare the posterior probabilities $P(\tilde{Q}_{c;\alpha} \geq \tilde{Q}_{k;\alpha}\ \forall k|\text{data})$ for different comunas and select the worst (best) comuna for which this posterior probability is the maximum (minimum). Thus, along with the identification of the worst (best) comuna, we also obtain these posterior probabilities suggesting a quality of the identification of worst (best) comuna. We can use $\tilde{Q}^s_\alpha$ matrix to approximate these posterior probabilities. For row c and column r of $\tilde{Q}^s_\alpha$ corresponding to comuna c and MCMC replicate r, respectively, we can create a binary variable indicating if

the comuna is the worst (best) among all comunas. The posterior probability for this comuna $P(\tilde{Q}_{c;\alpha} \geq \tilde{Q}_{k;\alpha}\ \forall k|\text{data})$ can then be approximated by the average of these binary observations over R columns.

2.5. Numerical Results

As mentioned in the introduction, a number of researchers focused on the problem of estimation and its measure on uncertainty. While our general approach can address this problem, we choose to illustrate the general Bayesian approach for the relatively understudied inferential problems related to the identification of areas with extreme poverty (e.g., the second and the third inferential problems mentioned in Section 2.4). The data analysis presented in this section is based on the hierarchical model stated in Section 2.3 implemented on CASEN 2009 data for a given region containing 54 comunas and comuna level auxiliary variables listed in Section 2.1 We illustrate our methodology for poverty rates ($\alpha = 0$) and poverty gaps ($\alpha = 1$), two important poverty measures in the FGT class of poverty indices. After 10,000 burn-in, we generate 54×10000 matrix $\tilde{Q}^s_{c;\alpha}$ for $\alpha = 0$ (corresponding to poverty rate index) and $\alpha = 1$ (poverty gap index). We checked the convergence of MCMC convergence using the potential scale reduction factor introduced by Gelman and Rubin (1992).

Numerical results are shown in Tables 1-4. We carry out the data analysis using WinBugs-R interface. Table 1 addresses the second inferential goal, i.e., flagging the comunas that do not meet certain pre-specified standard for poverty rate. Table 2 is similar to Table 1 except that this is for the poverty gap measure. We use three different standards based on three different multipliers ($1.10, 1.25$ and 1.50) of the regional direct estimate of the respective measure. These standards are for illustration only and our approach can use any other standards that are deemed reasonable. We need a cutoff for these posterior probabilities in order to flag comunas that do not meet the given standard. To illustrate our approach, we use 0.5 as the cutoff. In other words, a comuna is deemed out of the range with respect to the pre-specified standard if the posterior probability is more than 0.5. Comunas 33 and 13 do not meet all three standards for both poverty rate and poverty gap measures. Other comunas meet the more liberal standard (1.5 times the regional poverty measure) with respect to both poverty rate and poverty gap measures. In contrast, when the standard is very conservative (1.1 times the regional poverty measure) all the comunas are not in conformity with the given standard. For a moderate standard (1.25 times the regional poverty measure), comunas 33, 13, 22, 18, 2, 6, 45, 16, 30 do not satisfy the standard in terms of poverty rate measure. The comunas 21, 5, 17 and 15 are added to the list when we consider the poverty gap measure. The standard and the cut-off to be used are subjective, but the Bayesian approach with different standard and cutoff combinations should give policy makers some useful guidance in making certain policy decisions. In order to save space in Table 1 (Table 2), we report results for the 29 out of 54 comunas in the region with highest posterior probabilities of poverty rate (poverty gap) exceeding the most conservative threshold.

Table 1: The posterior probabilities that poverty rate for a comuna exceeds three different thresholds; $Q_{r,0}$ is direct estimate of regional poverty rate. The table presents results for the 29 comunas (out of 54 comunas in the region) with the highest $P(\widetilde{Q}_{c;0} > 1.10Q_{r,0}|\text{data})$

Comuna	$P(\widetilde{Q}_{c;0} > 1.10Q_{r,0}\|\text{data})$	$P(\widetilde{Q}_{c;0} > 1.25Q_{r,0}\|\text{data})$	$P(\widetilde{Q}_{c;0} > 1.50Q_{r,0}\|\text{data})$
33	1.0000	0.9995	0.6172
13	1.0000	0.9988	0.5636
22	0.9952	0.7962	0.0314
18	0.9904	0.6996	0.0100
2	0.9834	0.4939	0.0005
6	0.9809	0.5331	0.0006
45	0.9786	0.5755	0.0032
16	0.9721	0.5157	0.0015
30	0.9662	0.5086	0.0024
21	0.9362	0.3925	0.0013
5	0.9356	0.3878	0.0010
17	0.9258	0.3840	0.0012
15	0.9185	0.3643	0.0015
25	0.8822	0.2524	0.0002
43	0.8755	0.2266	0.0000
38	0.8612	0.2209	0.0003
27	0.8466	0.2139	0.0002
26	0.8425	0.3259	0.0022
51	0.8365	0.2941	0.0009
24	0.7835	0.1216	0.0000
29	0.7111	0.0995	0.0000
28	0.7030	0.1085	0.0000
31	0.6771	0.0700	0.0000
35	0.6694	0.1018	0.0000
36	0.6404	0.0731	0.0000
41	0.6142	0.0591	0.0001
37	0.6041	0.0775	0.0000
7	0.5705	0.0386	0.0000
47	0.5179	0.0417	0.0000

Table 2: Posterior probabilities that poverty gap for a given comuna exceeds three different thresholds; $Q_{r,1}$ is direct estimate of regional poverty gap. The table presents results for the 29 comunas (out of 54 comunas in the region) with the highest $P(\tilde{Q}_{c;1} > 1.10Q_{r,1}|\text{data})$ values.

Comuna	$P(\widetilde{Q}_{c;1} > 1.10Q_{r,1}\|\text{data})$	$P(\widetilde{Q}_{c;1} > 1.25Q_{r,1}\|\text{data})$	$P(\widetilde{Q}_{c;1} > 1.50Q_{r,1}\|\text{data})$
33	1.0000	0.9998	0.9266
13	1.0000	0.9994	0.9060
22	0.9966	0.9143	0.2635
18	0.9918	0.8327	0.1195
2	0.9893	0.7516	0.0395
45	0.9827	0.7577	0.0781
6	0.9824	0.7174	0.0300
16	0.9792	0.7189	0.0490
30	0.9693	0.6764	0.0489
21	0.9467	0.5871	0.0320
5	0.9420	0.5656	0.0240
17	0.9339	0.5592	0.0292
15	0.9333	0.5631	0.0337
38	0.9001	0.4329	0.0081
25	0.8923	0.4203	0.0079
43	0.8802	0.3812	0.0070
26	0.8751	0.5310	0.0657
27	0.8745	0.3970	0.0104
51	0.8540	0.4497	0.0305
24	0.8223	0.2674	0.0018
29	0.7700	0.2401	0.0026
28	0.7441	0.2390	0.0026
31	0.7321	0.1848	0.0005
35	0.6924	0.2091	0.0026
36	0.6671	0.1631	0.0006
37	0.6399	0.1772	0.0021
7	0.6376	0.1243	0.0002
41	0.6355	0.1365	0.0003
47	0.5586	0.1095	0.0003

Table 3 displays approximations (by MCMC) to the posterior probabilities of a comuna being the worst (Prob.Max) in terms of both poverty rate and poverty gap measures. According to the Prob.Max criterion, comuna 33 stands out as the worst comuna in terms of both poverty rate and poverty gap measures. Table 4 displays approximations (by MCMC) to the posterior probabilities of a comuna being the best (Prob.Min) in terms of both poverty rate and poverty gap measures. According to Prob.Min criterion, comuna 8 emerges as the best comuna in terms of both poverty rate and poverty gap measures. These probabilities are also giving us a good sense of confidence we can place on our decision, which is not possible with poverty rate and poverty gap estimates alone. Tables 3 and 4 do not report results for comunas with negligible posterior probabilities.

Table 3: Posterior probability that poverty rate or poverty gap for a given comuna is the maximum (Prob.Max). The table does not include comunas with negligible posterior probabilities.

Comuna	Prob.Max	
	Poverty Rate	Poverty Gap
33	0.5126	0.5246
13	0.4496	0.4301
22	0.0169	0.0215
18	0.0051	0.0044
45	0.0025	0.0031
17	0.0021	0.0021
26	0.0021	0.0042
30	0.0017	0.0017
21	0.0013	0.0016
15	0.0010	0.0011
16	0.0009	0.0012
51	0.0008	0.0009
6	0.0007	0.0006
5	0.0006	0.0006
2	0.0005	0.0009
27	0.0005	0.0004
38	0.0005	0.0006
25	0.0003	0.0001
35	0.0001	0.0002
41	0.0001	0.0000
43	0.0001	0.0000
28	0.0000	0.0001

Table 4: Posterior probability that poverty rate or poverty gap for a given comuna is the minimum (Prob.Min). The table does not include comunas with negligible posterior probabilities.

Comuna	Prob.Min	
	Poverty Rate	Poverty Gap
8	0.5310	0.5161
1	0.3929	0.3945
42	0.0240	0.0268
48	0.0186	0.0237
12	0.0121	0.0139
4	0.0075	0.0089
34	0.0057	0.0079
3	0.0052	0.0047
14	0.0009	0.0011
10	0.0009	0.0005
23	0.0008	0.0012
44	0.0002	0.0004
46	0.0001	0.0002
40	0.0001	0.0001

3. Concluding Remarks

We point out inappropriateness of using point estimates for all inferential purposes and propose a general Bayesian approach to solve different inferential problems in the context of poverty mapping. The proposed approach provides not only an action relevant to the inferential problem but also a way to assess the quality of such action. To make the methodology user-friendly one can store the Q_α^s matrix of size $C \times R$, where C is the number of comunas and R is the number of MCMC replications. This way the users do not need to know how to generate this matrix, which requires knowledge of advanced Bayesian computing. Once the user has access to this generated matrix, he/she can easily carry out a variety of statistical analysis such as the ones presented in the paper with greater ease. While we illustrate the approach for the FGT poverty indices, the approach is general and can deal with other important indices such as the ones given in sustainable development goals. We have taken one working model to illustrate the approach, but the approach is general and can be applied to other models that are deemed appropriate in other projects.

4. Acknowledgments

We thank Editor-in-Chief Professor Wlodzimierz Okrasa and an anonymous referee for reading an earlier version of the article carefully and offering a number of constructive suggestions, which led to a significant improvement of our article. The first author's research was partially supported by the U.S. National Science Foundation grant SES-1758808.

REFERENCES

BELL, W. R., BASEL, W. W., MAPLES, J. J., (2016). An overview of the U.S. Census Bureaus Small Area Income and Poverty Estimates Program. In M. Pratesi (Ed.) *Analysis of poverty data by small area estimation* (pp. 349-377). West Sussex: Wiley & Sons, Inc.

CASAS-CORDERO VALENCIA, C. ENCINA, J., LAHIRI, P., (2016). Poverty mapping in Chilean comunas. In M. Pratesi (Ed.) *Analysis of Poverty Data by Small Area Estimation* (pp. 379–403). West Sussex: Wiley & Sons, Inc.

CHATTERJEE, S., LAHIRI, P., LI, H., (2008). Parametric bootstrap approximation to the distribution of EBLUP and related prediction intervals in linear mixed models. *The Annals of Statistics*, 36, 3, pp. 1221–1245.

CLAYTON, D., KALDOR, J., (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, 43, pp. 671–681.

EFRON, B., MORRIS, C., (1975). Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association*, 70, pp. 311–319.

ELBERS, C., LANJOUW, J. O., LANJOUW, P., (2003). Micro-level estimation of poverty and inequality. *Econometrica*, 71, pp. 355–364.

FAY, R. E., HARRIOT, R., (1979). Estimation of income from small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74, pp. 269–277.

FOSTER, J., GREER, J., THORBECKE, E., (1984). A class of decomposable poverty measures. *Econometrica*, 52, pp. 761–766.

FRONCO, C., BELL, W. R., (2015). Borrowing information over time in Binomial/Logit normal models for small area estimation. *Joint Special Issue of Statistics in Transition and Survey Methodology,* 16, 4, pp. 563–584.

GELMAN, A., (2015). 3 New priors you can't do without, for coefficients and variance parameters in multilevel regression. *Statistical Modeling, Causal Inference, and Social Science blog, 7 Nov.* http://andrewgelman.com/2015/11/07/priors-for-coefficients-and-variance-parameters-in-multilevel-regression/

GELMAN, A., PRICE, P. N., (1999). All maps of parameter estimates are misleading. *Statistics in Medicine*, 18, pp. 3221–3234.

GELMAN, A., RUBIN, D. B., (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7,4, pp. 457–511.

GHOSH, M. (1992). Constrained Bayes estimation with applications. *Journal of the American Statistical Association*, 87, 418, pp. 533–540.

JIANG, J., LAHIRI, P., (2006). Mixed model prediction and small area estimation. *TEST*, 15, 1–96.

JONES, H. E., SPIEGELHALTER, D. J., (2011). The identification of unusual health-care providers from a hierarchical model. *American Statistician*, 65, pp. 154–163, DOI: 10.1198/tast.2011.10190.

LAHIRI, P., (1990), "Adjusted" Bayes and empirical Bayes estimation in finite population sampling, *Sankhya*, B, 52, pp. 50–66.

LANGFORD, I. H., LEWIS, T., (1998). Outliers in multilevel data. *Journal of the Royal Statistical Society, Ser. A*, 161, pp. 121–160.

LI, H. and LAHIRI, P., (2010). Adjusted maximum method for solving small area estimation problems, *Journal of Multivariate Analysis*, 101, pp. 882–892.

LOUIS, T. A., (1984). Estimating a population of parameter values using Bayes and empirical Bayes methods. *Journal of the American Statistical Association*, 79, 386, pp. 393–398.

MARSHALL, R. J., (1991). Mapping disease and mortality rates using empirical Bayes estimators. *Appllied Statistics*, 40, pp. 283–294.

MOLINA, I., NANDRAM, B., RAO, J. N. K., (2014). Small area estimation of general parameters with application to poverty indicators: a hierarchical Bayes approach. *The Annals of Applied Statistics*, Vol. 8, No. 2, pp. 852–885.

MOLINA, I., RAO, J. N. K., (2010). Small area estimation of poverty indicators. *Canadian Journal of Statistics*, 38, pp. 369–385.

MOLINA, A., RICHARDSON, S., (1991). Empirical Bayes estimates of cancer mortality rates using spatial models. *Statistics in Medicine*, 10, pp. 95–112.

MORRIS, C. N., CHRISTIANSEN, C. L., (1996). Hierarchical models for ranking and for identifying extremes, with applications, in Bayesian statistics 5, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F.M. Smith, Oxford: Oxford University Press, pp. 277–296.

NANDRAM, B., SEDRANSK, J., PICKLE, L. W., (2000). Bayesian analysis and mapping of mortality rates for chronic obstructive pulmonary disease. *Journal of American Statistical Association*, 95, pp. 1110–1118.

PFEFFERMANN, D., (2013). New important developments in small area estimation. *Statistical Science*, 28, pp. 40–68.

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

RAO, J.N.K. and MOLINA, A., (2015). *Small area estimation.* Wiley.

RAVALLION, M., (1994). *Poverty comparisons*. Chur, Switzerland: Harwood Academic Press.

SHEN, W., LOUIS, T. A., (1998). Triple-goal estimates in two-stage hierarchical models. *Journal of Royal Statistical Society Series B*, 60,Part 2, pp. 455–471.

TSUTUKAWA, R. K., SHOOP, G.L., MARIENFELD. C. J., (1985). Empirical Bayes estimation of cancer mortality rates. *Statistics in Medicine*, 4, pp. 201–212.