Hierarchical Bayesian Models for Noisy Size Responses From Small Areas

Binod Manandhar* Balgobin Nandram†

Abstract

We implement the techniques of small area estimation (SAE) to study a positively skewed welfare indicator, consumption. A logarithmic transformation that may be problematic is usually suggested for positively skewed data to build a model. In our study, we have developed hierarchical Bayesian models without log-transformation. We applied our model to the Nepal Living Standards Survey, 2003/04 consumption data, an aggregate of all food and nonfood items consumed in the past twelve months. Since, the respondent has to recall the consumptions in the past twelve months, we assume that data are recorded with noises. For the noisy data, we fit three special cases of generalized beta distribution of the second kind (GB2) models using the Metropolis Hastings sampler. After fitting Bayesian models for SAE, we show how to select the most plausible model and perform the Bayesian data analysis.

Key Words: GB2 distribution, Generalized gamma distribution, Logarithmic transformation, Markov chain Monte Carlo, Skewed distribution

1. Introduction

Continuous and positively skewed (CPS) data, such as consumption, income, insurance, and loss in numerous applications, are examples of size data. Such data are generally heavy-tailed and skewed to the right. The logarithmic transformation is the most widely used tool to meet the normality assumption for CPS size data. If we have used a transformation to build a model, the usual way to get estimates back to their original scale is to perform a back-transformation. Does back-transformation give a correct distribution of the response variable? Furthermore, what if the normality assumption fails? Feng et al. (2013, 2014) discussed the problems with using the logarithmic transformation for positively skewed data.

There are numerous research papers and reports which use logarithmic transformation for skewed data to build a model. The World Bank uses the Elbers, Lanjouw, and Lanjouw (ELL) method (2001, 2003) for small area estimation (SAE) of the poverty measures and the response is a logarithmic transformed welfare variable. It is a nested error model that decomposes the total error into the sum of the area error and the unit error (Battese, Harter, and Fuller, 1988). It had already been applied in 60 countries by 2011 (The World Bank, 2013). The empirical Bayesian nested error model for SAE also uses the logarithmic transformed welfare variable as the response variable (Molina and Rao 2010). The hierarchical Bayesian model for CPS data is shown in the paper, by Molina, Nandram, and Rao in 2014. It also uses a logarithmic transformed response variable. If there are two or more levels of hierarchies, a multi-level model could be another choice in SAE (Nguyen et. al. 2010).

We build hierarchical Bayesian models for the CPS data without logarithmic transformation and predict the responses for both sampled and non-sampled units. We focus on giving estimates for small areas. SAE is essential for different sectors like government agencies, developmental partners, planners, and researchers for many purposes like developmental planning.

*Department of Mathematical Sciences, Worcester Polytechnic Institute, Worcester, MA 01609
†Department of Mathematical Sciences, Worcester Polytechnic Institute, Worcester, MA 01609

We assume that the response data are recorded with noises. The noise could have been introduced into the response data as recalling errors. We fit models with the generalized beta distribution of the second kind (GB2), the mixture of two generalized gamma distributions, where the distribution of the response variable is mixed with the distribution of its rate parameter, both having generalized gamma distributions. We fit three special cases of the GB2: the mixture of exponential and gamma distributions, the mixture of two gamma distributions, and the mixture of two generalized gamma distributions. We apply our models to the welfare consumption CPS data from the second Nepal Living Standards Survey (NLSS-II), 2003/04.

Model adequacy has been checked by a Bayesian cross-validation approach. We have used summary statistics of the conditional predictive ordinates (CPO), the logarithm of the pseudo-marginal likelihood (LPML) to compare models. After selection of the best fit model, it is applied to the census data for the prediction of responses and then small area estimates are obtained. To calculate the poverty indicators, we compare the predicted per capita consumption against the national poverty line.

NLSS-II data are available for 3912 households enumerated in the national survey. In NLSS-II the welfare response variable per capita consumption is the aggregate of all food and non-food items consumed in the past twelve months. The responders had to recall all kinds of consumptions in monetary value throughout the reference year. In addition, for each food item the respondent had to recall the number of months consumed and quantity consumed in the typical month, then evaluate its market value at that time. It could also be possible that there could be bias of reporting quantity and money values of food and non-food consumed by some households. Hence, there could be the possibility of introducing noise in the data.

We chose nine relevant covariates, which can influence welfare status, and per capita consumption from the NLSS-II survey for modeling. These covariates have a moderate correlation with the response variable. To facilitate SAE we have population census 2001 data with these nine covariates. They are (i) “Household size” (*hhsize*), (ii) “proportion of kids aged 0 - 6 in the household” (*skids6*), (iii) “proportion of kids aged 7 - 14 in the household” (*skids714*), (iv) “abroad migrant” (*remtab*), (v) “House temporary” (*hutype3*), (vi) “House owned” (*huown2*), (vii) “proportion of households with cooking fuel LP/gas in Ward” (*ckfuel3w*), (viii) “proportion of households with land-owning females in municipality/VDC” (*pflandv*), and (ix) “proportion of kids 6-16 attending school in municipality/VDC” (*pschv*).

We organize the paper as follows: In Section 2 we discuss the GB2 distribution as the mixture of the exponential and gamma, the mixture of two gamma and the mixture of two generalized gamma distributions. In Section 3 we develop hierarchical Bayesian model and discuss parameters sampling. In Section 4, we show the application to NLSS-II and population census data and simulation results for SAE.

2. Modeling with GB2

We use GB2 to model size (positive values) data which can also be expressed as a mixture of two generalized gamma distributions. We exploit this property in model building. Let the probability density function of the response variable $y|\alpha, \lambda, \gamma$ and the probability density function of its rate parameter $\lambda|\phi, \theta, \gamma$ both have the generalized gamma distribution

$$y|\alpha, \lambda, \gamma \sim \text{GGamma}(\alpha, \lambda, \gamma),$$

$$\lambda|\phi, \theta, \gamma \sim \text{GGamma}(\phi, \theta, \gamma).$$

Mixing these generalized gamma densities, we get the GB2 density

$$\begin{aligned} f(y|\alpha,\phi,\theta) &= \int_0^\infty f(y|\lambda,\alpha,\gamma)\ \ g(\lambda|\theta,\phi,\gamma)\ d\lambda \\ &= \frac{\gamma}{\mathrm{B}(\frac{\alpha}{\gamma},\frac{\phi}{\gamma})}\ \frac{y^{\alpha-1}}{\theta^\alpha(1+\left(\frac{y}{\theta}\right)^\gamma)^{\frac{\alpha+\phi}{\gamma}}}, \quad \theta,\alpha,\phi>0,\ y>0. \end{aligned} \tag{1}$$

We note here that the rate parameter λ of the response variable $\boldsymbol{y}$ has been integrated out. We have developed three different mixtures of the GB2 distribution in an hierarchical order.

Exponential-Gamma Mixture Model: The moments do not exist for a mixture of the two exponential distributions so we use the simplest GB2 model as a mixture of the exponential and gamma distributions. If the response variable $Y|\lambda$ has an exponential distribution, and its rate parameter $\lambda|\alpha,\theta$ has a gamma distribution, mixing these two distributions we get GB2 density

$$f(y|\alpha,\theta) = \frac{\alpha}{\theta(1+\frac{y}{\theta})^{\alpha+1}}\ , \quad \alpha,\theta>0. \tag{2}$$

Mixture of two Gamma GB2 Model: Let the response variable have the gamma distribution $Y|\alpha,\lambda\ \sim$ Gamma(α,λ) and its rate parameter have the gamma distribution $\lambda|\phi,\theta\ \sim$ Gamma(ϕ,θ). Mixing these two gamma densities and integrating out λ, we get the GB2 model

$$\begin{aligned} f(y|\alpha,\phi,\theta) &= \frac{y^{\alpha-1}\theta^\phi}{\Gamma(\alpha)\Gamma(\phi)}\int_\lambda e^{-(\theta+y)\lambda}\ \ \lambda^{\alpha+\phi-1}d\lambda \\ &= \frac{y^{\alpha-1}}{\mathrm{B}(\alpha,\phi)}\ \frac{1}{\theta^\alpha(1+\frac{y}{\theta})^{\alpha+\phi}}, \quad \theta,\alpha,\phi>0. \end{aligned} \tag{3}$$

Its k^{th} moment is given by

$$E[Y^k|\alpha,\phi,\theta] = \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}\frac{\Gamma(\phi-k)}{\Gamma(\phi)}\theta^k. \tag{4}$$

For variance to exist in this density, we need $\phi>2$. If we consider α and ϕ to be two distinct shape parameters, they are not identifiable (see below). We consider that two rate parameters, α and ϕ, are related linearly as, $\phi=\alpha+2$. This also allows the variance to exist. Considering this linear relationship between two shape parameters, we have the GB2 density function from the mixture of two gamma distributions as

$$f(y|\alpha,\theta) = \frac{y^{\alpha-1}}{\mathrm{B}(\alpha,\alpha+2)}\ \frac{1}{\theta^\alpha(1+\frac{y}{\theta})^{2(\alpha+1)}}, \quad \theta,\alpha>0. \tag{5}$$

Non-identifiable Parameters Let us say we have n independent samples from the gamma distribution, $Y_i|\lambda_i,\alpha\sim$ Gamma(λ_i,α). We would like to find a maximum likelihood estimate (MLE) for the parameters $\alpha,\lambda_i,i=1,\cdots,n$. The likelihood function is

$$f(\boldsymbol{y}|\alpha,\boldsymbol{\lambda}) = \prod_{i=1}^n \frac{e^{-\lambda_i y_i}\ y^{\alpha-1}}{\Gamma(\alpha)}\ \lambda_i^\alpha, \quad \lambda_i,\alpha>0,\ i=1,\cdots,n.$$

The log-likelihood function is

$$\Delta = \sum_{i=1}^{n} [-\lambda_i y_i + (\alpha - 1)log(y_i) + \alpha log(\lambda_i)] - nlog\Gamma(\alpha).$$

The maximum likelihood estimator (MLE) for λ_i, is $\hat{\lambda}_i = \frac{\alpha}{y_i},\ i = 1, \cdots, \ell$. Substituting the MLE of λ_i in the log-likelihood function gives us

$$\Delta = \sum_{i=1}^{n} [-\alpha + (\alpha - 1)log(y_i) + \alpha log(\alpha/y_i)] - nlog\Gamma(\alpha).$$

Taking the partial derivative with respect to α gives

$$\frac{\partial \Delta}{\partial \alpha} = nlog(\alpha) - n\Psi(\alpha),$$

where $\Psi(\alpha) = \frac{d}{d\alpha} log\Gamma(\alpha)$. Setting $\frac{\partial \Delta}{\partial \alpha} = 0$, we have $log(\alpha) = \Psi(\alpha)$. It has no solution. Therefore, if each response y_i has its parameter $\lambda_i, i = 1, \cdots, n$, then the parameters (α, λ_i) together are not identifiable.

Mixture of two Generalized Gamma GB2 Model: Let the response variable have the generalized gamma distribution, $Y|\alpha, \lambda, \gamma \sim$ GGamma$(\alpha, \lambda, \gamma)$ and let its rate parameter also have the generalized gamma distribution, $\lambda|\phi, \theta, \gamma \sim$ GGamma(ϕ, θ, γ). Mixing these two distributions and integrating out λ, we have the following GB2 density

$$\begin{aligned} f(y|\alpha, \phi, \theta, \gamma) &= \gamma^2 \frac{y^{\alpha-1}\theta^{\phi}}{\Gamma(\frac{\alpha}{\gamma})\Gamma(\frac{\phi}{\gamma})} \int_{\lambda} e^{-(\theta^{\gamma}+y^{\gamma})\lambda^{\gamma}}\ \lambda^{\alpha+\phi-1}\, d\lambda, \\ &= \frac{\gamma\, y^{\alpha-1}}{\mathrm{B}\left(\frac{\alpha}{\gamma}, \frac{\phi}{\gamma}\right)} \frac{1}{\theta^{\alpha}(1 + (\frac{y}{\theta})^{\gamma})^{\frac{\alpha+\phi}{\gamma}}}, \quad \theta, \alpha, \phi, \gamma > 0. \end{aligned} \tag{6}$$

Its k^{th} moment is

$$E[Y^k|\alpha, \phi, \theta, \gamma] = \frac{\Gamma\left(\frac{\alpha+k}{\gamma}\right)}{\Gamma\left(\frac{\alpha}{\gamma}\right)} \frac{\Gamma\left(\frac{\phi-k}{\gamma}\right)}{\Gamma\left(\frac{\phi}{\gamma}\right)} \theta^k. \tag{7}$$

As before, in the mixture of the two gamma distributions, we need $\phi > 2$ for the variance to exist. We assume that the two shape parameters α and ϕ are related linearly as, $\phi = \alpha + 2$. Then the GB2 density can be written as

$$f(y|\alpha, \theta, \gamma) = \frac{\gamma y^{\alpha-1}}{\mathrm{B}(\frac{\alpha}{\gamma}, \frac{\alpha+2}{\gamma})} \frac{1}{\theta^{\alpha}(1 + (\frac{y}{\theta})^{\gamma})^{\frac{2(\alpha+1)}{\gamma}}}, \quad \theta, \alpha, \gamma > 0. \tag{8}$$

2.1 Model comparison

Bayesian cross-validation approach by Gelfand, Dey, and Chang (1992) has been used to evaluate the adequacy of a model. The cross-validation approach involves the prediction of subset y_i of the response data $\boldsymbol{y}$, when only the component $\boldsymbol{y}_{(i)}$ is used. Let $\boldsymbol{y}$ be the data vector of $N \times 1$ and $\boldsymbol{y}_{(i)}$ and denote the $(N-1) \times 1$ data vector with the i^{th} observation deleted. If we fit a model with $\boldsymbol{y}_{(i)}$ and if the model fits well, then it should predict y_i very well.

The cross-validation approach needs to find $p(y_i|y_{(i)})$, called the cross-validation predictive distribution or conditional predictive ordinate (CPO). We consider the CPO defined by Box (1980) and the studies under normal distribution by Pettit (1990)

$$CPO_i \approx \sum_{k=1}^{M} f(y_i|y_{(i)}, \mathbf{\Omega}) \left[\frac{\left[f(y_{(i)}|\mathbf{\Omega}^{(k)})\right]^{-1}}{\sum_{k=1}^{M} \left[f(y_{(i)}|\mathbf{\Omega}^{(k)})\right]^{-1}} \right].$$

A summary statistic of the $CPO_i's$ is the logarithm of the pseudo-marginal likelihood (LPML) defined as

$$LPML = \sum_{i=1}^{n} \log\left(CPO_i\right). \tag{9}$$

3. Hierarchical Bayesian Models

Consider sample data with n observations, response variable $\boldsymbol{y}_{n\times 1}$, and covariate $\boldsymbol{x}_{n\times p}$. We have ℓ small areas, $i = 1, \cdots, \ell$, and each small area has $j = 1, \cdots, n_i$, observations. Let y_{ij} and $\boldsymbol{x}_{ij}$, $i = 1 \cdots, \ell$, $j = 1 \cdots, n_i$ denote the response variable and the corresponding covariates in the i^{th} area and j^{th} observation.

Below, we show the GB2 model as the mixture of two generalized gamma distributions. In a similar way we have built models for the mixture of the exponential and gamma distributions and the mixture of two gamma distributions are not presented here.

3.1 Two Generalized Gamma Mixture GB2 Model

We assume that the responses $y_{ij}|\alpha, \theta, \gamma,\ i = 1, \cdots, n,\ j = 1, \cdots, n_i$ are random samples from the GB2 distribution. We considered two shape parameters α and ϕ which are linearly related, $\phi = \alpha + 2$. We introduce the covariates through the rate parameter by writing $e^{\boldsymbol{x'}_{ij}\boldsymbol{\beta}+\boldsymbol{\nu_i}}$ and assume ν_i follows the normal distribution with mean zero and variance σ^2. The likelihood function is

$$\pi(\boldsymbol{y}|\alpha, \boldsymbol{\beta}, \boldsymbol{\nu}) = \prod_{i=1}^{\ell}\prod_{j=1}^{n_i} \gamma \frac{y_{ij}^{\alpha-1}}{\mathrm{B}(\frac{\alpha}{\gamma}, \frac{\alpha+2}{\gamma})} \frac{e^{-\alpha(\boldsymbol{x'}_{ij}\boldsymbol{\beta}+\nu_i)}}{\left(1+\left[y_{ij}\, e^{-(\boldsymbol{x'}_{ij}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}\right)^{\frac{2(\alpha+1)}{\gamma}}}. \tag{10}$$

Let $\boldsymbol{\alpha}$ and β have non-informative priors and γ have an informative prior. The priors are independent. The hierarchical Bayesian GB2 model with random area effects is

$$\begin{aligned}
y_{ij}|\boldsymbol{\beta}, \alpha, \gamma, \nu_i &\overset{\text{ind}}{\sim} \text{GB2}\left(\alpha, e^{\boldsymbol{x'}_{ij}\boldsymbol{\beta}+\boldsymbol{\nu_i}}, \gamma\right),\ \theta_{ij} = e^{\boldsymbol{x'}_{ij}\boldsymbol{\beta}+\boldsymbol{\nu_i}},\ i = 1, \cdots, n,\ j = 1, \cdots, n_i, \\
\nu_i &\overset{\text{iid}}{\sim} N(0, \sigma^2), \\
\pi(\boldsymbol{\beta}, \alpha, \sigma^2) &\propto \frac{1}{(1+\alpha)^2\,(1+\sigma^2)^2} \\
\gamma &\sim \text{Gamma}(S, R), \qquad \text{where shape S and rate R are specified.}
\end{aligned} \tag{11}$$

Combining the likelihood in (10) and the priors in (11) via Bayes theorem, we get the joint posterior density of $\alpha, \boldsymbol{\beta}, \gamma, \boldsymbol{\nu}, \sigma^2|\boldsymbol{y}$ as

$$\pi(\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{\nu},\sigma^2|\boldsymbol{y}) \propto f(\boldsymbol{y}|\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{\nu},\sigma^2)\,\pi(\boldsymbol{\nu}|\sigma^2)\pi(\boldsymbol{\beta},\alpha,\sigma^2)$$

$$= \left[\frac{\gamma\, g^{\alpha-1}}{\mathrm{B}(\frac{\alpha}{\gamma},\frac{\alpha+2}{\gamma})}\right]^n e^{-\alpha\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x'_{ij}}\boldsymbol{\beta}} \prod_{i=1}^{\ell}\left[\frac{e^{-\alpha\, n_i\nu_i}}{\prod_{j=1}^{n_i}\left(1+\left[y_{ij}\, e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}\right)^{\frac{2(\alpha+1)}{\gamma}}}\right]$$

$$\times\prod_{i=1}^{\ell}\left[\left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}} e^{-\frac{\nu_i^2}{2\sigma^2}}\right]\times\frac{e^{-R\gamma}\,\gamma^{S-1}}{(1+\alpha)^2\,(1+\sigma^2)^2}. \tag{12}$$

Let the log-likelihood function be $G(\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{\nu}|\boldsymbol{y})$. For notational simplicity we write G only, then its first- and second-order partial derivatives with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\nu}$ are

$$\frac{\partial G}{\partial\boldsymbol{\beta}} = \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\left(-\alpha\;+2(\alpha+1)\frac{\left[y_{ij}\,e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}}{1+\left[y_{ij}\,e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}}\right)\boldsymbol{x_{ij}},$$

$$\frac{\partial G}{\partial\nu_i} = -\,\alpha\, n_i + 2(\alpha+1)\sum_{j=1}^{n_i}\frac{\left[y_{ij}\,e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}}{1+\left[y_{ij}\,e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}},$$

$$\frac{\partial^2 G}{\partial\boldsymbol{\beta}^2} = -\,2(\alpha+1)\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\gamma\frac{\left[y_{ij}\,e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}}{\left(1+\left[y_{ij}\,e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}\right)^2}\boldsymbol{x_{ij}}\boldsymbol{x'_{ij}},$$

$$\frac{\partial^2 G}{\partial\nu_i^2} = -\,2(\alpha+1)\sum_{j=1}^{n_i}\gamma\frac{\left[y_{ij}\,e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}}{\left(1+\left[y_{ij}\,e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}\right)^2},$$

$$\frac{\partial^2 G}{\partial\boldsymbol{\beta}\partial\nu_i} = -\,2(\alpha+1)\sum_{j=1}^{n_i}\gamma\frac{\left[y_{ij}\,e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}}{\left(1+\left[y_{ij}\,e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}\right)^2}\boldsymbol{x_{ij}}.$$

Using the first-order Taylor's series approximation for $\frac{\left(y_i\,e^{-(\boldsymbol{x'_i}\boldsymbol{\beta}+\boldsymbol{\nu_i})}\right)^{\gamma}}{1+\left(y_i\,e^{-(\boldsymbol{x'_i}\boldsymbol{\beta}+\boldsymbol{\nu_i})}\right)^{\gamma}}$ at $\boldsymbol{\beta}=\boldsymbol{0}$, the approximate MLE of $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta}^*|\alpha,\gamma,\boldsymbol{\nu} = \left[\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\frac{\gamma\left(y_{ij}\,e^{-\nu_i}\right)^{\gamma}}{\left(1+\left(y_{ij}\,e^{-\nu_i}\right)^{\gamma}\right)^2}(\boldsymbol{x_i}\boldsymbol{x'_i})\right]^{-1}\left(\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\left(\frac{\left(y_{ij}\,e^{-\nu_i}\right)^{\gamma}}{1+\left(y_{ij}\,e^{-\nu_i}\right)^{\gamma}}-\frac{\alpha}{2(\alpha+1)}\right)\boldsymbol{x_{ij}}\right). \tag{13}$$

Similarly, using the first-order Taylor's series approximation at $\nu_i=0$, we have the MLE of ν_i given by

$$\nu_i^*|\alpha,\gamma,\boldsymbol{\beta} = \left[\sum_{j=1}^{n_i}\frac{\gamma\left(y_{ij}\,e^{-\boldsymbol{x'_{ij}}\boldsymbol{\beta}}\right)^{\gamma}}{\left(1+\left(y_{ij}\,e^{-\boldsymbol{x'_{ij}}\boldsymbol{\beta}}\right)^{\gamma}\right)^2}\right]^{-1}\left[\sum_{j=1}^{n_i}\left(\frac{\left(y_{ij}\,e^{-\boldsymbol{x'_{ij}}\boldsymbol{\beta}}\right)^{\gamma}}{1+\left(y_{ij}\,e^{-\boldsymbol{x'_{ij}}\boldsymbol{\beta}}\right)^{\gamma}}-\frac{\alpha\, n_i}{2(\alpha+1)}\right)\right]. \tag{14}$$

Let the gradient vectors be $\nabla G(\tau^*)=(\boldsymbol{g'_\nu},\,\boldsymbol{g'_\beta})'$, where $\boldsymbol{g_\nu}=\left(\frac{\partial G}{\partial\nu_1}\;\cdots\;\frac{\partial G}{\partial\nu_\ell}\right)'|_{\boldsymbol{\nu}=\boldsymbol{\nu}^*,\,\boldsymbol{\beta}=\boldsymbol{\beta}^*}$ and $\boldsymbol{g_\beta}=\left(\frac{\partial G}{\partial\beta_0}\;\cdots\;\frac{\partial G}{\partial\beta_p}\right)'|_{\boldsymbol{\nu}=\boldsymbol{\nu}^*,\,\boldsymbol{\beta}=\boldsymbol{\beta}^*}$ and the Hessian matrix be H evaluated at the approximate mode values $\boldsymbol{\beta}^*$ and $\boldsymbol{\nu}^*$. Then using the second-order Taylor's series approximation, we can write the approximated likelihood function as

$$f(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\nu}) \approx e^{\left[G(\boldsymbol{\tau^*})+\frac{1}{2}(\nabla G(\tau^*))'\,(-H(\boldsymbol{\tau^*}))^{-1}\,\nabla G(\tau^*)\right]}$$
$$\times\,(2\pi)^{\frac{p+\ell}{2}}\,\left|(-H(\tau^*))^{-1}\right|^{\frac{1}{2}}\,N\left[\tau^*+(-H(\boldsymbol{\tau^*}))^{-1}\,\nabla G(\tau^*),\,(-H(\boldsymbol{\tau^*}))^{-1}\right],$$

where N denotes the multivariate normal distribution for the parameter set $\boldsymbol{\tau} = (\boldsymbol{\beta}',\, \boldsymbol{\nu}')'$. Following the *multivariate normal approximation theorem* we can write

$$\begin{pmatrix} \boldsymbol{\nu} \\ \boldsymbol{\beta} \end{pmatrix} \sim N \left\{ \begin{pmatrix} \boldsymbol{\mu}^*_{\boldsymbol{\nu}} \\ \boldsymbol{\mu}^*_{\boldsymbol{\beta}} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix} \right\},$$

where the Hessian matrix is $H = -\begin{pmatrix} A_{11} & A_{12} \\ A_{12}' & A_{22} \end{pmatrix}$. Let us denote

$$C_{\alpha\gamma}(\tau^*) = e^{\left[G(\boldsymbol{\tau}^*) + \frac{1}{2}(\nabla G(\tau^*))'\,(-H(\boldsymbol{\tau}^*))^{-1}\,\nabla G(\tau^*)\right]} \left|(-H(\tau^*))^{-1}\right|^{\frac{1}{2}}.$$

By applying the *multivariate normal approximation theorem* we can write the approximate joint posterior density as

$$\begin{aligned}
& f(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\nu}, \sigma^2 | \boldsymbol{y}) \\
& \quad \propto C_{\alpha\gamma}(\tau^*) \times N\left(\boldsymbol{\mu}^*_\beta,\ \Sigma_{22}\right) \times N\left(\boldsymbol{\mu}^*_\nu + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}^*_\beta),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'\right) \\
& \qquad \times N\left(\mathbf{0},\ \sigma^2 I_\ell\right) \times \frac{e^{-R\gamma}\,\gamma^{S-1}}{(1+\alpha)^2\,(1+\sigma^2)^2} \\
& \quad = C_{\alpha\gamma}(\tau^*) \times \frac{e^{-R\gamma}\,\gamma^{S-1}}{(1+\alpha)^2\,(1+\sigma^2)^2} \times \frac{|A_{11}|^{\frac{1}{2}}}{|\Sigma_{22}|^{\frac{1}{2}}\ |\sigma^2 I_\ell|^{\frac{1}{2}}} \times e^{-\frac{1}{2}\left[(\boldsymbol{\beta} - \boldsymbol{\mu}^*_\beta)'\,\Sigma_{22}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}^*_\beta)\right]} \\
& \qquad \times e^{-\frac{1}{2}\left[(\boldsymbol{\mu}^*_\nu - A_{11}^{-1}A_{12}(\boldsymbol{\beta} - \boldsymbol{\mu}^*_\beta))' A_{11}\left(\left(A_{11} + (\sigma^2 I_\ell)^{-1}\right)^{-1}\right)^{-1}\left(\sigma^2 I_\ell\right)^{-1}(\boldsymbol{\mu}^*_\nu - A_{11}^{-1}A_{12}(\boldsymbol{\beta} - \boldsymbol{\mu}^*_\beta))\right]} \qquad (15) \\
& \qquad \times e^{-\frac{1}{2}\left[\left[\boldsymbol{\nu} - \left(A_{11} + (\sigma^2 I_\ell)^{-1}\right)^{-1}(A_{11}\boldsymbol{\mu}^*_\nu - A_{12}(\boldsymbol{\beta} - \boldsymbol{\mu}^*_\beta))\right]'\left(A_{11} + (\sigma^2 I_\ell)^{-1}\right)\left[\boldsymbol{\nu} - \left(A_{11} + (\sigma^2 I_\ell)^{-1}\right)^{-1}(A_{11}\boldsymbol{\mu}^*_\nu - A_{12}(\boldsymbol{\beta} - \boldsymbol{\mu}^*_\beta))\right]\right]}.
\end{aligned}$$

From the above joint posterior density function (15), we see that $\boldsymbol{\nu}$ has the multivariate normal distribution

$$\boldsymbol{\nu} | \boldsymbol{\beta}, \alpha, \sigma^2 \sim N\left[\left(A_{11} + (\sigma^2 I_\ell)^{-1}\right)^{-1}(A_{11}\boldsymbol{\mu}^*_\nu - A_{12}(\boldsymbol{\beta} - \boldsymbol{\mu}^*_\beta)),\ \left(A_{11} + (\sigma^2 I_\ell)^{-1}\right)^{-1}\right]. \qquad (16)$$

There are numerous small areas. Integrating out $\boldsymbol{\nu}$, we have the joint density function of $\boldsymbol{\beta}, \alpha, \gamma, \sigma^2 | \boldsymbol{y}$ as follows:

$$\begin{aligned}
& f(\boldsymbol{\beta}, \alpha, \gamma, \sigma^2 | \boldsymbol{y}) \\
& \propto C_{\alpha\gamma}(\tau^*) \times \frac{e^{-R\gamma}\,\gamma^{S-1}}{(1+\alpha)^2\,(1+\sigma^2)^2} \times \frac{|A_{11}|^{\frac{1}{2}}\ \left|A_{11} + (\sigma^2 I_\ell)^{-1}\right|^{-\frac{1}{2}}}{|\Sigma_{22}|^{\frac{1}{2}}\ |\sigma^2 I_\ell|^{\frac{1}{2}}} \times e^{-\frac{1}{2}\left[(\boldsymbol{\beta} - \boldsymbol{\mu}^*_\beta)'\,\Sigma_{22}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}^*_\beta)\right]} \\
& \quad \times e^{-\frac{1}{2}\left[(\boldsymbol{\beta} - \widetilde{\boldsymbol{\mu}}_\beta)'\widetilde{\Sigma}(\boldsymbol{\beta} - \widetilde{\boldsymbol{\mu}}_\beta) - \widetilde{\boldsymbol{\mu}}_\beta'\widetilde{\Sigma}\widetilde{\boldsymbol{\mu}}_\beta + (\boldsymbol{\mu}^*_\nu + A_{11}^{-1}A_{12}\boldsymbol{\mu}^*_\beta)' S (\boldsymbol{\mu}^*_\nu + A_{11}^{-1}A_{12}\boldsymbol{\mu}^*_\beta)\right]},
\end{aligned}$$

where

$$\begin{aligned}
S &= A_{11}\left(A_{11} + (\sigma^2 I_\ell)^{-1}\right)^{-1}(\sigma^2 I_\ell)^{-1}, \\
\widetilde{\boldsymbol{\mu}}_\beta &= (A_{12}'A_{11}^{-1}SA_{11}^{-1}A_{12})^{-1}A_{12}'A_{11}^{-1}S\boldsymbol{\mu}^*_\nu + \boldsymbol{\mu}^*_\beta, \\
\widetilde{\Sigma}_\beta &= A_{12}'A_{11}^{-1}SA_{11}^{-1}A_{12}.
\end{aligned}$$

From the above joint density of $\boldsymbol{\beta}, \alpha, \gamma, \sigma^2 | \boldsymbol{y}$ we notice that $\boldsymbol{\beta}$ has a multivariate normal distribution

$$\boldsymbol{\beta} | \alpha, \gamma, \sigma^2, \boldsymbol{y} \sim N\left[\left(\Sigma_{22}^{-1} + \widetilde{\Sigma}_\beta\right)^{-1}\left(\Sigma_{22}^{-1}\boldsymbol{\mu}^*_\beta + \widetilde{\Sigma}_\beta\widetilde{\boldsymbol{\mu}}_\beta\right),\ \left(\Sigma_{22}^{-1} + \widetilde{\Sigma}_\beta\right)^{-1}\right]. \qquad (17)$$

Integrating out $\boldsymbol{\beta}$ from the above joint density function, we get the joint density of $\alpha, \gamma, \sigma^2 | \boldsymbol{y}$

$$
\begin{aligned}
&\pi(\alpha,\gamma,\sigma^2|\boldsymbol{y}) \\
&\quad \propto C_{\alpha\gamma}(\tau^*)\frac{e^{-R\gamma}\,\gamma^{S-1}}{(1+\alpha)^2\,(1+\sigma^2)^2}\,\frac{|A_{11}|^{\frac{1}{2}}\,\left|A_{11}+(\sigma^2 I_\ell)^{-1}\right|^{-\frac{1}{2}}\left|\Sigma_{22}^{-1}+\widetilde{\Sigma}_\beta\right|^{-\frac{1}{2}}}{|\Sigma_{22}|^{\frac{1}{2}}\,|\sigma^2 I_\ell|^{\frac{1}{2}}} \\
&\qquad \times e^{-\frac{1}{2}\left[(\boldsymbol{\mu}^*_\beta-\widetilde{\boldsymbol{\mu}}_\beta)'\Sigma_{22}^{-1}\left(\Sigma_{22}^{-1}+\widetilde{\Sigma}_\beta\widetilde{\boldsymbol{\mu}}_\beta\right)^{-1}\widetilde{\Sigma}_\beta(\boldsymbol{\mu}^*_\beta-\widetilde{\boldsymbol{\mu}}_\beta)\right]} \\
&\qquad \times e^{-\frac{1}{2}\left[-\widetilde{\boldsymbol{\mu}}'_\beta\widetilde{\Sigma}_\beta\widetilde{\boldsymbol{\mu}}_\beta+(\boldsymbol{\mu}^*_\nu+A_{11}^{-1}A_{12}\boldsymbol{\mu}^*_\beta)'S(\boldsymbol{\mu}^*_\nu+A_{11}^{-1}A_{12}\boldsymbol{\mu}^*_\beta)\right]}.
\end{aligned}
\tag{18}
$$

3.2 Approximation of Likelihood

The proposed models in this paper have posterior densities and conditional posterior densities in complex form, which makes parameter drawing tedious. To ease these difficulties, we have used the second-order Taylor's series approximation that will help us by providing approximated multivariate normal distribution for a large set of parameters. For more see Nandram, Fu and Manandhar (2017). Below, we provide the important results applicable for approximation.

Lemma: Let $\pi(\boldsymbol{\tau})$ be the unimodal density function. Then, $\boldsymbol{\tau}$ has an approximately multivariate normal distribution

$$\boldsymbol{\tau} \sim N\left\{\boldsymbol{\tau}^* - H^{-1}\boldsymbol{g},\ \ -H^{-1}\right\}, \tag{19}$$

where $\boldsymbol{\tau}^*, \boldsymbol{g},$ and H are the mode values, the gradient vector, and the Hessian matrix respectively of $\log\pi(\boldsymbol{\tau})$.

Multivariate Normal Approximation Theorem

Theorem 1. *Suppose $\Delta = G(\boldsymbol{\tau})$ is the log-likelihood function of unimodal density for the given data, response y_{ij} with corresponding covariates $\boldsymbol{x}_{ij},\ i=1,\cdots,\ell,\ j=1,\cdots,n_i$. Let $\boldsymbol{\tau}$ can be written as $(\boldsymbol{\beta}',\boldsymbol{\nu}')'$, then the joint posterior density of the parameters can be approximated by a multivariate normal density. Furthermore, the marginal density of $\boldsymbol{\beta}$ and the conditional density of $\boldsymbol{\nu}|\boldsymbol{\beta}$ can be approximated by multivariate normal densities.*

Proof. Given the log-likelihood function of $\boldsymbol{\tau}$ and $\Delta = G(\boldsymbol{\tau})$, let us write $\boldsymbol{\tau} = \begin{pmatrix}\boldsymbol{\nu}\\ \boldsymbol{\beta}\end{pmatrix}$, with the corresponding gradient vectors $\boldsymbol{g} = \begin{pmatrix}\boldsymbol{g_\nu}\\ \boldsymbol{g_\beta}\end{pmatrix}$ and Hessian matrix H, evaluated at the mode $\begin{pmatrix}\boldsymbol{\nu^*}\\ \boldsymbol{\beta^*}\end{pmatrix}$, then we have

$$\boldsymbol{g} = \left(\ \frac{\partial\Delta}{\partial\nu_1}\ \ \cdots\ \ \frac{\partial\Delta}{\partial\nu_\ell}\ \ \frac{\partial\Delta}{\partial\boldsymbol{\beta_0}}\ \ \cdots\ \ \frac{\partial\Delta}{\partial\boldsymbol{\beta_p}}\ \right)'\Big|_{\boldsymbol{\nu}=\boldsymbol{\nu}^*,\,\boldsymbol{\beta}=\boldsymbol{\beta}^*},$$

$$\boldsymbol{g_\nu} = \left(\ \frac{\partial\Delta}{\partial\nu_1}\ \ \cdots\ \ \frac{\partial\Delta}{\partial\nu_\ell}\ \right)'\Big|_{\boldsymbol{\nu}=\boldsymbol{\nu}^*,\,\boldsymbol{\beta}=\boldsymbol{\beta}^*},\quad \boldsymbol{g_\beta} = \left(\ \frac{\partial\Delta}{\partial\beta_0}\ \ \cdots\ \ \frac{\partial\Delta}{\partial\beta_p}\ \right)'\Big|_{\boldsymbol{\nu}=\boldsymbol{\nu}^*,\,\boldsymbol{\beta}=\boldsymbol{\beta}^*},$$

and

$$H = -\begin{pmatrix} A_{11} & A_{12}\\ A'_{12} & A_{22}\end{pmatrix}, \tag{20}$$

where

$$A_{11} = -\begin{pmatrix} \frac{\partial^2\Delta}{\partial{\nu_1}^2} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \frac{\partial^2\Delta}{\partial{\nu_\ell}^2}\end{pmatrix}, A_{12} = -\begin{pmatrix} \frac{\partial^2\Delta}{\partial\nu_1\partial\boldsymbol{\beta_0}} & \cdots & \frac{\partial^2\Delta}{\partial\nu_1\partial\boldsymbol{\beta_p}}\\ \vdots & \ddots & \vdots\\ \frac{\partial^2\Delta}{\partial\nu_\ell\partial\boldsymbol{\beta_0}} & \cdots & \frac{\partial^2\Delta}{\partial\nu_\ell\partial\boldsymbol{\beta_p}}\end{pmatrix}, A_{22} = \begin{pmatrix} \frac{\partial^2\Delta}{\partial\boldsymbol{\beta_0^2}} & \cdots & \frac{\partial^2\Delta}{\partial\boldsymbol{\beta_0}\partial\boldsymbol{\beta_p}}\\ \vdots & \ddots & \vdots\\ \frac{\partial^2\Delta}{\partial\boldsymbol{\beta_0}\partial\boldsymbol{\beta_p}} & \cdots & \frac{\partial^2\Delta}{\partial\boldsymbol{\beta_p^2}}\end{pmatrix}.$$

From the above Lemma, the multivariate normal approximation for the unimodal function, we have

$$\begin{pmatrix} \boldsymbol{\nu} \\ \boldsymbol{\beta} \end{pmatrix} \sim N\left\{\boldsymbol{\tau}^* - H^{-1}\boldsymbol{g}, -H^{-1}\right\}, \quad \boldsymbol{\tau}^* = \begin{pmatrix} \boldsymbol{\nu}^* \\ \boldsymbol{\beta}^* \end{pmatrix},$$

where $\boldsymbol{\tau}^*$ is the approximated mode, $\boldsymbol{g}$ and H are the gradient vector and the Hessian matrix evaluated at $(\boldsymbol{\nu}^*, \boldsymbol{\beta}^*)$ respectively. Then the approximated multivariate normal distribution of $\boldsymbol{\tau}$ can be written as

$$\begin{pmatrix} \boldsymbol{\nu} \\ \boldsymbol{\beta} \end{pmatrix} \sim N\left\{\begin{pmatrix} \boldsymbol{\nu^*} \\ \boldsymbol{\beta^*} \end{pmatrix} + \begin{pmatrix} A_{11} & A_{12} \\ A_{12}' & A_{22} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{g_\nu} \\ \boldsymbol{g_\beta} \end{pmatrix}, \begin{pmatrix} A_{11} & A_{12} \\ A_{12}' & A_{22} \end{pmatrix}^{-1}\right\}, \quad \text{which can be simplified as}$$

$$\begin{pmatrix} \boldsymbol{\nu} \\ \boldsymbol{\beta} \end{pmatrix} \sim N\left\{\begin{pmatrix} \boldsymbol{\mu^*_\nu} \\ \boldsymbol{\mu^*_\beta} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}\right\}, \tag{21}$$

where

$$\boldsymbol{\mu}^*_\nu = \boldsymbol{\nu}^* + \Sigma_{11}\boldsymbol{g_\nu} + \Sigma_{12}\boldsymbol{g_\beta}, \quad \text{and}$$
$$\boldsymbol{\mu}^*_\beta = \boldsymbol{\beta^*} + \Sigma_{12}'\boldsymbol{g_\nu} + \Sigma_{22}\boldsymbol{g_\beta}.$$

Now applying the multivariate normal theorem, we have

$$\boldsymbol{\beta}|\boldsymbol{y} \sim N\left(\boldsymbol{\mu}^*_\beta,\ \Sigma_{22}\right), \quad \text{and} \tag{22}$$
$$\boldsymbol{\nu}|\boldsymbol{\beta}, \boldsymbol{y} \sim N\left(\boldsymbol{\mu}^*_\nu + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}^*_\beta),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'\right). \tag{23}$$

□

3.3 Sampling from Joint Posterior Density

We have chosen the shape and rate parameters $S = R = 1$ for our prior distribution $\gamma \sim \text{Gamma}(S, R)$. Grid sampling and the Metropolis–Hastings sampling algorithm are used for drawing samples. For the Metropolis–Hastings algorithm, we use a multivariate *t*-distribution as our proposal distribution. We take $d = 3$ degrees of freedom so that variance will exist.

(i) We borrow α and γ from the previous model, *two generalized gamma mixture GB2 model without random area effects* ν_i. From these samples we pick a set of 100 quantiles by keeping the variable α and then γ in ascending order.

(ii) We draw $\sigma^2|\alpha, \gamma, \boldsymbol{y}$ using the grid sampling method with density function given by

$$\begin{aligned}\pi(\sigma^2|\alpha,\gamma,\boldsymbol{y}) \propto\ & \frac{1}{(1+\sigma^2)^2} \frac{|A_{11}|^{\frac{1}{2}}\ \left|A_{11} + (\sigma^2 I_\ell)^{-1}\right|^{-\frac{1}{2}} \left|\Sigma_{22}^{-1} + \widetilde{\Sigma}_\beta\right|^{-\frac{1}{2}}}{|\Sigma_{22}|^{\frac{1}{2}}\ |\sigma^2 I_\ell|^{\frac{1}{2}}} \\ & \times e^{-\frac{1}{2}\left[(\boldsymbol{\mu}^*_\beta - \widetilde{\boldsymbol{\mu}}_\beta)'\Sigma_{22}^{-1}\left(\Sigma_{22}^{-1} + \widetilde{\Sigma}_\beta\right)^{-1}\widetilde{\Sigma}_\beta(\boldsymbol{\mu}^*_\beta - \widetilde{\boldsymbol{\mu}}_\beta)\right]} \\ & \times e^{-\frac{1}{2}\left[-\widetilde{\boldsymbol{\mu}}_\beta'\widetilde{\Sigma}_\beta\widetilde{\boldsymbol{\mu}}_\beta + (\boldsymbol{\mu}^*_\nu + A_{11}^{-1}A_{12}\boldsymbol{\mu}^*_\beta)'S(\boldsymbol{\mu}^*_\nu + A_{11}^{-1}A_{12}\boldsymbol{\mu}^*_\beta)\right]}.\end{aligned} \tag{24}$$

The domain of σ^2 is $(0, \infty)$. So we transform σ^2 into η which has range $(0, 1)$, $\sigma^2 = \frac{\eta}{1-\eta}$. We took 100 grid values of η and computed transformed probability $\pi(\eta|\alpha, \boldsymbol{y})$ from (24). For each set of quantile values of α and γ we draw η and then transform it back to σ^2.

(iii) Using the information α, γ and $\sigma^2|\alpha, \gamma$ drawn above, we can draw $\boldsymbol{\beta}|\alpha, \gamma, \sigma^2, \boldsymbol{y}$. The Metropolis–Hastings algorithm is then used to draw $\boldsymbol{\beta}, \alpha, \gamma, \sigma^2|\boldsymbol{y}$ jointly. The proposal distributions are *t*-distributions. The proposal density for $\log(\alpha, \gamma, \sigma^2)|\boldsymbol{y}$ is the multivariate *t*-distribution with *d* degrees of freedom, $\log(\alpha, \gamma, \sigma^2) \sim \mathrm{t}_d(\boldsymbol{\mu}_{ln}, \Sigma_{ln})$, where $\boldsymbol{\mu}_{ln}$ and Σ_{ln} are estimated from the above step. The proposal distribution for $\boldsymbol{\beta}|\alpha, \gamma, \sigma^2, \boldsymbol{y}$ is a multivariate *t*-distribution with *d* degrees of freedom with corresponding mean and covariance matrix as in equation (17). The target density is as follows

$$
\begin{aligned}
&\pi(\boldsymbol{\beta}, \alpha, \gamma, \sigma^2|\boldsymbol{y}) \\
&\quad\propto \frac{\gamma^{S-1}}{(1+\alpha)^2\,(1+\sigma^2)^2} \left[\frac{\gamma\, g^{\alpha-1}}{\mathrm{B}(\frac{\alpha}{\gamma}, \frac{\alpha+2}{\gamma})}\right]^n e^{-(\alpha \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \boldsymbol{x'_{ij}}\boldsymbol{\beta} + R\gamma)} \\
&\quad\times \prod_{i=1}^{\ell} \left[\int_{\nu_i} \frac{e^{-\alpha\, n_i \nu_i}}{\prod_{j=1}^{n_i} \left(1 + \left[y_{ij}\, e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}\right)^{\frac{2(\alpha+1)}{\gamma}}} \left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}} e^{-\frac{\nu_i^2}{2\sigma^2}}\, d\nu_i\right].
\end{aligned}
$$

This integration is not in simple form. We apply a numerical integration. We divide the integration domain into m equal intervals $[t_k,\ t_{k-1}]$

$$
\begin{aligned}
&\pi(\boldsymbol{\beta}, \alpha, \gamma, \sigma^2|\boldsymbol{y}) \\
&\propto \frac{\gamma^{S-1}}{(1+\alpha)^2\,(1+\sigma^2)^2} \left[\frac{\gamma\, g^{\alpha-1}}{\mathrm{B}(\frac{\alpha}{\gamma}, \frac{\alpha+2}{\gamma})}\right]^n e^{-(\alpha \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \boldsymbol{x'_{ij}}\boldsymbol{\beta} + R\gamma)} \\
&\times \prod_{i=1}^{\ell} \left[\sum_{k=1}^{m} \frac{e^{-n_i \hat{z}_k \sigma}}{\prod_{j=1}^{n_i} \left(1 + \left[y_{ij}\, e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\hat{z}_k\sigma)}\right]^{\gamma}\right)^{\frac{2(\alpha+1)}{\gamma}}} \times (\Phi(t_k) - \Phi(t_{k-1}))\right].
\end{aligned}
$$

(iv) Parameters $\nu_i|\boldsymbol{\beta}, \alpha, \gamma, \sigma^2$ are drawn using the Metropolis–Hastings algorithm. The proposal density is a *t*-distribution with *d* degrees of freedom. We take the mean and variance for the proposal from the samples of ν_i while drawing $(\boldsymbol{\beta}, \alpha, \gamma, \sigma^2)$ in the above step. The target density is

$$
\pi(\nu_i|\boldsymbol{\beta}, \alpha, \gamma, \sigma^2) \propto \frac{e^{-\left(\alpha\, n_i \nu_i + \frac{\nu_i^2}{2\sigma^2}\right)}}{\prod_{j=1}^{n_i} \left(1 + \left[y_{ij}\, e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}\right)^{\frac{2(\alpha+1)}{\gamma}}}, \quad i = 1, \cdots, \ell.
$$

We keep the samples ν_i from the Metropolis–Hastings algorithm if the acceptance rate lies between 0.25 and 0.50, otherwise we discard them and sample again using the grid sampling method in the second attempt.

3.4 Prediction

After drawing all parameters from the GB2 distribution model as mentioned above, we predict the responses as follows.

(i) Find the rate parameters θ. We calculate the rate parameter using the information on random area effect ν_i and $\boldsymbol{\beta}$ as follows

$$
\theta_{ij} = e^{\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i}.
$$

(ii) Draw shape parameters λ. In the GB2 distribution let us consider a transformation $t = (\theta\lambda)^\gamma$. This gives $(\theta\lambda)^\gamma \sim \text{Gamma}\left(\frac{\alpha+2}{\gamma}, 1\right)$. Say, we draw a random sample G_1 from this distribution, and then we can calculate λ as follows

$$\mathrm{G}_1 = (\theta\lambda)^\gamma \sim \text{Gamma}\left(\frac{\alpha+2}{\gamma}, 1\right), \qquad \lambda_{ij} = \frac{\mathrm{G}_1^{\frac{1}{\gamma}}}{\theta_{ij}}.$$

(iii) Predict responses. In the GB2 distribution we consider a transformation $t = (\lambda y)^\gamma$. This gives $(\lambda y)^\gamma \sim \text{Gamma}\left(\frac{\alpha}{\gamma}, 1\right)$. Say we draw a random sample G_2 from this distribution; then we can predict $\hat{y}$ as follows:

$$\mathrm{G}_2 = (\lambda_{ij} y)^\gamma \sim \text{Gamma}\left(\frac{\alpha}{\gamma}, 1\right), \qquad \hat{y}_{ij} = \frac{\mathrm{G}_2^{\frac{1}{\gamma}}}{\lambda_{ij}}.$$

4. Application and Simulation

We have sampled the parameters using the grid sampling and the MCMC Metropolis–Hastings (MH) sampling method. We have applied the MH algorithm more than once; however we have tabulated the acceptance rates for the final MH algorithm only. For all the fitted models, we have taken a set of 2100 samples, “burn-in” 100 samples and thinning interval of one. The final set has 1000 samples. For model-comparison purposes, we have calculated LPML values. The larger the value of LPML the better the model.

Table 1 presents LPML values for GB2 models This table shows, that the mixture of the exponential and gamma models have much smaller LPML values compared to the mixture of two gamma and the mixture of two generalized gamma GB2 models. Therefore obviously the mixture of two gammas or the mixture of two generalized gammas GB2 model fits better in NLSS-II consumption data. The selected best fitted model is GB2 as a mixture of two generalized gamma distributions. For stratum two, the LPML values for GB2 models, the mixture of exponential and gamma, two gamma, and two generalized gamma distributions models are respectively -741.5, -614.7, and -591.8.

Table 2 presents the parameters-assignment results. It provides the MH sampler acceptance rate for parameters, Geweke convergence diagnostic test, and effective sample sizes. The acceptance rate for parameters $(\alpha, \gamma, \boldsymbol{\beta}, \sigma^2)$ for models are provided. This table has every p-value greater than 0.05. The effective sample sizes for all parameters are unity except a few which have less (about 0.80 to 0.90), and a few are a little larger than unity.

Table 1: *LPML values for three GB2 models*

Stratum	**GB2 Models**		
	Exponential-Gamma	**Gamma-Gamma**	**GGamma-GGamma**
1	-487.6	-195.2	-173.9
2	-741.5	-614.7	-591.8
3	-509.6	-383.2	-362.1
4	-1479.3	-809.4	-756.3
5	-530.8	-372.5	-349.0
6	-1766.2	-1063.5	-1008.0

Table 2: *Metropolis–Hastings acceptance rates, Geweke test p-values and effective sample sizes*

Models with area effects

Stratum	Alpha, Gamma, Beta, Sigma2 acceptance rate	P-values												
		Alpha	**Gamma**	**Sigma2**	**Beta0**	**Beta1**	**Beta2**	**Beta3**	**Beta4**	**Beta5**	**Beta6**	**Beta7**	**Beta8**	**Beta9**
1	0.526	0.11	0.10	0.61	0.26	0.19	0.11	0.06	0.90	0.66	0.39	0.18	0.14	0.34
2	0.547	0.30	0.22	0.27	0.24	0.79	0.48	0.16	0.62	0.24	0.18	0.37	0.29	0.25
3	0.533	0.81	0.99	0.46	0.18	0.68	0.45	0.64	0.56	0.98	0.38	0.18	0.35	0.60
4	0.513	0.84	0.90	0.60	0.83	0.05	0.08	0.70	0.02	0.81	0.07	0.15	0.69	0.89
5	0.545	0.58	0.80	0.49	0.42	0.91	0.34	0.36	0.09	0.61	0.38	0.79	0.73	0.61
6	0.501	0.32	0.23	0.97	0.66	0.71	0.49	0.78	0.92	0.10	0.14	0.05	0.14	0.56

Stratum	Effective sample size												
	Alpha	**Gamma**	**Sigma2**	**Beta0**	**Beta1**	**Beta2**	**Beta3**	**Beta4**	**Beta5**	**Beta6**	**Beta7**	**Beta8**	**Beta9**
1	1.00	1.00	1.10	1.00	1.00	1.00	1.00	1.12	1.00	1.00	1.00	1.33	1.00
2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.10	1.00	1.09	1.11	1.00	1.00
3	0.79	0.77	1.00	1.00	1.00	1.00	0.85	1.00	1.00	1.00	1.00	1.00	1.00
4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.15	1.00	1.12	1.03
5	0.84	0.83	1.00	1.00	1.00	1.00	1.00	1.00	0.88	0.93	1.00	1.00	1.00
6	1.00	1.00	1.00	1.14	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.37	0.91

Now, we show the trace and correlation plots. The trace plots for parameters are shown for the *Mountains stratum* (stratum one). The trace plots for parameters alpha, gamma, sigma square, and vector of beta coefficients are shown from Figure 1 to Figure 13. The correlation plots for parameters alpha, gamma, sigma square, and vector of beta coefficients are shown from Figure 14 to Figure 26. The trace and correlation plots for other strata are also similar and not shown here.

We have applied models to the welfare per capita consumption data with nine covariates from NLSS-II. To facilitate SAE, we have nine covariates both in the NLSS-II and the population census, and their consistencies were checked prior to use. Poverty indicators have been calculated using the poverty threshold of an average of 7,696 Nepalese rupees per capita per year in 2003, adjusted for spatial price variation as reported in NLSS-II documents. It is the same poverty threshold used in SAE of Poverty, Nepal (Haslett et al., 2006).

We present the SAE of poverty indicators (poverty incidence, poverty gap, and poverty severity) by applying the selected mixture of two generalized gamma (GB2) models in the 2001 Population Census data. We have given the district level estimates for *Mountains stratum* as an example. Estimates for all other strata can be calculated similarly. The small area estimations for VDC/Municipalities and for wards are also done in a similar way and not tabulated here.

4.1 Family of Poverty Measures

Let E_{ij} be the welfare measure for the j^{th} unit of the i^{th} area, $i = 1, \cdots, A,\ \ j = 1, \cdots, N_i$. The family of poverty measures for small area i given the predetermined poverty threshold $z > 0$ is given by

$$P_{\alpha i} = \frac{1}{N_i} \sum_{j=1}^{N_i} \left(\frac{z - E_{ij}}{z} \right)^{\alpha} I\left(E_{ij} < z\right), \qquad \alpha \geq 0, \quad i = 1, \cdots, A, \tag{25}$$

where, $I(E_{ij} < z)$ is an indicator function (Foster, Greer, & Thorbecke, 1984). For, $\alpha = 0, \mathrm{P}_{0i}$ gives poverty incidence, the proportion of poor, $\mathrm{P}_{1i}, (\alpha = 1)$ gives the poverty gap and $\mathrm{P}_{2i}, (\alpha = 2)$ gives the poverty severity. The larger the value of the parameter α the greater emphasis given to the poorest poor.

4.2 Small Area Estimation

For SAE, first we get the area effects, $\nu_i, i = 1 \cdots, L$. We draw the random area effects from the Bayesian bootstrap sampling method with prior Dirichlet($\mathbf{0}$) given all other parameters. Once we draw the random area effects, ν_i, we predict the responses in the population census data as follows:

Prediction

(i) Find the rate parameters
$$\theta_{ij} = e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \nu_i}.$$

(ii) Draw shape parameter λ. In GB2 distribution consider a transformation $t = (\theta\lambda)^{\gamma}$. This gives $(\theta\lambda)^{\gamma} \sim$ Gamma $\left(\frac{\alpha+2}{\gamma}, 1\right)$. If we draw a random sample G_1 from this distribution, then we can calculate λ as
$$\mathrm{G}_1 = (\theta\lambda)^{\gamma} \sim \text{Gamma}\left(\frac{\alpha+2}{\gamma}, 1\right), \quad \lambda_{ij} = \frac{\mathrm{G}_1^{\frac{1}{\gamma}}}{\theta_{ij}}.$$

(iii) Predict responses. In the GB2 distribution consider a transformation $t = (\lambda y)^\gamma$. This gives $(\lambda y)^\gamma \sim$ Gamma $\left(\frac{\alpha}{\gamma}, 1\right)$. If we draw a random sample G_2 from this distribution, then we can predict $\hat{y}$ as

$$\mathrm{G}_2 = (\lambda y)^\gamma \sim \text{Gamma}\left(\frac{\alpha}{\gamma}, 1\right), \quad \hat{y}_{ij} = \frac{\mathrm{G}_2^{\frac{1}{\gamma}}}{\lambda_{ij}}.$$

After predicting the responses, the family of poverty measures for small area i with poverty threshold of z is given by $P_{\alpha i}$, equation (25).

4.3 Small Area Estimation at District Level

We provide the SAE for poverty indicators at the district level for the *Mountains stratum* as an example. We have also estimated indicators in municipalities/VDC level and ward levels but are not tabulated here; there are large numbers of those small areas. We have mapped the poverty indicators at the district level. It would be better to provide maps of these indicators in the municipality/VDC level. [Unfortunately we do not have *shape files* in the municipality/VDC level for mapping.]

Table 3 provides Mountains stratum district level poverty indicators, their standard errors and highest posterior density intervals at district levels. The poverty rate estimate is 0.364 (SE 0.035), poverty gap estimate is 0.102 (SE 0.013) and poverty severity estimate is 0.041 (SE 0.006) at the Mountatins stratum. There are seven districts which have higher poverty incidences than average stratum incidences. Poverty rates in those districts are: Mustang 0.411 (SE 0.036), Sankhuwasabha 0.393 (SE 0.036), Taplejung 0.391 (SE 0.036), Rasuwa 0.387 (SE 0.036), Darchula 0.384 (SE 0.038), Manang 0.379 (SE 0.038), and Kalikot 0.376 (SE 0.037). The four largest poverty gaps estimated districts are Mustang 0.132 (SE 0.014), Rasuwa 0.117 (SE 0.014), Taplejung 0.112 (SE 0.014) and Sankhuwasabha 0.112 (SE 0.014). Similarly four largest poverty severities estimated districts are Mustang 0.063 (SE 0.007), Rasuwa 0.051 (SE 0.007), Taplejung 0.045 (SE 0.007), and Sankhuwasabha 0.045 (SE 0.007).

4.4 Simulation Study

In the census data, covariates are available, but it does not have responses. We simulate the response values in the census data with a multivariate linear regression using the nine covariates as we have used for the model building. To simulate responses in the census, we fit the multivariate linear regression model with these nine covariates in NLSS-II data with log-transformed responses

$$\log(y_{ij}) = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + e_{ij},\ i = 1, \cdots, \ell,\ j = 1, \cdots, n_i.$$

Then we predict simulated responses in the census data

$$y_{ij}^{(s)} = e^{\boldsymbol{x}'_{ij}\hat{\boldsymbol{\beta}} + \hat{e}_{ij}},\ i = 1, \cdots, \mathrm{L},\ j = 1, \cdots, N_i,$$

where $\hat{e}_{ij}$ are generated from the assumption that residuals are distributed normally.

After generating simulated responses in the census data, we draw samples of size n from the census data with simulated responses. We picked the same wards for the simulated sample as in the NLSS-II data and the same number of households (12 households) by systematic random sampling as it was done in the NLSS-II. There are four Wards of NLSS-II (Mountains stratum), two wards from the Dolakha *(ward codes 2200606 and 2204407)* district and one ward from the Sankhuwasabha *(ward code 901303)* and Kalikot *(ward code 6400401)* districts, where the 2001 population census was unable to enumerate because of a Maoist insurgency at that time in those wards. For those wards where census data are not available, we randomly replace with the next ward from the same district. So in total we have 384 households from 32 wards in our simulated samples, as in NLSS-II.

We fit our selected model to the simulated sample data and predict the responses in the census data. If the models fit well, then they should predict the poverty indicators well. We have calculated the three poverty indicators for both simulated responses and predicted responses.

Figures 30 through 38 show diagonal plots of the poverty indicators from the simulation study in district, municipality/VDC and ward levels. The census poverty indicators are calculated from the simulated responses and the predicted poverty indicators are calculated from the model we have fitted. We have also provided the linear relationship between census simulated indicators and model predicted indicators with their R^2 values in their respective plots.

Figure 30 contains simulation study poverty incidences for the district level. The R^2 value of the linear relationship is 0.926. Figure 31 is simulation study poverty incidences for municipality/VDC level. The R^2 value of the linear relationship is 0.843. Figure 32 is simulation study poverty incidences at the ward level. The R^2 values of the linear relationship is 0.659.

Figures 33 through 35 show diagonal plots of the poverty gaps from the simulated responses versus predicted poverty gaps from the fitted models in district, municipality/VDC and ward levels. We have also provided the linear relationship between the simulated response gaps in the census data and predicted responses gap by the fitted model with the R^2 values in their respective plots. Figures 36 through 38 show diagonal plots of the poverty severities from the simulated responses in the census data versus predicted poverty severities from the fitted model in district, municipality/VDC and ward levels. We have provided the linear relationship value R^2. In Figure 36, severities at the district level, there are some observations below the diagonal line, though it is up-lifted.

5. Concluding Remarks

We propose the GB2 models for the noisy responses. We assume that in our responses, the noise has been introduced as recalling errors. We have fitted three hierarchical Bayesian models and the mixture of two generalized gamma GB2 model has been selected as the best model. Our response variable is CPS and the logarithmic transformation is suggest. However, to avoid the problem due to logarithmic transformation, we have fitted the models without a logarithmic transformation.

We have applied our models to the CPS consumption data from NLSS-II, 2003/04 survey and provided the estimates for small areas. SAE estimation is needed for planning and research purposes. To provide the SAE we have used population census data, 2001 where we do have information of the covariates but not the responses. We have predicted the responses in the census data using the best fitted model.

Table 3: *Poverty incidences by districts, SAE 2001*

District	Region	Poverty Rates				Poverty Gaps				Poverty Sevirities			
		Rate	**SE**	**Hpd lower**	**Hpd Upper**	**Gap**	**SE**	**Hpd lower**	**Hpd Upper**	**Severity**	**SE**	**Hpd lower**	**Hpd Upper**
Taplejung	Eastern	0.391	0.036	0.324	0.463	0.112	0.014	0.083	0.137	0.045	0.007	0.032	0.058
Sankhuwasabha	Eastern	0.393	0.036	0.326	0.464	0.112	0.014	0.083	0.137	0.045	0.007	0.032	0.058
Solukhumbu	Eastern	0.327	0.036	0.257	0.395	0.086	0.013	0.061	0.110	0.033	0.006	0.022	0.044
Dolakha	Central	0.351	0.034	0.284	0.418	0.097	0.013	0.072	0.121	0.039	0.006	0.028	0.051
Sindhupalchok	Central	0.359	0.035	0.295	0.426	0.101	0.013	0.078	0.127	0.041	0.006	0.030	0.053
Rasuwa	Central	0.387	0.036	0.322	0.458	0.117	0.014	0.092	0.148	0.051	0.007	0.038	0.065
Manang	Western	0.379	0.038	0.313	0.456	0.106	0.014	0.082	0.138	0.042	0.007	0.030	0.057
Mustang	Western	0.411	0.036	0.348	0.489	0.132	0.014	0.104	0.160	0.063	0.007	0.049	0.078
Dolpa	Mid Western	0.335	0.035	0.264	0.399	0.090	0.013	0.067	0.117	0.036	0.006	0.024	0.048
Jumla	Mid Western	0.352	0.038	0.282	0.427	0.094	0.014	0.069	0.123	0.037	0.006	0.024	0.049
Kalikot	Mid Western	0.376	0.037	0.294	0.439	0.103	0.014	0.075	0.130	0.041	0.007	0.028	0.054
Mugu	Mid Western	0.357	0.035	0.292	0.428	0.098	0.013	0.074	0.123	0.039	0.006	0.027	0.051
Humla	Mid Western	0.351	0.036	0.285	0.422	0.095	0.013	0.073	0.122	0.037	0.006	0.026	0.049
Bajura	Far Western	0.352	0.035	0.285	0.417	0.095	0.013	0.071	0.120	0.037	0.006	0.026	0.049
Bajhang	Far Western	0.360	0.037	0.295	0.436	0.098	0.014	0.075	0.127	0.038	0.006	0.025	0.050
Darchula	Far Western	0.384	0.038	0.308	0.458	0.109	0.015	0.081	0.137	0.044	0.007	0.030	0.057
Mountains		**0.364**	**0.035**	**0.291**	**0.430**	**0.102**	**0.013**	**0.075**	**0.128**	**0.041**	**0.006**	**0.029**	**0.054**

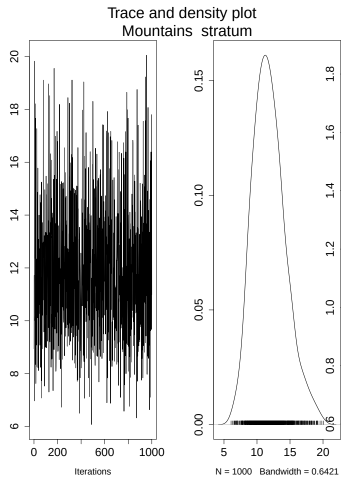

Figure 1: Alpha

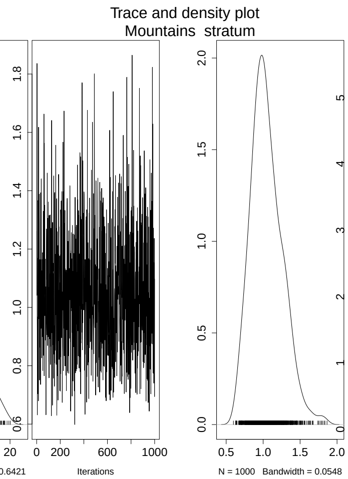

Figure 2: Gamma

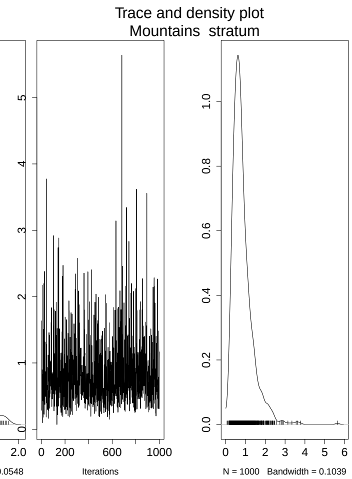

Figure 3: Sigma Square

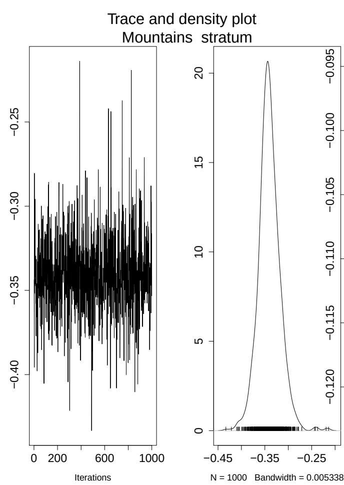

Figure 4: Beta0

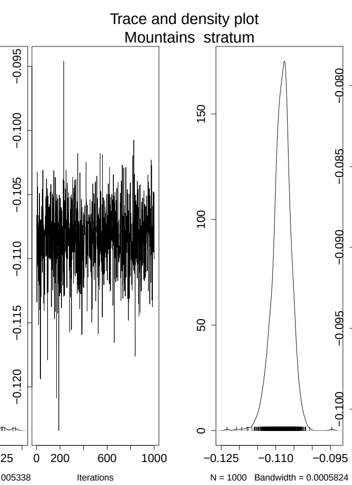

Figure 5: Beta1

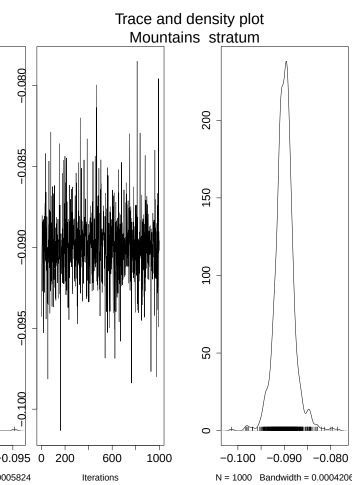

Figure 6: Beta2

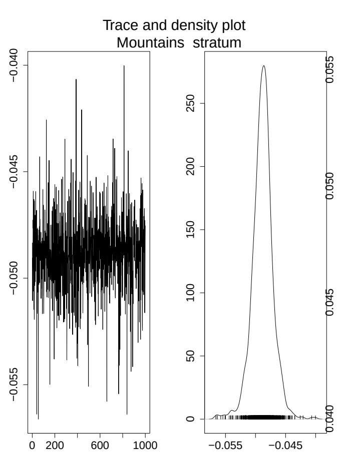

Figure 7: Beta3

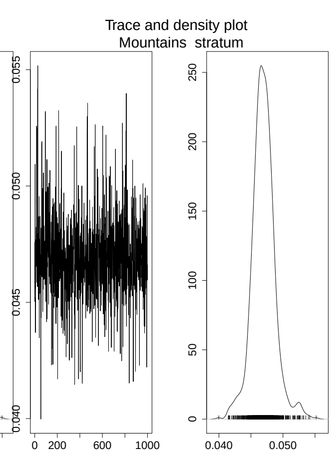

Figure 8: Beta4

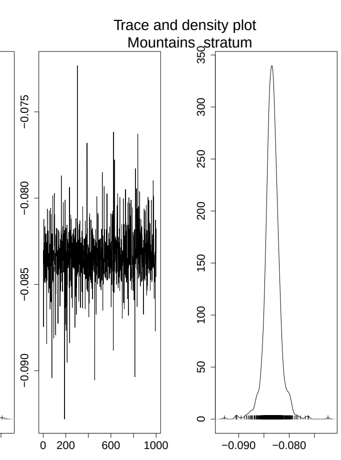

Figure 9: Beta5

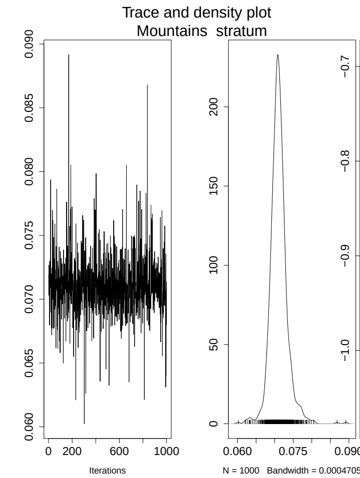

Figure 10: Beta6

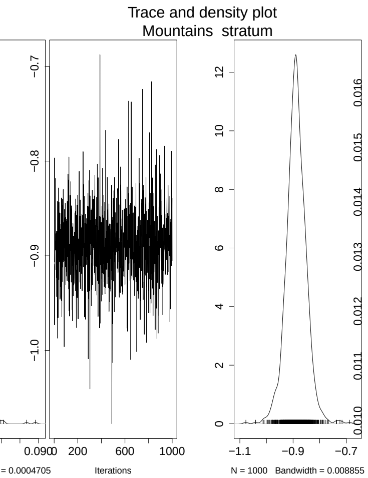

Figure 11: Beta7

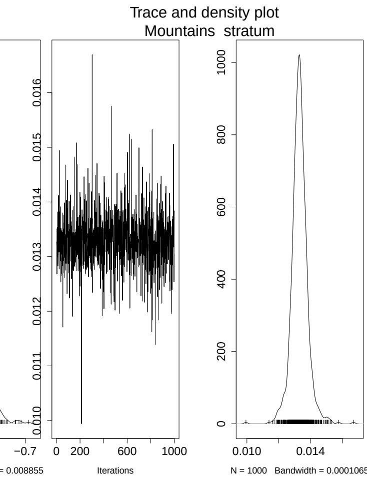

Figure 12: Beta8

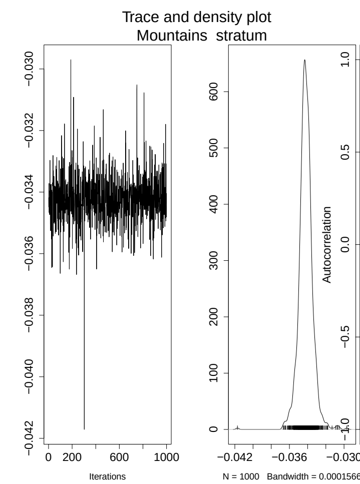

Figure 13: Beta9

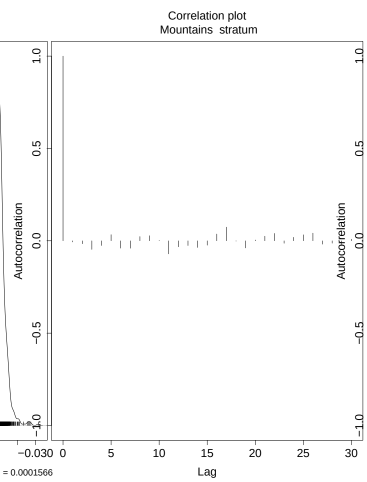

Figure 14: Alpha

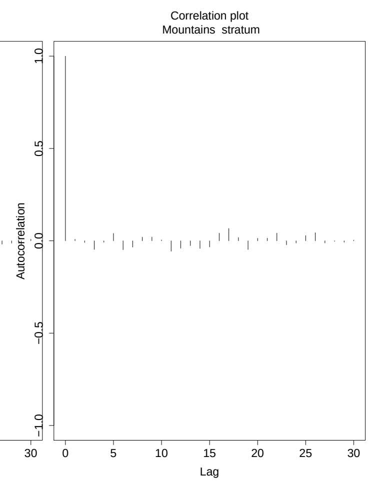

Figure 15: Gamma

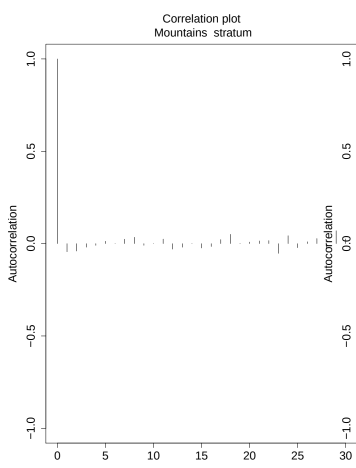

Figure 16: Sigma Square

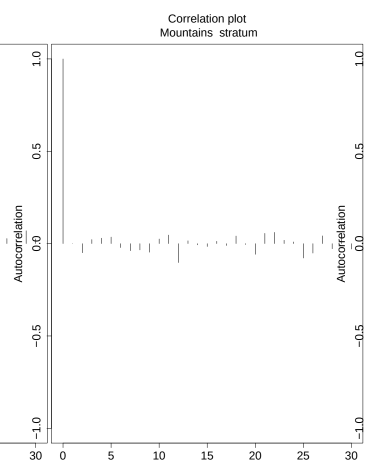

Figure 17: Beta0

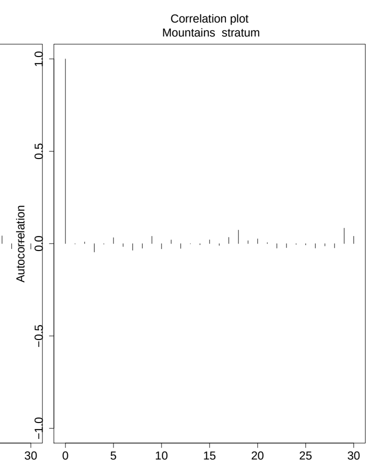

Figure 18: Beta1

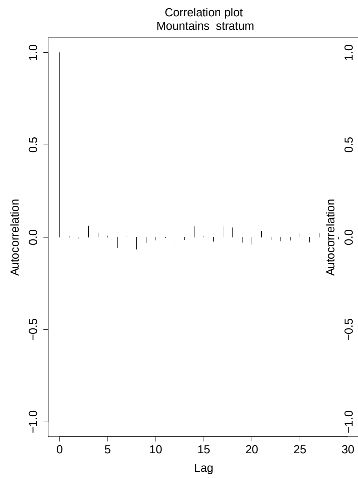

Figure 19: Beta2

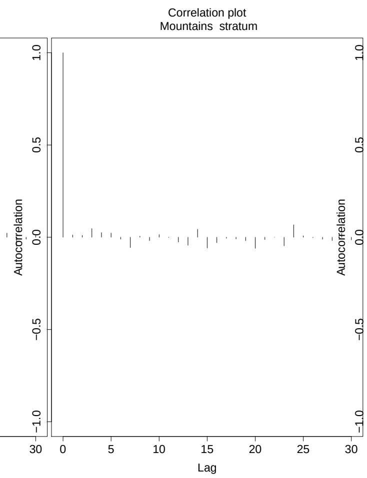

Figure 20: Beta3

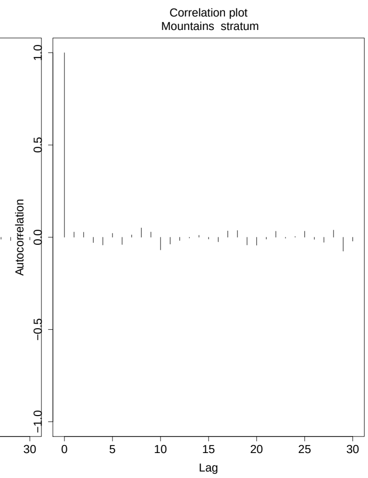

Figure 21: Beta4

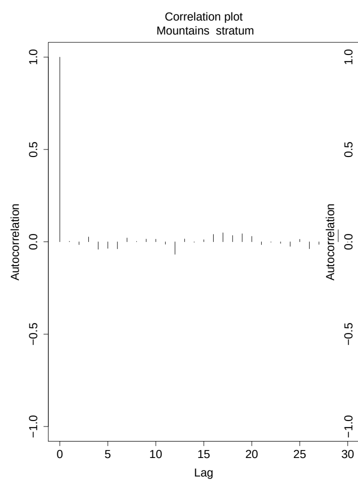

Figure 22: Beta5

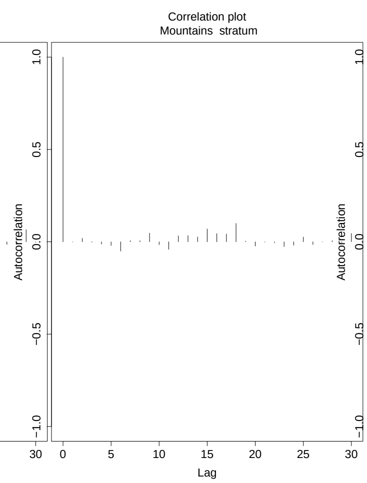

Figure 23: Beta6

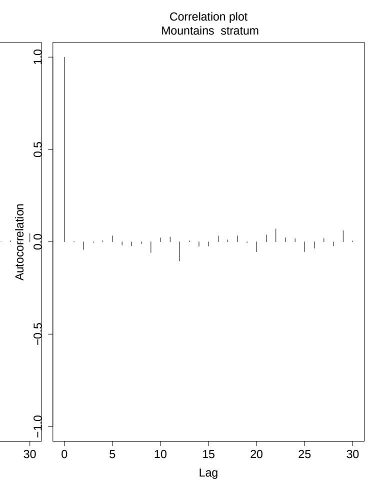

Figure 24: Beta7

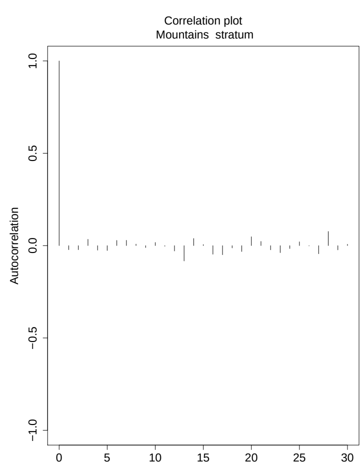

Figure 25: Beta8

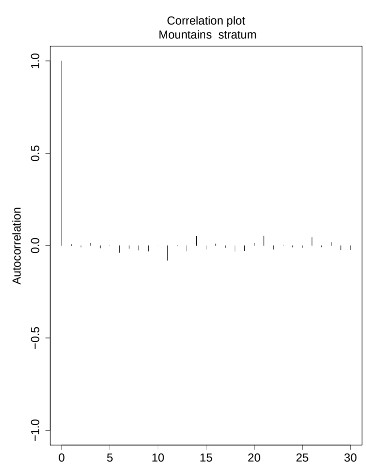

Figure 26: Beta9

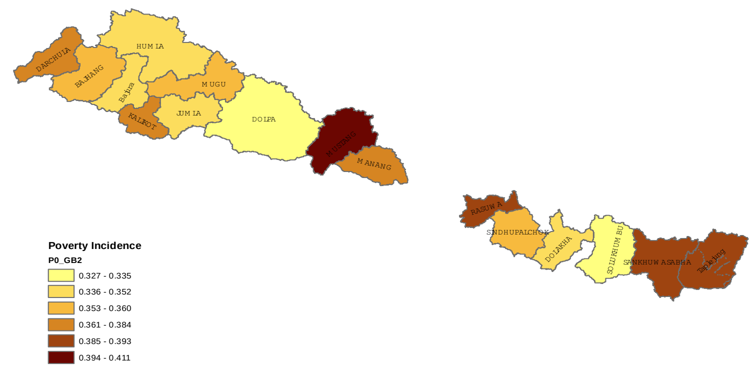

Figure 27: Poverty incidence (P0) at the district level (Mountains stratum)

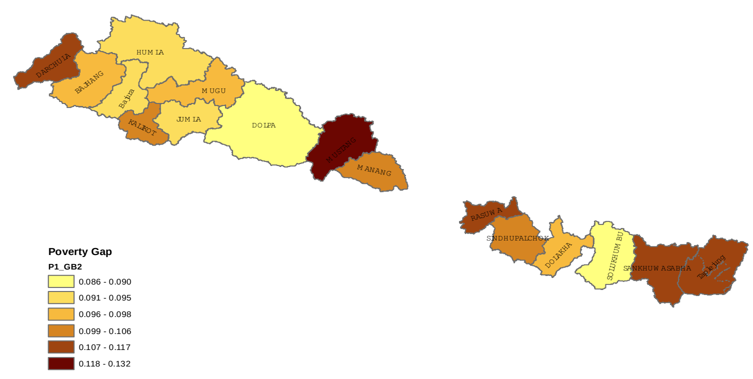

Figure 28: Poverty gap (P1) at the district level (Mountains stratum)

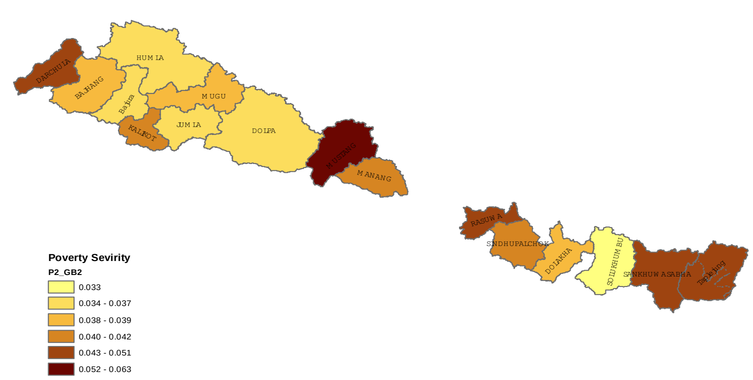

Figure 29: Poverty severity (P2) at the district level (Mountains stratum)

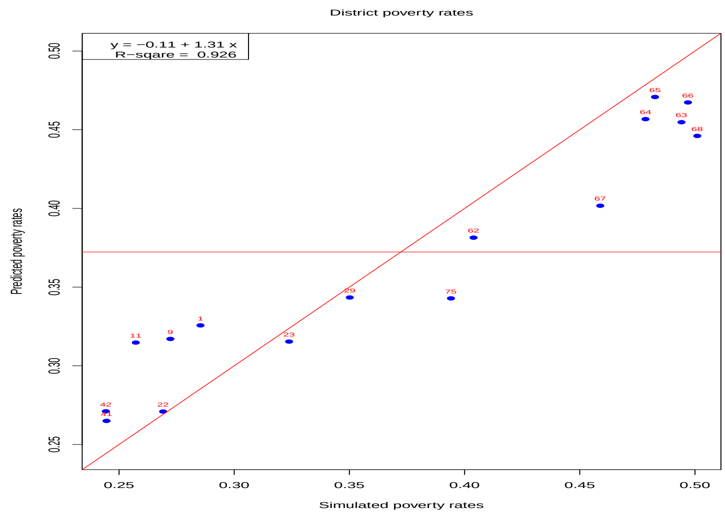

Figure 30: Poverty incidences in the simulation study by district (Mountains stratum)

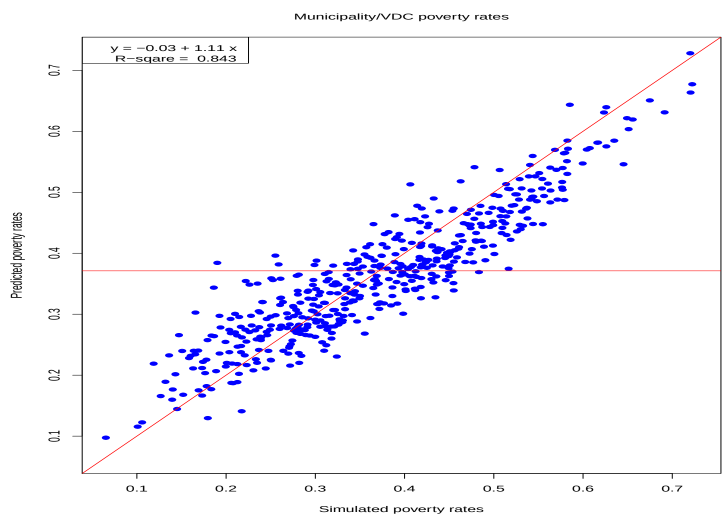

Figure 31: Poverty incidences in the simulation study by municipality/VDC (Mountains stratum)

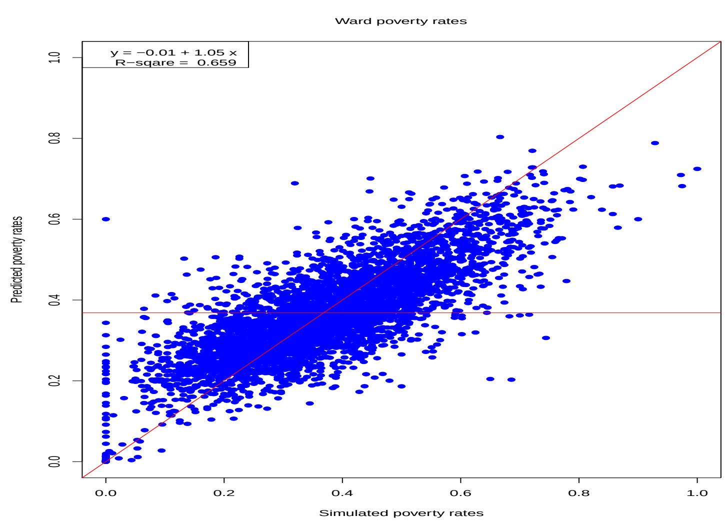

Figure 32: Poverty incidences in the simulation study by ward (Mountains stratum)

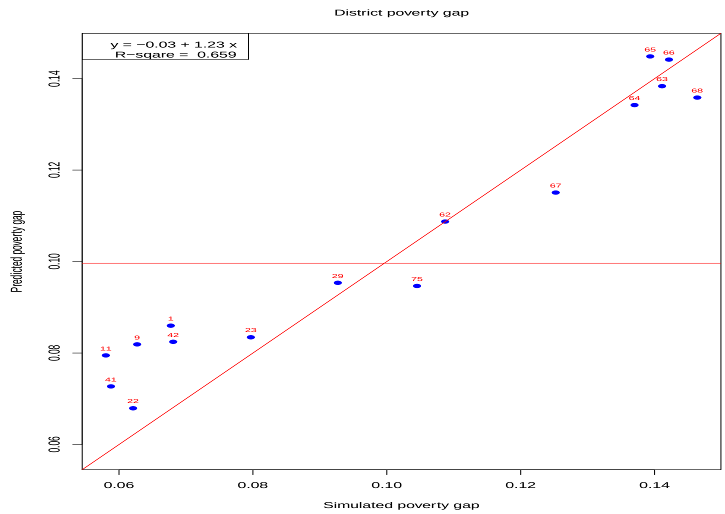

Figure 33: Poverty gaps in the simulation study by district (Mountains stratum)

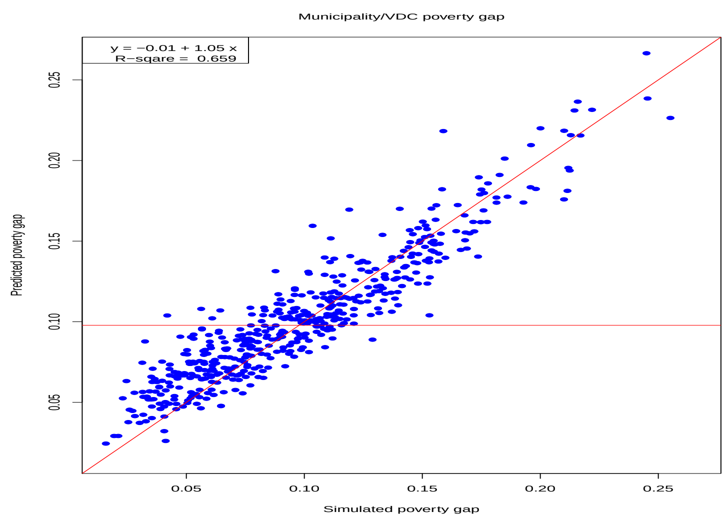

Figure 34: Poverty gaps in the simulation study by municipality/VDC (Mountains stratum)

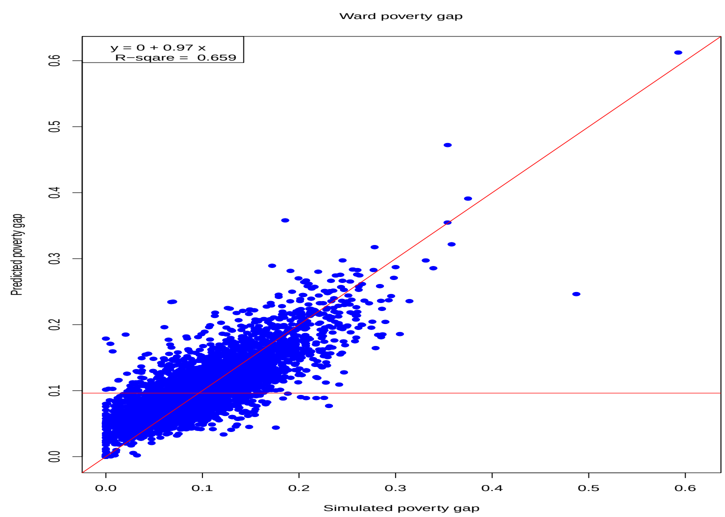

Figure 35: Poverty gaps in the simulation study by ward (Mountains stratum)

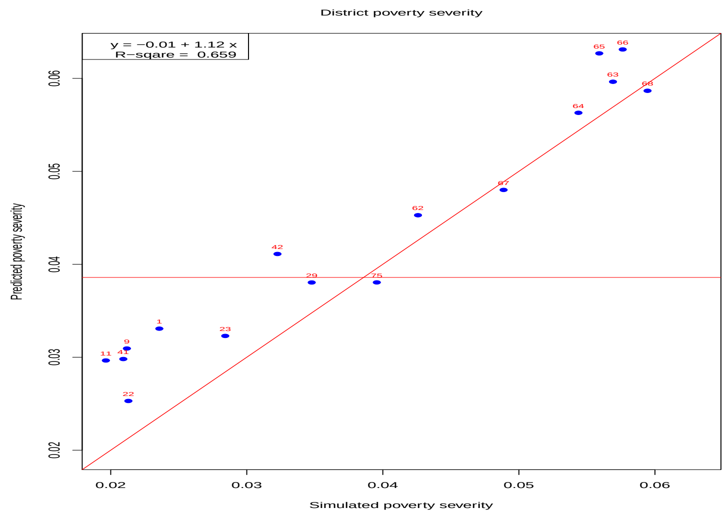

Figure 36: Poverty severities in the simulation study by district (Mountains stratum)

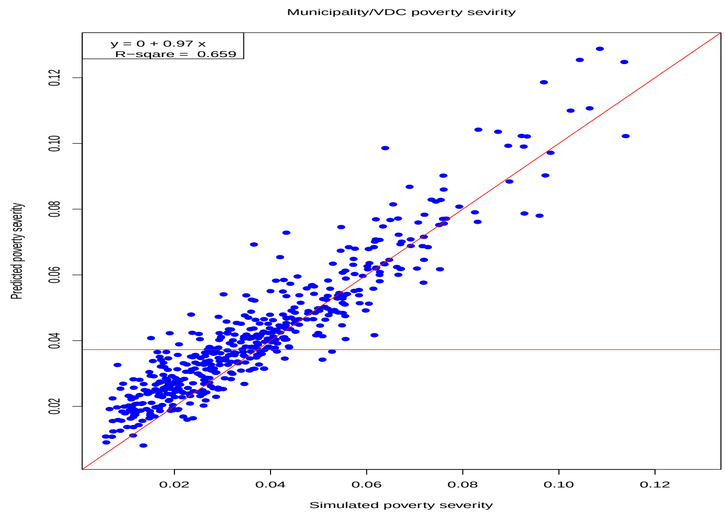

Figure 37: Poverty severities in the simulation study by municipality/VDC (Mountains stratum)

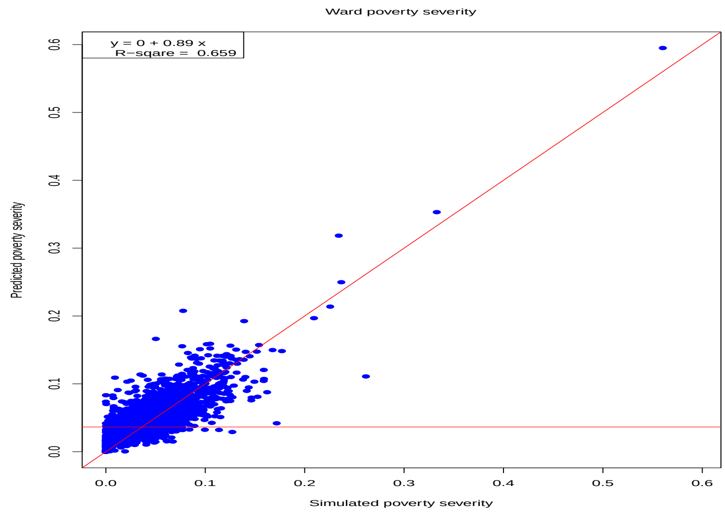

Figure 38: Poverty severities in the simulation study by ward (Mountains stratum)

REFERENCES

Battese, G. E., Harter, R. M., 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*, 83.401, 28-36.

Berliner, L. M (1996), “Hierarchical Bayesian time series models,” *Maximum entropy and Bayesian methods*, 15-22, Springer.

Besag, J. (1974), “Spatial interaction and the statistical analysis of lattice systems,” *Journal of the Royal Statistical Society*, Series B (Methodological), 192-236.

Box, G. E. (1980), “Sampling and Bayes’ Inference in Scientific Modelling and Robustness,” *Journal of the Royal Statistical Society*, Series A (General), 4, 383-430, Royal Statistical Society, Wiley, ISSN 00359238.

Central Bureau of Statistics, Nepal, “Nepal living standards survey report 2003/04,” *CBS, Nepal*, Vol 1.

Central Bureau of Statistics, Nepal, “Nepal living standards survey report 2003/04,” *CBS, Nepal*, Vol 2.

Central Bureau of Statistics, Nepal, “Poverty trends in Nepal (1995-96 and 2003-04),” *CBS Nepal*.

Cressie, N., and Wikle, C. K. (2015), “Statistics for spatio-temporal data,” *John Wiley & Sons*.

Dong, A. X., and Chan, J. S. K. (2013), “Bayesian analysis of loss reserving using dynamic models with generalized beta distribution,” *Insurance: Mathematics and Economics*, 53, 2, 355-365.

Feng, C., Wang, H., Lu, N., and Tu, X. M. (2013), “Log transformation: application and interpretation in biomedical research,” *Statistics in medicine*, 32, 230-239.

Feng, C., Wang, H., Lu, N., Chen, T., He, H., Lu, Y., and Tu, X. M. (2014), “Log-transformation and its implications for data analysis,” *Shanghai archives of psychiatry*, 26, 105-109.

Foster, J., Greer, J., and Thorbecke, E. (1984), “A Class of Decomposable Poverty Measures,” *Econometrica*, 3, 761-766, Wiley, Econometric Society, 52.

Gelfand, A. E., Dey, D. K., and Chang, H. (1992), “Model determination using predictive distributions with implementation via sampling-based methods,” *DTIC Document*.

Ghosh, M. and Rao, J. N. K. (1994), “Small Area Estimation: An Appraisal,” *Statistical Science*, 9, 55-76.

Graf, M. and Nedyalkova, D. (2014), “Modeling of income and indicators of poverty and social exclusion using the generalized beta distribution of the second kind,” *Review of Income and Wealth*, 60, 821-842.

Haslett, S. J., and Jones, G. (2006), “Small area estimation of poverty, caloric intake and malnutrition in Nepal,” *Nepal Central Bureau of Statistics/World Food Programme, United Nations/World Bank*, Kathmandu, 184, 999337018-5.

Manandhar, B. (2009), “Statistics of Newar and Nepalmandal Territory: Analyzing National Population Censuses Data 1991-2001 of Newaa Civilization,” *Ilohan Publication*, Kathmandu Nepalmandal, ISBN: 978-9937-9015-2-9

Manandhar, B. and Nandram, B. (2016), “Bayesian predictive inference for consumption data from small areas,” *In JSM 2016 proceedings*, Section on Bayesian Statistical Science, Alexandria, VA: American Statistical Association, 2437-2446.

Manandhar, B. and Nandram, B. (2017a), “Hierarchical Bayesian models for size responses from small areas: An application to poverty estimation,” *(under preparation)*.

Molina, I., and Rao, J. N. K. (2010), “Small area estimation of poverty indicators,” *Canadian Journal of Statistics*, 38, 3, 369-385.

Molina, I., Nandram, B., and 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*, 8, 852-885.

Nandram, B. and Sedransk, J and Pickle, L. W. (2000), “Bayesian analysis and mapping of mortality rates for chronic obstructive pulmonary disease,” *Journal of the American Statistical Association*, 95, 452, 1110-1118, Taylor & Francis Group.

Nandram, B. and Choi, J. W. (2005), “Hierarchical Bayesian nonignorable nonresponse regression models for small areas: An application to the NHANES data,” *Survey Methodology*, 32, 1, 73-84, Statistics Canada.

Nandram, B. and Choi, J. W. (2010), “A Bayesian analysis of body mass index data from small domains under nonignorable nonresponse and selection,” *Journal of the American Statistical Association*, 105, 120-135.

Nandram, B., Fu, S., and Manandhar, B. (2017), “Bayesian logistic regression for small areas with numerous households,” *(under preparation)*.

Nandram, B., Xu, Z., and Manandhar, B. (2017), “Bayesian inference of a finite population under length bias,” *(under preparation)*.

Nguyen, P., Haughton, D., Hudson, I., and Boland, J. (2010), “Multilevel models ans small area estimation in the context of Vietnam living standards surveys,” 42emes *Journes de Statistique*.

Pettit, L. I. (1990), “The conditional predictive ordinate for the normal distribution,” *Journal of the Royal Statistical Society*, 175-184.

Rubin, D. B. (1981), “The Bayesian bootstrap,” The annals of statistics, 9(1), 130-134.

The World Bank (2006), “Nepal - Resilience amidst conflict: an assessment of poverty in Nepal, 1995-96 and 2003-04,” *The World Bank*.

The World Bank (2013), “Nepal small area estimation of poverty 2011,” *The World Bank*, vol. 1.