Bayesian Predictive Inference for Consumption Data from Small Areas

Binod Manandhar* Balgobin Nandram†

Abstract

We developed a robust Bayesian method, based on the generalized beta distribution of the second kind (GB2) to analyze consumption data from Nepal. Our objective is to predict the poverty rates of small areas. The consumption data are positively-skewed and this suggests transforming the data using a logarithmic transformation, which however could be problematic. We use a standard small area model with two covariates and we assume that the consumption data have a flexible distribution that can be conveniently expressed as the scale mixture of generalized gamma distributions with another generalized gamma distribution being the mixing distribution. We have constructed a hierarchical Bayesian model and we have incorporated the covariates in an innovative manner. We have applied this model to the second Nepal Living Standards Survey (NLSS-II). We have compared our model with the hierarchical Bayesian nested error regression (NER) model which uses normality assumption. Under the GB2 density the joint posterior density is complex, so we have used Markov chain Monte Carlo (MCMC) methods to fit it. The NER model does not need MCMC methods but, as indicated, it could be problematic under the logarithmic transformation.

Key Words: GB2 distribution, Hierarchical Bayes, Logarithmic transformation, Noninformative priors, Non-normality, Poverty, Small area estimation.

1. Introduction

Positively skewed continuous data are seen in many situations. Size data such as income, consumption, insurance and loss data are examples of continuous positively skewed data. In many situations, such data could be heavy-tailed as well. There are numerous statistical models and tests which have been developed under the normality assumption of a variable under study. If the variable under study is not normal or approximately normal, then the usual way to meet the normality assumption is by transformation. If we have positively-skewed data, log transformation is the widely-used tool to meet the normality assumption and proceed for model-building. Once the normality assumption holds, it makes model-building, further analysis, and computation easier. The usual practice, frequentist as well as in Bayesian paradigm, is to get estimates by the log transformation of a response variable. For example, the small area estimation (SAE) of poverty applied by the World Bank in numerous countries uses a nested error model with a logarithmic-transformed response variable. The hierarchical Bayesian small area estimation paper (Molina, Nandram, and Rao, 2014) used the logarithm transformation of a response for estimating SAE of poverty. Nandram and Choi (2010) also used the logarithmic transformation for a proposed nonignorable nonresponse model, a spline regression, to estimate the finite population means of small domains formed by crossing age, race and sex within counties.

Log transformation is the most popular tool used to meet normality assumption when a response variable has a positively-skewed distribution. Once we use transformation to build a model, then the usual way to get back to the original scale estimates is to perform back transformation. Does back transformation give the correct distribution of the response variable? Furthermore, what if the normality assumption fails? There are numerous positively-skewed distributions, one of them is a log-normal distribution. If the original data follow

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

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

a log-normal or approximately log-normal distribution, then the log-transformed data follow a normal or approximately normal distribution. However, back transformation does not give a correct distribution (Feng et al. 2014). It would be better if we had a model that could give better estimates and prediction, without logarithmic transformation of the variable under study. Out concept is to develop a statistical model without the need for a logarithmic transformation for positively-skewed data. This will definitely avoid the need for transformation since we are not obligated to meet normality assumption of the response variable.

There are many distributions which are positively-skewed. Which distribution should we choose? Exponential, gamma, and generalized gamma distributions are some standard examples of continuous positively-skewed distributions. Since we do not know the distribution of the data, we may need to assume that the data follow some particular distribution for model-building. If we do not know much about the data, it could be better to choose a generalized distribution.

Which generalized distribution would be better to choose for positively-skewed data? $GB2$ is a generalized distribution which can give many distributions as special cases. A three-parameter distribution includes log-t, generalized gamma, and beta of type two; two-parameter distribution includes log-Cauchy, log-normal, Weibull, gamma, and shifted Pareto; one-parameter distribution includes half normal, exponential, and Chisquare (Dong and Chan, 2013).

$GB2$ could be a useful distribution for continuous positively-skewed data with a heavy tail. The shape and rate parameters included in GB2 distribution contribute in capturing those heavy tails and can be used in modeling positively-skewed data. The $GB2$ distribution is acknowledged to give an excellent description of income distribution (McDonald, 1984; McDonald and Xu, 1995). Graf and Nedyalkova (2014) used $GB2$ distribution to model income and as an indicators of poverty data. They fit the $GB2$ model to the income distribution and to study income inequality at the country level in the contest of the European Union Statistics on Income and Living Conditions (EU-SILC). Dong and Chan (2013) presented a Bayesian approach in order to model long-tail loss-reserving data using $GB2$.

A GB2 distribution could be a useful distribution, if the data have errors and bias. The errors in the data could arise as a measurement error, a response error, a recalling error or a bias along with many other possibilities. In a living standards survey the welfare variable consumption is the aggregate of all food and all non-food items consumed. The respondent has to recall all kind of consumptions (in monitory value) throughout the whole reference year. In such cases the recorded data may not be the true response and GB2 could be helpful for modeling, since GB2 assumes that its rate parameter has a distribution.

This paper focuses on giving estimates for small areas. Small-area estimation refers to a collection of statistical techniques designed for improving sample survey estimates through the use of auxiliary information (Rao and Molina, 2015). It begins with a response variable for which we require estimates over a range of small areas. To facilitate SAE we introduced covariates in our model with the random area effect in hierarchical Bayesian paradigm. In this paper, we will demonstrate use of the GB2 model for continuous positively-skewed data using the NLSS-II, mountains stratum. The response variable is per capita consumption and two covariates were used. We will show the results using the sampled data and a prediction within the sampled units. An extension to a non-sampled units prediction can also be done, which we will discuss in Section 5. The goal of this paper is to avoid log transformation of a response variable, this is accommodated using the GB2 for a positively-skewed data, and introduce covariates in the model.

We organize the paper as follows. In Section 2, we discuss problems that arise with logarithmic transformation for continuous, positively-skewed data. In Section 3, we present GB2 distribution as a mixture of two generalized gamma distributions. In Section 4, we develop a Bayesian hierarchical model and introduce covariates into the model. In Section 5, we show how to predict a response variable for sampled as well as non-sampled observations from GB2 model. In Section 6, we talk about the computation and drawing samples. In Section 7, we show the results from NLSS-II mountains stratum, consumption as the response variable with two covariates, followed by a discussion in Section 8.

2. Problems with log transformation

Let us assume we have original response observations from the log-normal distribution, $y' \sim LN(\mu, \sigma^2)$. The mean of the original response variable is

$$E[Y'] = e^{\mu + \frac{\sigma^2}{2}}.$$

Let us take a log transformation, $y = \ln(y')$. Now, the model with log-transformed response variable follows a normal distribution with mean μ_{ln}. The mean estimate of the log-transformed variable is

$$\hat{\mu}_{ln} = \frac{1}{n}\sum_{i=1}^{n} \log(y_i').$$

Transferring back to the mean we get $e^{\hat{\mu}_{ln}}$, a maximum likelihood estimator but not an unbiased estimate of e^{μ}. However, the mean of the original response data y' is $e^{\mu+\frac{\sigma^2}{2}}$, not e^{μ}. Thus, log transformation can not give correct estimates of the log-normal distribution (See Feng et al., 2013, 2014).

There is yet another problem pertinent with the log transformation. We will show an example of the non-existence of moments in the Bayesian paradigm. Let us consider a population with size N, with response vector $\boldsymbol{y}$, where we have sampled $\boldsymbol{y_s} = y_1, \cdots, y_n$. Let us take the log transformation of the observed variable

$$z_i = \log(y_i), \quad i = 1, \cdots, N.$$

Consider a model and its prior

$$\begin{aligned} z_1, \cdots, z_n, z_{n+1}, \cdots, z_N | \mu, \sigma^2 &\overset{\text{iid}}{\sim} N(\mu, \sigma^2), \quad \sigma^2 > 0 \\ \pi(\mu, \sigma^2) &\propto \frac{1}{\sigma^2}. \end{aligned}$$

It gives

$$\begin{aligned} \frac{(n-1)s^2}{\sigma^2} | \boldsymbol{z_s} &\sim \Gamma\left(\frac{n-1}{2}, \frac{1}{2}\right), \quad \boldsymbol{z_s} = (z_1, \cdots, z_n)', \\ \frac{\mu - \bar{z}}{s/\sqrt{n}} | \boldsymbol{z}_s &\sim t_{n-1}. \end{aligned}$$

Let us find the expected value of response variable by integrating out parameters.

$$\begin{aligned} I &= E[y_i | \boldsymbol{z_s}] = E_{\mu,\sigma^2}\left[E(y_i | \mu, \sigma^2, \boldsymbol{z_s})\right] \\ &= E_{\mu,\sigma^2}[e^{\mu + \frac{\sigma^2}{2}} | \boldsymbol{z_s}] \\ &= \int_{-\infty}^{\infty} \int_0^{\infty} e^{\mu + \frac{\sigma^2}{2}} \quad \pi(\mu | \sigma^2, \boldsymbol{z_s}) \, \pi(\sigma^2 | \boldsymbol{z_s}) \, d\sigma^2 d\mu. \end{aligned}$$

Since $\frac{\sigma^2}{2} > 0$,

$$\begin{aligned} I &\geq \int_{-\infty}^{\infty}\int_{0}^{\infty} e^{\mu} \quad \pi(\mu|\sigma^2, \boldsymbol{z_s})\, \pi(\sigma^2|\boldsymbol{z_s})\; d\sigma^2 d\mu \\ &= \int_{-\infty}^{\infty} e^{\mu} \int_{0}^{\infty} \quad \pi(\mu|\sigma^2, \boldsymbol{z_s})\, \pi(\sigma^2|\boldsymbol{z_s})\; d\sigma^2 d\mu \\ &= \int_{-\infty}^{\infty} e^{\mu} \quad t_{n-1}\left(\frac{\mu - \bar{z}}{s/\sqrt{n}}|\boldsymbol{z_s}\right) d\mu \\ &= \infty. \end{aligned}$$

Thus the expected value of y_i does not exist. This is indeed problematic for inference.

3. The generalized beta distribution of the second kind (GB2)

GB2 distribution has four parameters, which can be expressed as a mixture of the generalized gamma distributions. The probability density function of a response variable y and the probability density function of the rate parameter λ both have the generalized gamma distribution:

$$f(y|\lambda, \alpha, \gamma) = \frac{\gamma e^{-(\lambda y)^{\gamma}} y^{\alpha-1} \lambda^{\alpha}}{\Gamma(\frac{\alpha}{\gamma})}, \quad \lambda, \alpha, \gamma > 0, \tag{1}$$

$$g(\lambda|\eta, \phi, \gamma) = \frac{\gamma e^{-(\eta\lambda)^{\gamma}} \lambda^{\phi-1} \eta^{\phi}}{\Gamma(\frac{\phi}{\gamma})}, \quad \eta, \phi, \gamma > 0. \tag{2}$$

Mixing the generalized gamma density of y with the generalized gamma density of the rate parameter λ gives the $GB2$ distribution with four parameters, $Y \sim GB2(\alpha, \phi, \gamma, \eta)$. In the GB2, the shape parameters α and ϕ determine the skewness of the distribution, the shape parameter γ controls the overall shape, and η is the rate parameter

$$f(y|\alpha, \phi, \gamma, \eta) = \int_{0}^{\infty} f(y|\lambda, \alpha, \gamma) g(\lambda|\eta, \phi, \gamma)\; d\lambda \tag{3}$$

$$= \gamma \eta^{\phi} \frac{\Gamma(\frac{\alpha+\phi}{\gamma})}{\Gamma(\frac{\alpha}{\gamma})\Gamma(\frac{\phi}{\gamma})} \frac{y^{\alpha-1}}{(y^{\gamma} + \eta^{\gamma})^{\frac{\alpha+\phi}{\gamma}}}, \quad \alpha, \phi, \gamma, \eta > 0. \tag{4}$$

The k^{th} moment of the $GB2$ distribution exists if $\phi > k$, and it is given by

$$E[Y^k|\alpha, \phi, \gamma, \eta] = \frac{\Gamma\left(\frac{\alpha+1}{\gamma}\right)}{\Gamma\left(\frac{\alpha}{\gamma}\right)} \frac{\Gamma\left(\frac{\phi-k}{\gamma}\right)}{\Gamma\left(\frac{\phi}{\gamma}\right)} \eta^{k}, \quad \phi > k.$$

For the mean to exist, we need $\phi > 1$. If the mean exists, we have

$$\eta = \frac{\Gamma\left(\frac{\alpha}{\gamma}\right)}{\Gamma\left(\frac{\alpha+1}{\gamma}\right)} \frac{\Gamma\left(\frac{\phi}{\gamma}\right)}{\Gamma\left(\frac{\phi-1}{\gamma}\right)} E[Y|\alpha, \phi, \gamma, \eta], \quad \phi > 1.$$

Let $E[Y|\alpha, \phi, \gamma, \eta] = e^{\mu}$. Then $\mu = \ln(E[Y|\alpha, \phi, \gamma, \eta])$, and

$$f(y|\alpha, \phi, \gamma, \mu) = \gamma \left[\frac{\Gamma\left(\frac{\alpha+1}{\gamma}\right)}{\Gamma\left(\frac{\alpha}{\gamma}\right)} \frac{\Gamma\left(\frac{\phi-1}{\gamma}\right)}{\Gamma\left(\frac{\phi}{\gamma}\right)} e^{-\mu} \right]^{\alpha} \frac{\Gamma(\frac{\alpha+\phi}{\gamma})}{\Gamma(\frac{\alpha}{\gamma})\Gamma(\frac{\phi}{\gamma})} \frac{y^{\alpha-1}}{\left(1 + \left[y \frac{\Gamma\left(\frac{\alpha+1}{\gamma}\right)}{\Gamma\left(\frac{\alpha}{\gamma}\right)} \frac{\Gamma\left(\frac{\phi-1}{\gamma}\right)}{\Gamma\left(\frac{\phi}{\gamma}\right)} e^{-\mu} \right]^{\gamma} \right)^{\frac{\alpha+\phi}{\gamma}}},$$

with $\alpha, \gamma > 0,\ \phi > 1$.

We adjust the density function such that the mean exists with $\phi > 0$, and for $y \geq 0$

$$f(y|\alpha,\phi,\gamma,\mu) = \gamma\left[\frac{\Gamma\left(\frac{\alpha+1}{\gamma}\right)}{\Gamma\left(\frac{\alpha}{\gamma}\right)}\frac{\Gamma\left(\frac{\phi}{\gamma}\right)}{\Gamma\left(\frac{\phi+1}{\gamma}\right)}e^{-\mu}\right]^{\alpha}\frac{\Gamma(\frac{\alpha+\phi+1}{\gamma})}{\Gamma(\frac{\alpha}{\gamma})\Gamma(\frac{\phi+1}{\gamma})}\frac{y^{\alpha-1}}{\left(1+\left[y\frac{\Gamma\left(\frac{\alpha+1}{\gamma}\right)}{\Gamma\left(\frac{\alpha}{\gamma}\right)}\frac{\Gamma\left(\frac{\phi}{\gamma}\right)}{\Gamma\left(\frac{\phi+1}{\gamma}\right)}e^{-\mu}\right]^{\gamma}\right)^{\frac{\alpha+\phi+1}{\gamma}}},$$

with $\alpha, \phi, \gamma > 0$.

4. The GB2 hierarchical Bayes model

Let the sample data have n observations, and the response variable y have p covariates. There are ℓ small areas, $i = 1, \cdots, \ell$, and each small area has $j = 1, \cdots, n_i$, observations. Let us consider $\mu_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \nu_i$, where $\boldsymbol{x}$ is the $p \times 1$ vector of covariates, $\boldsymbol{\beta}$ is the weight parameter for each covariate and ν_i is the area effect. The likelihood function is

$$f(\boldsymbol{y}|\alpha,\phi,\gamma,\boldsymbol{\beta},\boldsymbol{\nu}) = \prod_{i=1}^{\ell}\left[\frac{\gamma\left[\frac{\Gamma\left(\frac{\alpha+1}{\gamma}\right)}{\Gamma\left(\frac{\alpha}{\gamma}\right)}\frac{\Gamma\left(\frac{\phi}{\gamma}\right)}{\Gamma\left(\frac{\phi+1}{\gamma}\right)}\right]^{\alpha}(g_i)^{\alpha-1}}{\frac{\Gamma(\frac{\alpha}{\gamma})\Gamma(\frac{\phi+1}{\gamma})}{\Gamma(\frac{\alpha+\phi+1}{\gamma})}}\right]^{n_i} \times \prod_{i=1}^{\ell}\left[\frac{e^{-\alpha\left[\sum_{j=1}^{n_i}(\boldsymbol{x}'_{ij}\boldsymbol{\beta})+n_i\nu_i\right]}}{\prod_{j=1}^{n_i}\left(1+\left[y_{ij}\frac{\Gamma\left(\frac{\alpha+1}{\gamma}\right)}{\Gamma\left(\frac{\alpha}{\gamma}\right)}\frac{\Gamma\left(\frac{\phi}{\gamma}\right)}{\Gamma\left(\frac{\phi+1}{\gamma}\right)}e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}\right)^{\frac{\alpha+\phi+1}{\gamma}}}\right].$$

Let

$$R_{(\alpha,\phi,\gamma)} = \left[\frac{\Gamma\left(\frac{\alpha+1}{\gamma}\right)}{\Gamma\left(\frac{\alpha}{\gamma}\right)}\frac{\Gamma\left(\frac{\phi}{\gamma}\right)}{\Gamma\left(\frac{\phi+1}{\gamma}\right)}\right],\ B_{(\alpha,\phi,\gamma)} = \frac{\Gamma(\frac{\alpha}{\gamma})\Gamma(\frac{\phi+1}{\gamma})}{\Gamma(\frac{\alpha+\phi+1}{\gamma})},$$

$$g_i = \left[\prod_{j=1}^{n_i} y_{ij}\right]^{1/n_i}, \quad i = 1, \cdots, \ell.$$

We have

$$f(\boldsymbol{y}|\alpha,\phi,\gamma,\boldsymbol{\beta},\boldsymbol{\nu}) = \prod_{i=1}^{\ell}\left[\frac{\gamma\left[R_{(\alpha,\phi,\gamma)}\right]^{\alpha}(g_i)^{\alpha-1}}{B_{(\alpha,\phi,\gamma)}}\right]^{n_i} \times \prod_{i=1}^{\ell}\left[\frac{e^{-\alpha\left(\sum_{j=1}^{n_i}(\boldsymbol{x}'_{ij}\boldsymbol{\beta})+n_i\nu_i\right)}}{\prod_{j=1}^{n_i}\left(1+\left[y_{ij}R_{(\alpha,\phi,\gamma)}e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}\right)^{\frac{\alpha+\phi+1}{\gamma}}}\right]. \tag{5}$$

We assume the area effects follow a normal distribution

$$\nu_i|\sigma^2 \sim N(0,\sigma^2), \quad i=1,\cdots,\ell.$$

Using non-informative independent priors for $\sigma^2, \alpha, \phi, \gamma$, and β given by

$$\begin{aligned}\pi(\sigma^2) &= \frac{1}{(1+\sigma^2)^2}, \quad \pi(\alpha)=\frac{1}{(1+\alpha)^2}, \quad \pi(\phi)=\frac{1}{(1+\phi)^2}, \quad \pi(\gamma)=\frac{1}{(1+\gamma)^2},\\ \pi(\boldsymbol{\beta}) &\propto 1,\end{aligned}$$

the joint posterior distribution is

$$\begin{aligned}f(\boldsymbol{\beta},\sigma^2,\alpha,\phi,\gamma,\boldsymbol{\nu}|\boldsymbol{y}) \propto&\\ &\frac{1}{(1+\sigma^2)^2}\times\frac{1}{(1+\alpha)^2}\times\frac{1}{(1+\phi)^2}\times\frac{1}{(1+\gamma)^2}\\ &\times\prod_{i=1}^{\ell}\left[\frac{\gamma\left[R_{(\alpha,\phi,\gamma)}\right]^{\alpha}(g_i)^{\alpha-1}}{B_{(\alpha,\phi,\gamma)}}\right]^{n_i}\\ &\times\prod_{i=1}^{\ell}\left[\frac{e^{-\alpha\left(\sum_{j=1}^{n_i}(\boldsymbol{x}'_{ij}\boldsymbol{\beta})+n_i\nu_i\right)}}{\prod_{j=1}^{n_i}\left(1+\left[y_{ij}R_{(\alpha,\phi,\gamma)}e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}\right)^{\frac{\alpha+\phi+1}{\gamma}}}\right]\\ &\times\prod_{i=1}^{\ell}\left[\frac{1}{\sqrt{\sigma^2}}e^{-\frac{1}{2}\frac{\nu_i^2}{\sigma^2}}\right].\end{aligned} \tag{6}$$

5. Prediction

If we have the set of parameters $\alpha, \phi, \gamma, \boldsymbol{\beta}$ and $\nu_i, i=1,\cdots,\ell$, we can find μ_{ij} for each sampled PSU as

$$\mu_{ij} = \boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i, \quad i=1,\cdots,\ell, \quad j=1,\cdots,n_i.$$

For the non-sampled PSUs, we do not have information for $\nu_i, i=\ell+1,\cdots,L$. However, we can use the prior distribution, $\nu_i \sim N(0,\sigma^2)$, to draw area effects with information of parameter σ^2. (Note, this paper has shown results only for sampled observations.) Now, we can calculate parameter η as

$$\eta_{ij} = \frac{\Gamma\left(\frac{\alpha}{\gamma}\right)\Gamma\left(\frac{\phi+1}{\gamma}\right)}{\Gamma\left(\frac{\alpha+1}{\gamma}\right)\Gamma\left(\frac{\phi}{\gamma}\right)}e^{\mu_{ij}}.$$

From the generalized gamma density of response $y|\lambda,\alpha,\gamma$ in (1) and parameter $\lambda|\eta,\alpha,\gamma$ in (2), we can write these densities with transformation, $t=(\lambda y)^\gamma$ and $t=(\eta\lambda)^\gamma$ respectively as

$$\begin{aligned}G1 &= (\eta\lambda)^\gamma \sim \Gamma\left(\frac{\phi+1}{\gamma},1\right), \text{ and}\\ G2 &= (\lambda y)^\gamma \sim \Gamma\left(\frac{\alpha}{\gamma},1\right).\end{aligned}$$

If two gamma random samples G_1 and G_2 are drawn from above two distributions respectively, then we can predict the response, $\hat{y}_{ij}$ as follows

$$
\begin{aligned}
\lambda_{ij} &= \frac{(G_1)^{\frac{1}{\gamma}}}{\eta_{ij}}, \text{ and} \\
\hat{y}_{ij} &= \frac{(G_2)^{\frac{1}{\gamma}}}{\lambda_{ij}}.
\end{aligned}
$$

The $\hat{y}_{ij}$ are smoothed estimates of the y_{ij}, the observed response values.

6. Computation

There could be numerous small areas and it may not be feasible to sample area effects within MCMC. So, we integrate out area parameters $\nu_i, i = 1, \cdots, \ell$, from the joint posterior density function which will reduce the large number of parameters in the MCMC sequence, and help to provide better mixing in MCMC. However, posterior density is not in a simple form, so we integrate out ν_i numerically. We transform ν_i to standard normal $z_i = \nu_i/\sigma$, for numerical integration, and divide the domain into m equal intervals $[t_{k-1},\ t_k], k = 1, \cdots, m$. We can approximate joint posterior density after integrating out $\nu_i, i = 1, \cdots, \ell$ as follows

$$
\begin{aligned}
&f(\boldsymbol{\beta}, \sigma^2, \alpha, \phi, \gamma|\boldsymbol{y}) \\
&\propto \frac{1}{(1+\sigma^2)^2} \times \frac{1}{(1+\alpha)^2} \times \frac{1}{(1+\phi)^2} \times \frac{1}{(1+\gamma)^2} \times \prod_{i=1}^{\ell} \left\{ \left[\frac{\gamma \left[R_{(\alpha,\phi,\gamma)}\right]^{\alpha} (g_i)^{\alpha-1}}{B_{(\alpha,\phi,\gamma)}} \right]^{n_i} \right\} \\
&\times \prod_{i=1}^{\ell} \left\{ \sum_{k=1}^{m} \int_{t_{k-1}}^{t_k} \left[\frac{e^{-\alpha\left(\sum_{j=1}^{n_i} (\boldsymbol{x}'_{ij}\boldsymbol{\beta}) + n_i z \sigma\right)}}{\prod_{j=1}^{n_i} \left(1 + \left[y_{ij} R_{(\alpha,\phi,\gamma)} e^{-(x'\boldsymbol{\beta}+z\sigma)}\right]^{\gamma}\right)^{\frac{\alpha+\phi+1}{\gamma}}} \right] \frac{1}{\sqrt{\sigma^2}} e^{-\frac{1}{2}z^2} \ dz \right\}.
\end{aligned}
$$

For numerical integration we take 50 grid points between $(3, -3)$. The standard normal density domain covers 99.74% of the distribution within this range. The grid points are the middle points of the intervals, $z_k = \frac{t_{k-1}+t_k}{2},\ k = 1, \cdots, m$. Using large number grid values for ν_i: 100, 200, 1000 does not make much difference with the number of grids 50.

$$
\begin{aligned}
&f(\boldsymbol{\beta}, \sigma^2, \alpha, \phi, \gamma|\boldsymbol{y}) \\
&\propto \frac{1}{(1+\sigma^2)^2} \times \frac{1}{(1+\alpha)^2} \times \frac{1}{(1+\phi)^2} \times \frac{1}{(1+\gamma)^2} \times \prod_{i=1}^{\ell} \left\{ \left[\frac{\gamma \left[R_{(\alpha,\phi,\gamma)}\right]^{\alpha} (g_i)^{\alpha-1}}{B_{(\alpha,\phi,\gamma)}} \right]^{n_i} \right\} \\
&\times \prod_{i=1}^{\ell} \left\{ \sum_{k=1}^{m} \left[\frac{e^{-\alpha\left(\sum_{j=1}^{n_i} (\boldsymbol{x}'_{ij}\boldsymbol{\beta}) + n_i z_k \sigma\right)}}{\prod_{j=1}^{n_i} \left(1 + \left[y_{ij} R_{(\alpha,\phi,\gamma)} e^{-(x'\boldsymbol{\beta}+z_k\sigma)}\right]^{\gamma}\right)^{\frac{\alpha+\phi+1}{\gamma}}} \right] (\Phi(t_k) - \Phi(t_{k-1})) \right\}.
\end{aligned} \tag{7}
$$

We have posterior density in complex form and none of the conditional posterior densities are in simple form either. We have used the MCMC method for sampling parameters. We used grid sampling and the Metropolis Hastings method to draw parameters. The grid sampling method is used to draw σ^2, α, ϕ and γ parameters and the Metropolis Hastings algorithm is used to draw parameters $\boldsymbol{\beta}, \nu_i, i = 1, \cdots, \ell$. The number of grids used to draw parameters σ^2, α, ϕ and γ is 100. Using more grids like 200, 500, 1000 does not make a difference compared to the number of grids 100.

The proposal density for $\boldsymbol{\beta}$ is a multivariate normal, $\boldsymbol{\beta}| \sum \sim MN(\hat{\boldsymbol{\beta}},\ \sum)$, and a proposal density $\nu_i, i = 1, \cdots, \ell$, is a univariate normal, $\nu_i \overset{\text{iid}}{\sim} N(\hat{\nu_i}, \delta^2)$. We obtain $\hat{\boldsymbol{\beta}}$ and Σ as follows. Given other parameters, the proposal mean vector $\hat{\boldsymbol{\beta}}$ is obtained numerically from the Nelder-Mead simplex method. For a covariance matrix, the variances and covariances are calculated numerically and tuned by some positive constant (multiplied by two in our case). Similarly, for the parameter ν_i, the proposal univariate normal distribution's mean and variance ($\hat{\nu_i}$, δ^2) are obtained using numerical integration given other parameters. The target densities are respectively

$$f(\boldsymbol{\beta}|\sigma^2, \alpha, \phi, \gamma, \boldsymbol{y}) \propto \prod_{i=1}^{l} \int_{\nu_i} \left[\frac{e^{-(x'\boldsymbol{\beta}+\nu_i)} e^{-\frac{1}{2}\frac{\nu_i^2}{\sigma^2}} \qquad d\nu_i}{\prod_{j=1}^{n_i} \left(1 + \left[y_{ij} R_{(\alpha,\phi,\gamma)} e^{-(x'\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}\right)^{\frac{\alpha+\phi+1}{\gamma}}} \right],$$

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

The jumping probability from step $k-1$ to k for ν_i is shown below (the jumping probability for $\boldsymbol{\beta}$ can be written similarly)

$$\rho(k-1, k) = \min\left(1, \quad \frac{\pi(\nu_i^{(k)}) \; q(\nu_i^{(k-1)} \mid \nu_i^{(k)})}{\pi(\nu_i^{(k-1)}) \; q(\nu_i^{(k)} \mid \nu_i^{(k-1)})}\right),$$

where $\pi(.)$ is the target density function and $q(.)$ is the proposal density function. We choose a new k^{th} sample with probability $U \sim Unif(0, 1)$. Let us say, we draw a random number u from the standard uniform distribution, then accept the new sample $\nu_i^{(k)}$ if $u < \rho(k-1, k)$, else keep the previous sample $\nu_i^{(k-1)}$.

We used Metropolis Hastings sampler for parameter $\boldsymbol{\beta}$ within the Gibbs sampler. The acceptance rate for the Metropolis Hastings sampling of $\boldsymbol{\beta}$ is about 20%. After drawing all other rate and shape parameters we draw area effect parameter ν_i for each area separately. We draw parameter ν_i using the Metropolis Hastings sampler that has an acceptance rate ranging from 40% to 75%.

We have used NLSS-II consumption data as our response variable with two covariates: household size and share of kids aged between zero to seven. We show results using sampled data in NLSS-II survey for the mountains stratum. We drew 5200 parameter sets burn-in 270 samples with final 400 samples left. For consistency, we have also checked the results of total Gibbs samples remaining 1000, 500, and 100 Gibbs and those estimates are consistent with 400 Gibbs remaining. We went through diagnostic procedure of Geweke test, auto-correlation test and trace plots to get the final samples.

7. Results

Figure 1 is a density plot of the observed response and estimated responses for each enumeration unit in the sampled data. Dark bold black line is the observed response from the mountains stratum, NLSS-II and red thin lines are GB2 estimates of the responses from the model. This plot shows that the GB2 model can give a better density estimate of the observed density.

Figure 2 shows the a posterior poverty rates of the mountains stratum of Nepal using the

GB2 model and hierarchical Bayes nested error regression model (NER). The GB2 model estimates overall poverty rate of 0.301 with standard error of 0.034 and an hierarchical Bayes NER model estimate of 0.28 with a standard error of 0.025. The estimate by direct method in the NLSS-II report and the World bank report (2006) is 0.326.

Poverty rates are calculated by comparing per capita consumption against the national poverty line for Nepal of $7,696$ rupees per year in average (2003) as given in the report, SAE of Poverty, Caloric Intake and Malnutrition in Nepal (2006). It represents the percentage of the population below the national poverty line. Table 1 shows the district level poverty estimates of 14 districts of the mountains stratum using sampled observations only. We note that NLSS-II is not designed to give estimates in district levels because of the small sample sizes. Therefore, direct estimates for district level from NLSS-II do not make sense. If we do direct poverty estimates, the Manang and Dolpa districts have zero percentage poverty rates and the Humla district has a 91.3% poverty rate. Fitting a model will help to smooth out prediction by borrowing strength. By fitting the GB2 and hierarchical Bayes NER model these extreme poverty estimates have been smoothed out. Table 1 shows the results obtained from the GB2 model, hierarchical Bayes model as described in Molina, Nandram and Rao (2014) and direct estimates from NLSS-II survey. For both GB2 and hierarchical Bayes NER models, some of the districts such as Solukhumbu (22.8% from GB2 vs 22.7% from NER model) and Bajura (36.1% from GB2 vs 36.9% from NER model) gave close poverty estimates of poverty rates but most of districts have different poverty estimates. Here, we have shown results using only sampled households in the survey with two covariates, to demonstrate GB2 model for positively-skewed continuous data.

Table 1: District poverty rates and standard errors by NLSS-II direct method, GB2 and Bayesian nested error regression (NER) model for the mountains stratum.

Districts	Sample size	NLSS direct		GB2 model		NER model	
		Pov. Rate	SE	Pov. Rate	SE	Pov. Rate	SE
Taplejung	36	0.288	0.075	0.260	0.083	0.199	0.072
Sankhuwasabha	48	0.290	0.066	0.290	0.072	0.252	0.069
Solukhumbu	24	0.166	0.076	0.228	0.094	0.227	0.098
Dolakha	48	0.192	0.057	0.244	0.068	0.218	0.064
Sindhupalchok	84	0.433	0.054	0.342	0.062	0.318	0.051
Manang	12	0	0	0.046	0.069	0.003	0.017
Dolpa	12	0	0	0.133	0.112	0.066	0.078
Jumla	12	0.258	0.126	0.206	0.124	0.256	0.149
Kalikot	12	0.593	0.142	0.402	0.153	0.532	0.166
Mugu	12	0.300	0.132	0.321	0.136	0.211	0.136
Humla	12	0.913	0.082	0.460	0.175	0.666	0.158
Bajura	24	0.288	0.092	0.361	0.104	0.369	0.105
Bajhang	24	0.116	0.065	0.319	0.107	0.227	0.088
Darchula	24	0.536	0.102	0.369	0.110	0.418	0.111
Mountains	384	0.326	0.024	0.301	0.034	0.280	0.025

8. Discussion

Logarithmic transformation for positively-skewed data could be problematic. We fit the GB2 distribution to continuous positively-skewed consumption data within the Bayesian paradigm using the information of covariates without logarithmic transformation. This paper fits hierarchical the Bayesian GB2 model using NLSS-II sampled data and predicts within the same sampled data. In the future we will extend this GB2 model to give estimates for non-sampled observation as we have discussed in Section 5. Here we have shown results only using two covariates, our next extension is to use dimension reduction when we have numerous covariates.

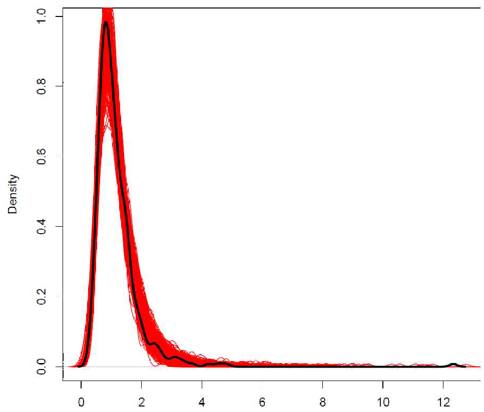

Figure 1: Density plot: Y vs $\hat{Y}$

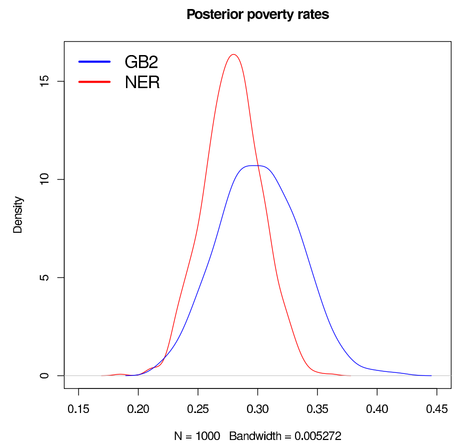

Figure 2: Posterior poverty rates

REFERENCES

Central Bureau of Statistics (CBS), Nepal (2004), "Nepal Living Standards Survey Report," *CBS Nepal*.

Central Bureau of Statistics (CBS), Nepal (2005), "Poverty trends in Nepal (1995-96 and 2003-04)," *CBS Nepal*.

Dong, A. and Chan, J. S. K. (2013), "Bayesian analysis of loss reserving using dynamic models with generalized beta distribution," *Insurance: Mathematics and Economics* , 53, 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.

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.

McDonald, J. B. and Xu, Y. J. (1995), "A generalization of the beta distribution with applications," *Journal of Econometrics*, 66, 133-152.

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 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.

Rao, J. N. K. and Molina, I. (2015), "Small Area Estimation, 2nd Edition," *Wiley*, ISBN: 978-1-118-73578-7.

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