

## Research Article

# Anomaly Detection in Health Insurance Claims Using Bayesian Quantile Regression

**Ezekiel N. N. Nortey, Reuben Pometsey, Louis Asiedu , Samuel Iddi ,  
and Felix O. Mettle **

*Department of Statistics & Actuarial Science, School of Physical and Mathematical Sciences, University of Ghana, Legon, Accra, Ghana*

Correspondence should be addressed to Louis Asiedu; lasiedu@ug.edu.gh

Received 7 October 2020; Revised 8 January 2021; Accepted 12 February 2021; Published 24 February 2021

Academic Editor: Sergejs Solovjovs

Copyright © 2021 Ezekiel N. N. Nortey et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Research has shown that current health expenditure in most countries, especially in sub-Saharan Africa, is inadequate and unsustainable. Yet, fraud, abuse, and waste in health insurance claims by service providers and subscribers threaten the delivery of quality healthcare. It is therefore imperative to analyze health insurance claim data to identify potentially suspicious claims. Typically, anomaly detection can be posited as a classification problem that requires the use of statistical methods such as mixture models and machine learning approaches to classify data points as either normal or anomalous. Additionally, health insurance claim data are mostly associated with problems of sparsity, heteroscedasticity, multicollinearity, and the presence of missing values. The analyses of such data are best addressed by adopting more robust statistical techniques. In this paper, we utilized the Bayesian quantile regression model to establish the relations between claim outcome of interest and subject-level features and further classify claims as either normal or anomalous. An estimated model component is assumed to inherently capture the behaviors of the response variable. A Bayesian mixture model, assuming a normal mixture of two components, is used to label claims as either normal or anomalous. The model was applied to health insurance data captured on 115 people suffering from various cardiovascular diseases across different states in the USA. Results show that 25 out of 115 claims (21.7%) were potentially suspicious. The overall accuracy of the fitted model was assessed to be 92%. Through the methodological approach and empirical application, we demonstrated that the Bayesian quantile regression is a viable model for anomaly detection.

## 1. Introduction

Providing quality and accessible healthcare is a major political decision that governments make all over the world. A major challenge in achieving this is funding. According to Novignon et al. [1], majority of countries in developing regions, especially sub-Saharan Africa (SSA), rely on donor grants and loans to finance healthcare. Such expenditures are not only unsustainable but also inadequate considering the enormous healthcare burden in the region. Apart from direct government investment into health services, health insurance schemes have been instituted to reduce the proportion of health service cost borne by the populace in many countries. However, fraud, abuse, and waste in health

insurance claims by service providers and subscribers threaten the delivery of quality healthcare. The National Healthcare Antifraud Association (NHCAA) of the United States defines healthcare fraud as an intentional deception or misrepresentation made by a person or an entity, with the knowledge that the deception could result in some kinds of unauthorized benefits to that person or entity [2, 3]. This “intentional deception or misrepresentation” can produce observations in the data that do not conform to normal observed patterns. These can therefore be regarded as anomalies or outliers. There are several statistical methods that are employed in the identification of fraud, especially in health insurance claims. Wilson [4] applied the logistic regression model to detect auto insurance fraud. This

method is only possible if there is information on both legitimate and fraudulent claims. Moreover, the labeling process may be fraught with error such that any model that is built on this may not automatically capture these errors. Ekina et al. [2] used Bayesian co-clustering model to detect health insurance fraud. This method has been used for the analysis of dyadic data, and since “conspiracy fraud” involves a connection between two entities, the Bayesian co-clustering is an appropriate method. Liu and Vasarhelyi [5] used a clustering method incorporating geo-locating information to detect fraud in Medicare/Medicaid data in the US.

Ekin et al. [6] provided a comprehensive assessment of medical fraud research. They mentioned that health fraud research is an emerging field which has gained some attention from researchers but acknowledged that more research needs to be conducted especially using Bayesian techniques. Bayesian methods have generally been proven to be an effective approach of identifying fraud. This is because Bayesian methods do not make strong statistical assumption as the frequentist counterparts.

Below are some reasons why Bayesian techniques would be appropriate for health fraud study [6].

- (i) Medical data are sparse because fraudulent cases are rare, as compared to legitimate cases.
- (ii) Data are also dynamic. Both fraudulent and legitimate patterns evolve over time.
- (iii) Fraudulent claims are not homogeneous.
- (iv) Missing values are present, and these can lead to over- or undersampling, nonrepresentativeness, and potential bias in inference.
- (v) Medical data are skewed and non-normal.

To address these challenges, this study proposes the use of Bayesian quantile regression techniques and Bayesian mixture model to identify fraud in health insurance data.

This study seeks to specify the posterior distribution of the parameters of interest given the data; identify the distribution of anomalous claims; and estimate the probability of each claim to identify potential anomalous claims.

## 2. Materials and Methods

**2.1. Bayesian Method.** From the Bayesian perspective, parameter vector of a given model is assumed to have a probability distribution referred to as the prior probability distribution. This prior probability distribution is informed by the prior knowledge about the parameter. The data from which the parameter is to be estimated are generated by a probability process, and this generating process is referred to as the likelihood. Of interest is the posterior probability which is the probability of the unknown parameter given the data. Given  $Y_i$  as a vector of observations, independently and identically distributed, and  $\theta_i$  a vector of unknown parameters, then the posterior probability distribution of the parameters is represented as

$$P(\theta_i|Y_i) = \frac{P(\theta_i) \times P(Y_i|\theta_i)}{\int_{\theta_i} P(\theta_i) \times P(Y_i|\theta_i) d\theta_i}, \quad (1)$$

where  $P(\theta_i)$  is the prior probability of the parameter  $\theta_i$  and  $P(\theta_i|Y_i)$  is the posterior probability of  $\theta_i$  given the observed data  $Y_i$ .

Generally, the prior may be chosen based on some information available and hence called informative prior. It may be chosen based on little or no information available. Such priors are called noninformative prior. The uniform distribution or Jeffrey’s prior is mostly used when there is no prior information about the unknown parameter. A prior may also either be a proper prior or an improper prior. A chosen proper prior results in a proper posterior probability. An improper prior is a distribution that does not integrate to one. However, if an improper prior is chosen properly, the resulting posterior distribution may be a proper posterior distribution.

There are other ways of choosing the prior. The prior can be chosen such that the resultant posterior distribution has the same form as the prior probability. Such priors are called conjugate.

**2.2. Quantile Regression.** This study used the quantile regression to capture the relationship between the variables and to estimate parameters of the model. This is because of its ability to capture the relationship across the entire data without imposing strong parametric assumptions about the response variable.

It is therefore an adequate technique which captures all dynamics in the data and addresses challenges associated with health insurance claims data.

Furthermore, given a set of response variables  $Y$  and explanatory variables  $X$ , the conditional cumulative distribution function is

$$F_{Y|X}(Y|X = x) = P(Y \leq y|X = x), \quad (2)$$

and for any  $\tau$ ,  $0 < \tau < 1$ , the  $\tau$ th conditional quantile function is

$$Q_\tau(Y|X = x) = \inf\{y: F_{Y|X}(y) \geq \tau\}. \quad (3)$$

The quantile function  $Q_\tau$  is monotonic increasing such that for any two quantiles  $\tau_1$  and  $\tau_2$  where  $\tau_1 < \tau_2$ ,  $Q_{\tau_1} < Q_{\tau_2}$ .

**2.3. Description of the Method.** To specify the density distribution that describes the anomalous claims, a posterior distribution of the parameters of interest ( $\beta$ ,  $\sigma$ , and  $m_i$ s) is first constructed using the quantile regression model:

$$Y = X\beta + \theta m_i + w\sqrt{(\sigma m_i u_i)}, \quad (4)$$

where  $E(Y) = Q_\tau(Y) = X\beta + \theta m_i$  and  $u_i$  has a standard normal distribution. The parameters to be estimated are the regression coefficients  $\beta$ , a scale parameter  $\sigma$ , and the mixture component  $m_i$ .

From the Bayesian perspective, given that  $Y_i$  are independently and identically random variables from which the parameters  $\beta$ ,  $\sigma$ , and  $m_i$ 's are to be estimated, the posterior distribution  $f(\beta, \sigma, m_i | y_1, y_2, \dots, y_n)$  is constructed to be proportional to some function  $g(\beta, \sigma, m_i)$  expressed as

$$f(\beta, \sigma, m_i | y_1, y_2, \dots, y_n) \propto g(\beta, \sigma, m_i). \quad (5)$$

Now

$$g(\beta, \sigma, m_i) = f(Y | \beta, \sigma, m_i) \times f(\beta | Y, \sigma, m_i) \times f(\sigma | Y, \beta, m_i) \times f(m_i | Y, \beta, \sigma), \quad (6)$$

by the factorization theorem. To estimate say  $\beta$ ,  $\sigma$  and  $m_i$  are treated as constants and their functions are included as a normalizing constant.

Thus,

$$g(\beta, \sigma, m_i) \propto f(\beta | Y, \sigma, m_i), \quad (7)$$

so that

$$f(\beta, \sigma, m_i | y_1, y_2, \dots, y_n) \propto f(\beta | Y, \sigma, m_i). \quad (8)$$

We then use the Gibbs sampler to sample  $\beta$  from the conditional distribution  $f(\beta | Y, \sigma, m_i)$ . This is the same as sampling  $\beta$  from the posterior distribution  $f(\beta, \sigma, m_i | y_1, y_2, \dots, y_n)$ . Samples are also drawn from the conditional distributions  $f(\sigma | Y, \beta, m_i)$  and  $f(m_i | Y, \beta, \sigma)$  to estimate  $\sigma$  and  $m_i$ , respectively. It is assumed that the estimates for  $m_i$ 's form a mixture distribution of unknown densities.

**2.4. The Model.** Suppose  $Y_i$ ,  $i = 1, 2, \dots, n$  are the set of independent variables identically and independently distributed,  $X_i$  are the regressors or covariates, and  $\beta_j$  are the regression coefficients.

The model for the Bayesian quantile regression as proposed by Kozumi and Kobayashi [7] is

$$\begin{aligned} Y_i &= X_{ij}\beta_j + \varepsilon_i, \quad i = 1, 2, \dots, n, \\ Y_i &= X_{ij}\beta_j + \sigma(\theta z_i + w\sqrt{z_i}u_i), \quad i = 1, 2, \dots, n, \end{aligned} \quad (9)$$

where  $\sigma$  is a scale parameter,  $\varepsilon_i = \theta\sigma z_i + w\sigma\sqrt{z_i}u_i$  are the error terms,  $\theta = ((1 - 2\tau)/(1 - \tau))$ , and  $w^2 = (2/\tau(1 - \tau))$ .  $z_i$  has a standard exponential distribution, and  $u_i$  has the standard normal distribution. We replace  $\sigma z_i$  with  $m_i$  and equation (9) becomes

$$Y_i = X_{ij}\beta_j + \theta m_i + w\sqrt{\sigma m_i}u_i, \quad (10)$$

which is the same as equation (4).

Given a random error variable  $\varepsilon$  with scale parameter  $\sigma$ , asymmetry parameter  $\tau$ , and location parameter  $\mu$ ,  $\varepsilon$  is said to have the asymmetry Laplace distribution (assumed by Yu and Moyeed [8]) given by the density

$$f(\varepsilon) = \frac{\tau(1 - \tau)}{\sigma} \exp\left[-\rho_\tau\left(\frac{\varepsilon}{\sigma}\right)\right], \quad (11)$$

where  $\rho_\tau(u) = [\tau I(u \geq 0) + (1 - \tau)I(u < 0)]|u|$  is the loss or check function,  $\tau$  is on the interval  $[0, 1]$ , and the  $\tau$ th quantile of  $\varepsilon$  is zero. The loss function assigns  $\tau$  of the errors to positive values and  $1 - \tau$  of the errors to negative values. Now if we let  $\varepsilon = Y_i - X_{ij}\beta_j - \theta m_i$ , then equation (11) becomes

$$f(Y, X, \beta, v_i) = \frac{\tau(1 - \tau)}{\sigma} \exp\left[-\rho_\tau\left(\frac{Y_i - X_{ij}\beta_j - \theta m_i}{\sigma}\right)\right]. \quad (12)$$

From equation (12),  $\mu = X_{ij}\beta_j + \theta m_i$  is the location parameter.  $v_i = \theta m_i$  can only be zero through  $\theta$  when  $\tau = (1/2)$ . The term  $m_i$ , however, is an exponential random variable and cannot be zero. The presence of anomaly in  $Y$  can be demonstrated through  $m_i$  by assuming that the predicted values of  $m_i$  have a mixture distribution.

It is somewhat intractable to derive the full conditional posterior distributions for the parameters from equation (12). Markov Chain Monte Carlo (MCMC) algorithms such as Gibbs sampler [9] and Metropolis-Hastings algorithm [10] enable one to estimate the parameters from the posterior distribution.

MCMC algorithms generate random numbers that will have stationary transition probabilities, hence an equilibrium distribution. In the Bayesian perspective, this equilibrium distribution, given some initial conditions, mimics the posterior distribution. Having done this, we can then compute statistics such as mean and variance from this distribution as estimates of the posterior distribution.

Assuming  $\varepsilon_i$  has an exponential normal mixture distribution, it can be shown that  $Y_i$  will be normally distributed with mean,  $X_{ij}\beta_j + \theta m_i$ , variance,  $w^2\sigma m_i$ , and a normal likelihood density of the form:

$$\begin{aligned} L(Y_i) &= \prod_{i=1}^n f(Y | X, \beta, m_i, \sigma) \\ &= \prod_{i=1}^n \frac{1}{w\sqrt{\sigma m_i}} \exp\left[-\frac{(Y_i - X_{ij}\beta_j - \theta m_i)^2}{2w^2\sigma m_i}\right] = \exp\left[-\sum_{i=1}^n \frac{(Y_i - X_{ij}\beta_j - \theta m_i)^2}{2w^2\sigma m_i}\right] \prod_{i=1}^n \frac{1}{w\sqrt{\sigma m_i}}. \end{aligned} \quad (13)$$

The variables  $\beta$ ,  $m_i$ , and  $\sigma$  are estimated by assuming the following priors:

$$\begin{aligned}\beta_j &\sim N(B_o, V_o), \\ m_i &\sim \exp\left(\frac{1}{\sigma}\right), \\ \sigma &\sim \text{IG}\left(\frac{n_o}{2}, \frac{s_o}{2}\right).\end{aligned}\quad (14)$$

By specifying the above conjugate priors for the parameters, their full marginal posterior distributions can be derived and we can subsequently sample from them.

From the likelihood function in equation (13) and the above prior distributions, the posterior distribution is given as

$$\begin{aligned}f(\beta, m_i, \sigma | Y_i) &\propto N(X\beta + \theta m_i, w^2 \sigma m_i) \times N(B_o, V_o) \\ &\times \exp\left(\frac{1}{\sigma}\right) \times \text{IG}\left(\frac{n_o}{2}, \frac{s_o}{2}\right),\end{aligned}\quad (15)$$

which becomes

$$\begin{aligned}f(\beta, m_i, \sigma | Y_i) &\propto \exp\left[-\sum_{i=1}^n \frac{(Y_i - X_{ij}\beta_j - \theta m_i)^2}{2w^2 \sigma m_i}\right] \prod_{i=1}^n \frac{1}{w\sqrt{\sigma m_i}} \times \frac{1}{\sqrt{2\pi V_o}} \exp\left[-\frac{(\beta_j - B_o)^2}{2V_o}\right] \\ &\times \left(\frac{1}{\sigma}\right)^n \exp\left(-\sum_{i=1}^n \frac{m_i}{\sigma}\right) \times \left(\frac{1}{\sigma}\right)^{(n_o/2)-1} \exp\left(-\frac{s_o}{2\sigma}\right).\end{aligned}\quad (16)$$

The marginal conditional posterior distributions of the parameters are derived from equation (16) as follows:

The full conditional posterior distribution of  $\beta$  is a normal distribution given as

$$\begin{aligned}f(\beta | \sigma, m_i, Y_i, X_i) &\propto \prod_{i=1}^n \frac{1}{\sqrt{\sigma m_i}} \times \frac{1}{\sqrt{2\pi V_o}} \exp\left[-\sum_{i=1}^n \frac{(Y_i - X_{ij}\beta_j - \theta m_i)^2}{2w^2 \sigma m_i}\right] \exp\left[-\frac{(\beta_j - B_o)^2}{2V_o}\right] \\ &= \frac{\sigma^{-n/2} m_i^{n/2}}{\sqrt{2\pi V_o}} \exp\left[-\sum_{i=1}^n \frac{(Y_i - X_{ij}\beta_j - \theta m_i)^2}{2w^2 \sigma m_i} - \frac{(\beta_j - B_o)^2}{2V_o}\right] \\ &= \frac{\sigma^{-n/2} m_i^{n/2}}{\sqrt{(2\pi V_o)}} \exp\left[-\frac{1}{2} \sum \left(\frac{X' X \beta^2}{w^2 \sigma V_o} + \frac{\beta^2}{V_o} - \frac{2XY\beta}{w^2 \sigma m_i} + \frac{2X\beta\theta m_i}{w^2 \sigma m_i} - \frac{2\beta\beta_o}{V_o} + K\right)\right] \\ f(\beta | \sigma, m_i, Y_i, X_i) &\propto \frac{\sigma^{-n/2} m_i^{n/2}}{\sqrt{(2\pi V_o)}} \exp\left[-\frac{1}{2} \sum \left(\frac{X' X}{w^2 \sigma m_i} + \frac{1}{V_o}\right) \beta^2 - \sum \left(\frac{X(Y - \theta m_i)}{w^2 \sigma m_i} + \frac{\beta_o}{V_o}\right) \beta + K\right],\end{aligned}\quad (17)$$

where  $K$  is a constant that does not depend on  $\beta$ .

Note that a function  $x \mapsto \exp(ax^2 + bx + c)$  with domain on the whole real line is an unnormalized probability density function (pdf) if and only if  $a < 0$ . In this case, the unnormalized pdf of the normal distribution with mean ( $\mu$ ) and variance ( $\sigma^2$ ) is

$$\begin{aligned}\mu &= -\frac{b}{2a}, \\ \sigma^2 &= -\frac{1}{2a}.\end{aligned}\quad (18)$$

Comparing equation (17) to  $x \mapsto \exp(ax^2 + bx + c)$  yields

$$\sigma^2 = \frac{1}{\sum \left( (X'X/w^2\sigma m_i) + (1/V_o) \right)} = \left[ \sum \left( \frac{X'X}{w^2\sigma m_i} + \frac{1}{V_o} \right) \right]^{-1}, \quad \bar{V}^{-1} = \sum_{i=1}^n \left( \frac{X'_i X_i}{w^2\sigma m_i} + \frac{1}{V_o} \right), \quad (20)$$

$$\mu = b\sigma^2 = \sum \left( \frac{X(Y - \theta m_i)}{w^2\sigma m_i} + \frac{\beta_o}{V_o} \right) \left[ \sum \left( \frac{X'X}{w^2\sigma m_i} + \frac{1}{V_o} \right) \right]^{-1}. \quad \bar{B} = \bar{V} \left[ \sum_{i=1}^n \left( \frac{X_i(Y_i - \theta m_i)}{w^2\sigma m_i} + \frac{\beta_o}{V_o} \right) \right].$$

If we let  $\sigma^2 = \bar{V}$  and  $\mu = \bar{B}$ , then the distribution of  $\beta$  is normal with parameters  $\bar{B}$  and  $\bar{V}$  where

It can also be derived from equation (16) that the full conditional distribution of  $m_i$  is

$$\begin{aligned} f(m_i|\beta, \sigma, \lambda_i, \gamma, Y_i) &\propto \left(\frac{1}{\sigma}\right)^n \prod \frac{1}{w\sqrt{\sigma m_i}} \exp \left[ -\sum_{i=1}^n \frac{(Y_i - X_{ij}\beta_j - \theta m_i)^2}{2w^2\sigma m_i} \right] \exp \left( -\sum_{i=1}^n \frac{m_i}{\sigma} \right) \\ &\propto \frac{1}{m_i^{n/2}} \exp \left[ -\sum_{i=1}^n \left( \frac{(Y - X\beta)^2 - 2(Y - X\beta)\theta m_i - (\theta m_i)^2}{2w^2\sigma m_i} + \frac{m_i}{\sigma} \right) \right] \\ &\propto m_i^{-n/2} \exp \left[ -\sum_{i=1}^n \left( \frac{2(Y - X\beta)\theta m_i}{2w^2\sigma m_i} + \frac{(Y - X\beta)^2}{2w^2\sigma m_i} - \frac{\theta^2 m_i^2}{2w^2\sigma m_i} - \frac{m_i}{\sigma} \right) \right] \\ &\propto m_i^{-n/2} K \exp \left[ \sum_{i=1}^n \frac{(Y - X\beta)^2}{2w^2\sigma} m_i^{-1} - \sum_{i=1}^n \left( \frac{\theta}{2w^2\sigma} + \frac{1}{\sigma} \right) m_i \right] \\ f(m_i|\beta, \sigma, \lambda_i, \gamma, Y_i) &\propto m_i^{-n/2} \exp \left[ -\frac{1}{2} \sum_{i=1}^n \frac{(Y - X\beta)^2}{w^2\sigma} m_i^{-1} + \sum_{i=1}^n \left( \frac{\theta}{w^2\sigma} + \frac{2}{\sigma} \right) m_i \right]. \end{aligned} \quad (21)$$

The full conditional posterior distribution of the  $m_i$  is the generalized inverse Gaussian distribution with parameters  $(1/2)$ ,  $\hat{\mu}$ , and  $\hat{\nu}$  given as

$$f(m_i|\beta, \sigma, Y_i, X_i) \propto m_i^{-(n/2)} \exp \left[ -\frac{1}{2} (\hat{\mu}_i^2 m_i^{-1} + \hat{\nu}^2 m_i) \right], \quad (22)$$

where  $\hat{\mu}_i^2 = \sum [(Y_i - X_{ij}\beta_j)^2 / \sigma w^2]$  and  $\hat{\nu}^2 = \sum ((2/\sigma) + (\theta/\sigma w^2))$ .

Again from equation (16), the full conditional distribution of  $\sigma$  follows the inverse gamma distribution given as

$$\begin{aligned} f(\sigma|\beta, m_i, Y, X) &\propto \left(\frac{1}{\sigma}\right)^{n/2} \left(\frac{1}{\sigma}\right)^n \left(\frac{1}{\sigma}\right)^{(n_o/2)-1} \exp \left[ -\sum_{i=1}^n \frac{(Y - X\beta - \theta m_i)^2}{2w^2\sigma m_i} - \sum_{i=1}^n \frac{m_i}{\sigma} - \frac{s_o}{2\sigma} \right], \\ f(\sigma|\beta, m_i, Y, X) &\propto \left(\frac{1}{\sigma}\right)^{((3n+n_o)/2)-1} \exp \left[ -\frac{1}{\sigma} \left( \frac{s_o}{2} + \sum m_i + \sum \frac{(Y_i - X_{ij}\beta_j - \theta m_i)^2}{2w^2 m_i} \right) \right]. \end{aligned} \quad (23)$$

This is an inverse gamma distribution with shape parameter  $(3n + n_o)/2$  and scale parameter  $((s_o/2) + \sum m_i + \sum ((Y_i - X_{ij}\beta_j - \theta m_i)^2 / (2w^2 m_i)))$ .

**2.5. Estimating the Mixture Components of  $m_i$ .** We used the R package for the Bayesian mixture model, "BayesMix," to estimate the mixture components of  $m_i$ . BayesMix provides Gibbs sampling of the posterior distribution, a method

to set up the model and specify the priors and initial values required for the Gibbs sampler.

To run the Gibbs sampler, we specify the prior distributions of the parameters and initial values for the hyperparameters in the proposed models. The regression coefficients are normally distributed with mean  $\beta_o$  and variance  $V_o$ . We set initial values for these parameters as 0 and 10 (with precision as 1/10), respectively.



$m_i$ 's are exponentially distributed with rate parameter  $1/\sigma$ . The prior distribution of the scale parameter  $\sigma$  is an inverse gamma distribution with hyperparameters  $(n_o/2)$  and  $(s_o/2)$ . The quantity  $n_o$  is the prior effective sample size for  $\sigma$ , and  $s_o$  is prior point estimate for  $\sigma$ . For our analysis, we set  $1/\sigma$  to prec which is the rate of the exponential variable  $m_i$ . The initial values for prec are generated randomly as uniform values with interval 0 and 1 and 0 and 2. This choice is subjective and arbitrary but with the hope that the data will be informative enough to produce a stationary posterior distribution.

**2.6. Model Evaluation.** The proposed model is evaluated by replicating new values of  $Y$ ,  $Y^{\text{rep}}$  and determining how close these values are to the original ones. The approach is called posterior consistency and involves using a posterior predictive distribution.

The posterior predictive distribution is given by

$$p(Y^{\text{rep}}|Y) = \int_{\Theta} f(Y^{\text{rep}}|\theta_i) * f(\theta_i|Y) d\theta_i, \quad (24)$$

where  $\Theta$  is our parameter space and  $\theta_i$  is the set of parameters we estimated from our model. We sample from this distribution as follows:

- (1) Sample  $\theta_i$  from posterior distribution  $f(\theta_i|Y)$ .
- (2) Substitute the values of  $\theta_i$  above in  $f(Y|\theta_i)$  and draw  $Y^{\text{rep}}$ .

To compare the predictive posterior distribution to our model, we can compute some statistics such as the mean, standard deviation, and order statistics and compare the values. Graphical approaches may also be used. The histogram of  $Y^{\text{rep}}$  gives the distribution of the predictive posterior distribution, and this can be compared to the model distribution of  $Y$ .

**2.6.1. Computing the Probability Scores.** The densities of the component mixtures are estimated using the Bayesian method. Now, let  $\lambda = 0$  and  $\lambda = 1$  correspond to the density distribution with normal claims and density distribution with anomalous claims (which will have the higher mean), respectively.

From the above preamble, we compute the probability of each of the claim belonging to anomalous claim class,  $P(Y|\lambda = 1)$ . These probabilities are ranked for all observations, and those with values greater than 0.5 are classified into the anomalous claim class.

We treat  $m_i$  as a mixture of two normal densities and estimate the proportion and density of each component. For  $\mu_1$  and  $\mu_2$  (anomalous claims) while the smaller one will be associated with  $\lambda = 1$  (anomalous claims) while the smaller one will be associated with  $\lambda = 0$  (normal claims).  $p(Y|\lambda_i = 1)$ , which describes the outlier distribution, can be computed as

$$p(Y|\lambda = 1) = \frac{p(\lambda = 1)p(\lambda = 1, Y, \sigma, m_i)}{p(\lambda = 1)p(\lambda = 1, Y, \sigma, m_i) + p(\lambda = 0)p(\lambda = 0, Y, \sigma, m_i)}. \quad (25)$$

**2.6.2. Sensitivity, Specificity, Precision, and Accuracy.** The other metrics of performance evaluation can be calculated from Table 1.

$$\begin{aligned} \text{Sensitivity} &= \frac{TP}{TP + FN} = \frac{\text{number of true positive classifications}}{\text{number of all positive classifications}}, \\ \text{specificity} &= \frac{TN}{FP + TN} = \frac{\text{number of true negative classifications}}{\text{number of all negative classifications}}, \\ \text{accuracy} &= \frac{TP + TN}{TP + FN + FP + TN} = \frac{\text{number of correct classifications}}{\text{number of all classifications}}, \\ \text{precision} &= \frac{TP}{TP + FP}. \end{aligned} \quad (26)$$

**2.7. Data Acquisition.** Secondary data on health insurance claims by 115 people suffering from various cardiovascular diseases located in different states in the USA were used to run the proposed model. The response variable is the claim amount while the covariates/independent variables are location (state), diagnosis, gender, type of care, pharmacy, and type of health facility. All the covariates are categorical

variables. Table 2 shows the claims data but with the nominal values for the independent variables.

### 3. Results and Discussion

We run the Gibbs sampler at quantile values,  $\tau = 0.95, 0.75, 0.50, 0.25$ , to observe the trend of the changes in  $\beta$ 's (regression coefficients) and  $\sigma$  (scale parameter). Table 3 shows the estimates of the regression coefficients at the specified quantiles ( $\tau = 0.95, 0.75, 0.50, 0.25$ ).

From Table 3, the quantile value  $\tau = 0.75$  is preferred to estimate the anomalous claims in the data. The values of the estimated parameters at  $\tau = 0.75$  help satisfy the assumption that an anomalous/fraudulent health insurance claim is usually located at the right of a legitimate claim. This suggests that we can estimate the parameters of the proposed model even at  $\tau = 0.95$ . However, the potential reduction factor at  $\tau = 0.75$  and  $\tau = 0.95$  is 1.01 and 1.08, respectively. This makes  $\tau = 0.75$  the preferred quantile value.

The mean, standard deviation, and standard error of the mean for each parameter estimated at  $\tau = 0.75$  are given in Table 4.

TABLE 1: Confusion matrix.

| Classification | Actual   |          | Total             |
|----------------|----------|----------|-------------------|
|                | Positive | Negative |                   |
| Positive       | TP       | FP       | TP + FP           |
| Negative       | FN       | TN       | FP + TN           |
| Total          | TP + FN  | FP + TN  | TP + TN + FP + FN |

TABLE 2: Claims data with other covariate information.

| Claim amount | Location | Diagnosis | Gender | Type of care | Pharmacy | Type of health facility |
|--------------|----------|-----------|--------|--------------|----------|-------------------------|
| 7257         | 1        | 1         | 2      | 1            | 1        | 1                       |
| 5835         | 2        | 1         | 1      | 1            | 2        | 1                       |
| 7023         | 2        | 1         | 1      | 1            | 3        | 2                       |
| 9284         | 2        | 1         | 1      | 1            | 1        | 2                       |
| 8148         | 2        | 1         | 1      | 1            | 3        | 2                       |
| 9024         | 3        | 1         | 1      | 1            | 4        | 3                       |
| ⋮            | ⋮        | ⋮         | ⋮      | ⋮            | ⋮        | ⋮                       |
| 8256         | 3        | 1         | 1      | 1            | 3        | 1                       |
| 8941         | 2        | 1         | 1      | 1            | 3        | 2                       |
| 8385         | 3        | 1         | 1      | 1            | 4        | 2                       |
| 11731.2      | 3        | 2         | 1      | 2            | 5        | 4                       |

Location: 1, California; 2, Georgia; 3, Alabama; 4, Virginia. Diagnosis: 1, acute myocardial infarction; 2, heart failure. Gender: 1, female; 2, male. Type of care: 1, outpatient; 2, inpatient. Pharmacy: 1, heparin; 2, nitroglycerin; 3, eptifibatide; 4, clopidogrel; 5, furosemide; 6, metoprolol; 7, bisoprolol; 8, hydrochlorothiazide. Type of health facility: 1, ambulatory service; 2, outpatient clinic; 3, acute care; 4, hospital (source: <https://data.cms.gov>).

TABLE 3: Posterior parameter estimate at various quantile values.

| Parameter | $\tau = 0.95$ | $\tau = 0.75$ | $\tau = 0.5$ | $\tau = 0.25$ |
|-----------|---------------|---------------|--------------|---------------|
| $\beta_1$ | 8.832456      | 8.81072       | 8.586895     | 6.582426      |
| $\beta_2$ | 0.044465      | 0.056238      | 0.010074     | 0.09637       |
| $\beta_3$ | 0.416721      | 0.273557      | -0.03593     | 0.23371       |
| $\beta_4$ | 0.094641      | 0.127956      | -0.00163     | 0.229348      |
| $\beta_5$ | 0.030974      | 0.217873      | 0.346222     | 0.393446      |
| $\beta_6$ | -0.03138      | -0.03801      | -0.0054      | -0.06249      |
| $\beta_7$ | -0.00134      | 0.000169      | -0.00819     | 0.009342      |
| $\sigma$  | 8.980103      | 3.062348      | 3.313797     | 0.803441      |

TABLE 4: Empirical mean and standard deviation for each parameter, with standard error of the mean at  $\tau = 0.75$ .

| Parameter  | Mean       | SD      | Naive SE | Time-series SE |
|------------|------------|---------|----------|----------------|
| $\beta[1]$ | 8.8107204  | 0.24925 | 0.000557 | 0.005705       |
| $\beta[2]$ | 0.0562377  | 0.06056 | 0.000135 | 0.000989       |
| $\beta[3]$ | 0.2735569  | 0.78952 | 0.001765 | 0.084209       |
| $\beta[4]$ | 0.1279555  | 0.10345 | 0.000231 | 0.001989       |
| $\beta[5]$ | 0.2178729  | 0.78366 | 0.001752 | 0.075442       |
| $\beta[6]$ | -0.0380131 | 0.04904 | 0.00011  | 0.000912       |
| $\beta[7]$ | 0.0001692  | 0.08409 | 0.000188 | 0.001818       |
| $m_i[1]$   | 3.5536795  | 1.6551  | 0.003701 | 0.008894       |
| $m_i[2]$   | 3.8391055  | 1.75822 | 0.003932 | 0.008311       |
| $m_i[3]$   | 3.2403607  | 1.48188 | 0.003314 | 0.005923       |
| $m_i[4]$   | 2.8009853  | 1.29663 | 0.002899 | 0.005247       |
| $m_i[5]$   | 2.9037541  | 1.33627 | 0.002988 | 0.005209       |
| ⋮          | ⋮          | ⋮       | ⋮        | ⋮              |
| $\sigma$   | 3.062346   | 32698   | 0.00073  | 0.00568        |

We provide the residual and normal Q-Q plots of the model in Figure 1 and also the trace plot and density of  $\beta$  and  $\sigma$  in Figure 2.

The results from Figure 2 (diagram of densities) satisfy the assumption of inverse gamma and normal prior distributions made for  $\sigma$  and  $\beta$ , respectively.

Our objective is to identify potential fraudulent claims through the estimated values of the unobserved variable  $m_i$ . Figure 3 shows the histogram and density of  $m_i$  at  $\tau = 0.75$ .

**3.1. Identifying Anomalous Claims.** The “BayesMix package” in R was used to set up an MCMC sampler to identify the two components of the mixture of  $m_i$ 's. The mean ( $\mu$ ), sigma ( $\sigma$ ), and proportion ( $\gamma$ ) of each component were estimated. Note that the probability of a claim being anomalous is  $P(\lambda = 1) = \gamma[2]$  and the probability of a claim being normal is  $P(\lambda = 0) = \gamma[1]$ . The results are shown in Table 5.

The trace plots and densities of the mixture parameters  $\mu$  and  $\sigma^2$  are shown in Figures 4 and 5, respectively.

The probability of each of the claims belonging to anomalous class,  $P(Y|\lambda = 1)$ , is through  $m_i$ 's, so we use  $\mu[2] = 14.17$  for density of  $P(Y, m, \lambda = 1)$  and  $P(\lambda = 1) = 0.1651$ . Moreover, the density of  $P(\lambda = 0)P(Y, m, \lambda = 0)$  is  $\mu[1] = 10.48$  and  $P(\lambda = 0) = 0.8349$ . Assuming the mixture components are

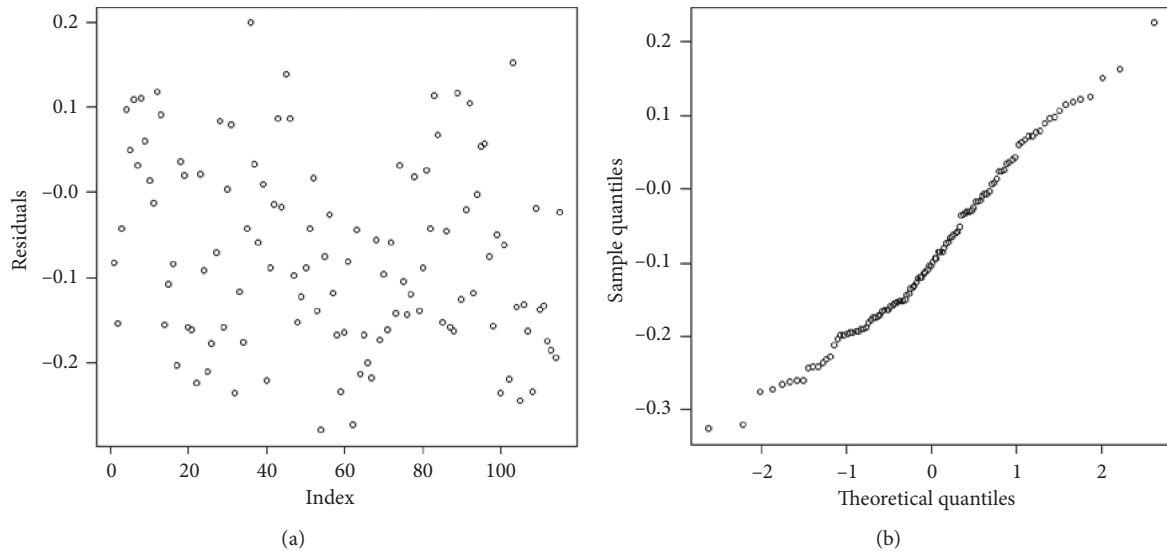
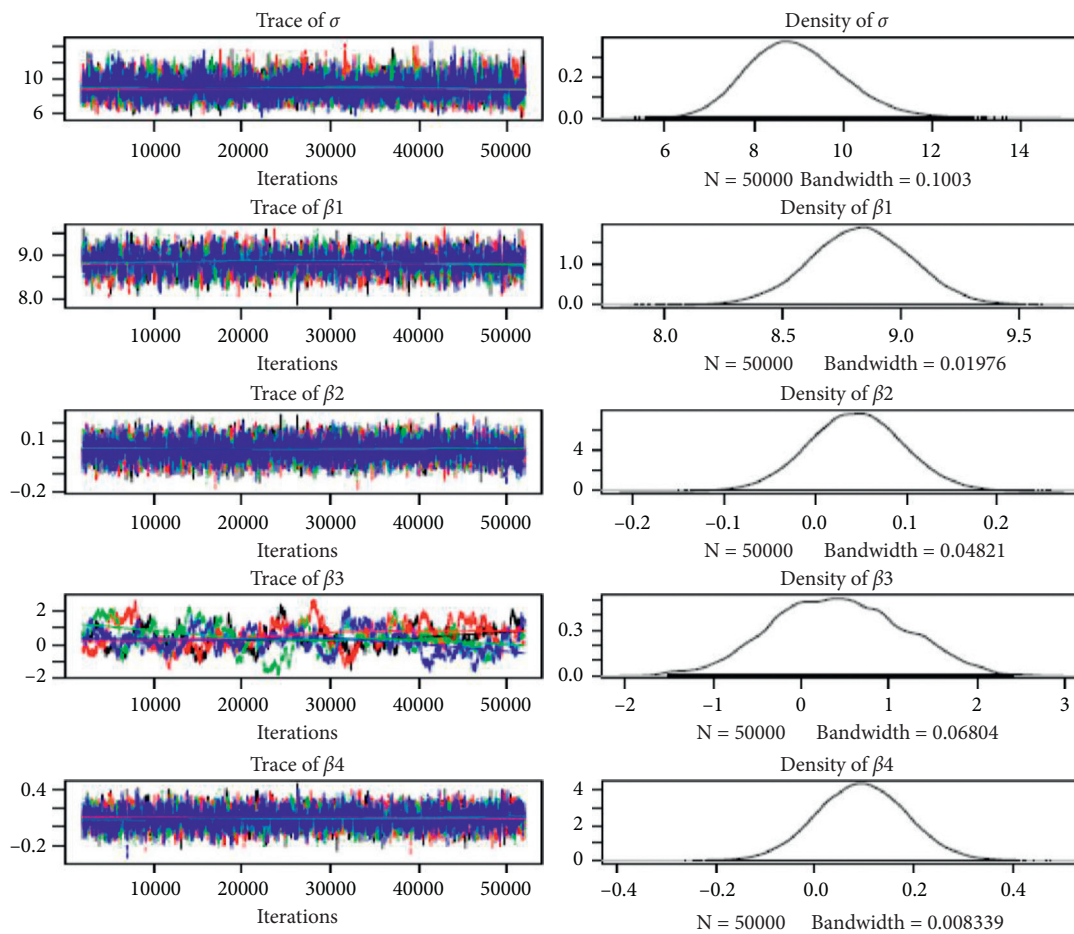


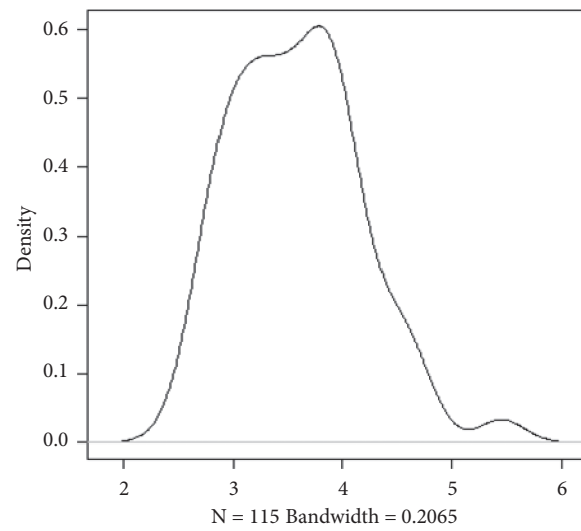
FIGURE 1: Residuals and normal Q-Q plots.

FIGURE 2: Trace plot and density of  $\beta$  and  $\sigma$  at  $\tau = 0.75$ .

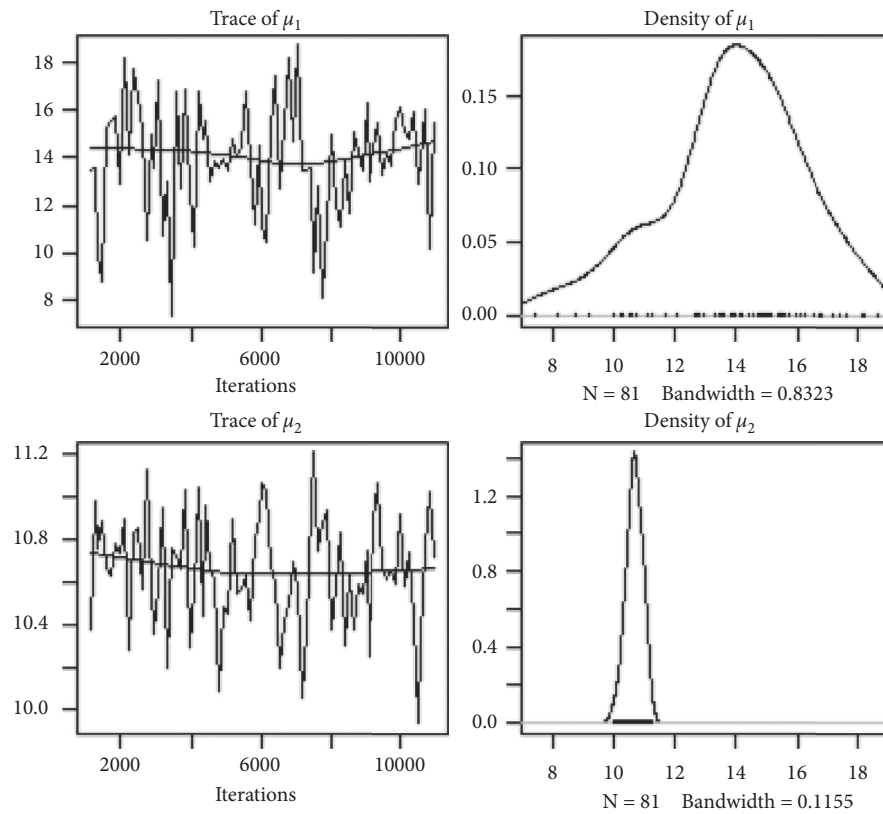
normally distributed, we compute  $P(Y|\lambda = 1)$  using equation (25) and provide ranked results with claims in Table 6. It is evident from Table 6 that each of the first 10 claims with claim

amounts 9440, 9638, 9316,  $\dots$ , 9855 has  $P(Y|\lambda = 1)$  greater than 0.5; hence, they would each be classified into the anomalous class.



FIGURE 3: Density of  $m_i$ .TABLE 5: Empirical mean, standard deviation, and 95%  $\mu$ ,  $\gamma$ , and  $\sigma^2$ .

| Parameter   | Mean   | SD     | 2.5%     | 97.5%  |
|-------------|--------|--------|----------|--------|
| $\mu[1]$    | 10.48  | 0.6272 | 9.138    | 11.08  |
| $\mu[2]$    | 14.17  | 2.0379 | 10.964   | 18.73  |
| $\gamma[1]$ | 0.8349 | 0.25   | 0.029028 | 0.9926 |
| $\gamma[2]$ | 0.1651 | 0.25   | 0.007377 | 0.9710 |
| $\sigma^2$  | 3.4257 | 0.8005 | 2.2058   | 4.8034 |

FIGURE 4: Trace plot and density of  $\mu$ .

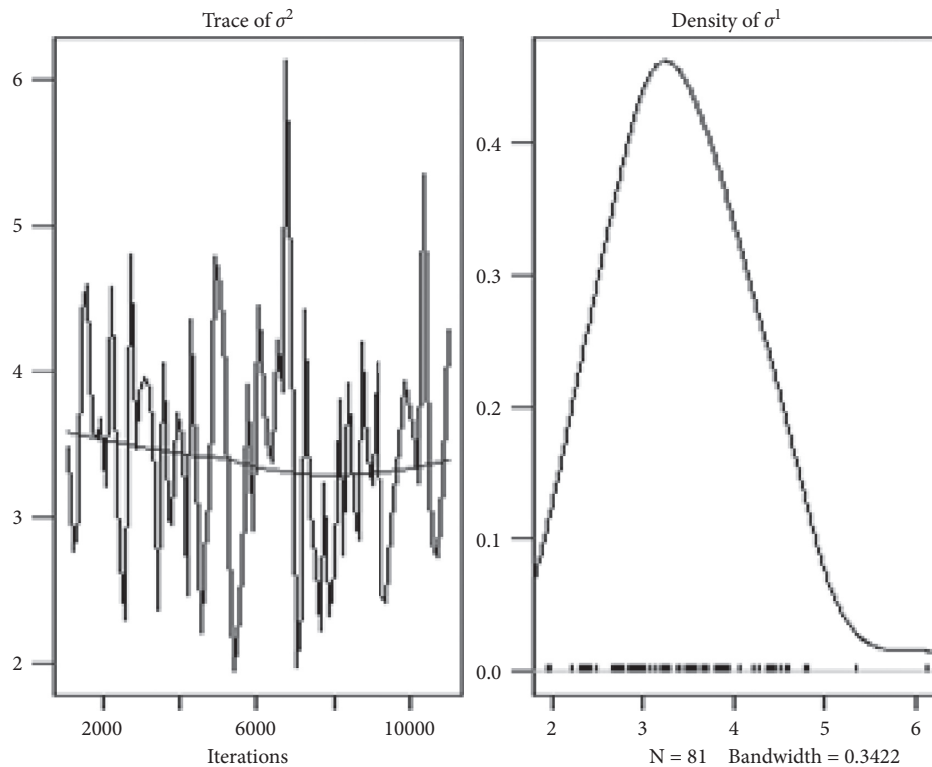
FIGURE 5: Trace plot and density of  $\sigma^2$ .

TABLE 6: Claims and their probabilities of being anomalous.

| $m_i$      | Claim amount | $p(Y, \lambda = 0, m_i)$ | $p(Y, \lambda = 1, m_i)$ | $p(Y \lambda = 1)$ |
|------------|--------------|--------------------------|--------------------------|--------------------|
| $m_i[36]$  | 9440         | $2.67\text{E} - 15$      | 0.000233                 | 1                  |
| $m_i[103]$ | 9638         | $3.17\text{E} - 11$      | 0.00063                  | 1                  |
| $m_i[45]$  | 9316         | $1.24\text{E} - 10$      | 0.000735                 | 0.999999           |
| $m_i[8]$   | 8941         | $2.95\text{E} - 09$      | 0.001063                 | 0.999982           |
| $m_i[89]$  | 9117         | $3.30\text{E} - 09$      | 0.001078                 | 0.999981           |
| $m_i[83]$  | 9241         | $3.65\text{E} - 09$      | 0.001091                 | 0.999979           |
| $m_i[6]$   | 9024         | $1.55\text{E} - 08$      | 0.001301                 | 0.999924           |
| $m_i[12]$  | 9937         | $4.02\text{E} - 08$      | 0.001465                 | 0.999826           |
| $m_i[4]$   | 9284         | $8.22\text{E} - 08$      | 0.001604                 | 0.999675           |
| $m_i[92]$  | 9855         | $2.92\text{E} - 07$      | 0.00189                  | 0.999019           |
| $\vdots$   | $\vdots$     | $\vdots$                 | $\vdots$                 | $\vdots$           |
| $m_i[41]$  | 6661         | 0.407214                 | 0.030185                 | 0.011533           |
| $m_i[93]$  | 6624         | 0.935536                 | 0.066375                 | 0.011045           |
| $m_i[82]$  | 12272        | 0.466905                 | 0.032471                 | 0.010828           |
| $m_i[55]$  | 7356         | 0.523207                 | 0.034726                 | 0.010339           |
| $m_i[75]$  | 9562.8       | 0.885195                 | 0.05808                  | 0.010222           |
| $m_i[90]$  | 6115         | 0.881762                 | 0.057646                 | 0.010186           |
| $m_i[80]$  | 7939         | 0.878415                 | 0.057234                 | 0.010152           |
| $m_i[1]$   | 7257         | 0.599579                 | 0.038033                 | 0.009886           |
| $m_i[24]$  | 6522         | 0.62609                  | 0.039274                 | 0.009777           |
| $m_i[70]$  | 9712.3       | 0.811345                 | 0.050588                 | 0.009719           |
| $m_i[16]$  | 7471         | 0.651635                 | 0.040528                 | 0.009695           |
| $m_i[61]$  | 10472.8      | 0.80352                  | 0.049954                 | 0.009691           |
| $m_i[50]$  | 7206         | 0.675306                 | 0.04175                  | 0.009638           |
| $m_i[97]$  | 11065.6      | 0.781346                 | 0.04827                  | 0.009631           |
| $m_i[47]$  | 6831         | 0.687836                 | 0.042424                 | 0.009615           |

TABLE 7: Performance metrics.

| Metrics     | Value (%) |
|-------------|-----------|
| Sensitivity | 100       |
| Specificity | 90        |
| Accuracy    | 92        |
| Precision   | 74        |

The result of our analysis using classification metrics is summarised in Table 7.

#### 4. Conclusions and Recommendations

The model identified 25 claims (about 22%) estimated at quantile level  $\tau = 0.75$  as belonging to the anomalous class and can thus be assumed to be fraudulent claims. The first 10 of these are shown in Table 6.

The model correctly identified all anomalous claims (sensitivity of 100%) and gave a high specificity value of 90%. The overall accuracy of the model was 92% with a precision of 74%. This study therefore recommends Bayesian quantile regression as a viable model for anomaly detection. It is worthy of note that the study assumed a normal density mixture for the  $m_i$ 's to be able to estimate the anomalous cluster in the data. Therefore, a choice of a different distribution for the  $m_i$ 's could result in different findings. Also, instead of using a mixture model, a variable selection

method through data augmentation could be used to aid in selecting only appropriate variables for the model.

### Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

### Conflicts of Interest

The authors declare that there are no conflicts of interest.

### References

- [1] J. Novignon, S. A. Olakojo, and J. Nonvignon, "The effects of public and private health care expenditure on health status in sub-saharan africa: new evidence from panel data analysis," *Health Economics Review*, vol. 2, p. 22, 2012.
- [2] T. Ekin, F. Ieva, F. Ruggeri, and R. Soyer, "Application of bayesian methods in detection of healthcare fraud," *Chemical Engineering Transaction*, vol. 33, 2013.
- [3] W.-S. Yang and S.-Y. Hwang, "A process-mining framework for the detection of healthcare fraud and abuse," *Expert Systems with Applications*, vol. 31, no. 1, pp. 56–68, 2006.
- [4] J. H. Wilson, "An analytical approach to detecting insurance fraud using logistic regression," *Journal of Finance and Accountancy*, vol. 1, p. 1, 2009.
- [5] Q. Liu and M. Vasarhelyi, "Healthcare fraud detection: a survey and a clustering model incorporating geo-location information," in *Proceedings of the 29th World Continuous Auditing and Reporting Symposium (29WCARS)*, Brisbane, Australia, November 2013.
- [6] T. Ekin, F. Ieva, F. Ruggeri, and R. Soyer, "Statistical medical fraud assessment: exposition to an emerging field," *International Statistical Review*, vol. 86, no. 3, pp. 379–402, 2018.
- [7] H. Kozumi and G. Kobayashi, "Gibbs sampling methods for bayesian quantile regression," *Journal of Statistical Computation and Simulation*, vol. 81, no. 11, pp. 1565–1578, 2011.
- [8] K. Yu and R. A. Moyeed, "Bayesian quantile regression," *Statistics & Probability Letters*, vol. 54, no. 4, pp. 437–447, 2001.
- [9] M. A. Tanner and W. H. Wong, "The calculation of posterior distributions by data augmentation," *Journal of the American Statistical Association*, vol. 82, no. 398, pp. 528–540, 1987.
- [10] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.