Research Article

A Cost of Misclassification Adjustment Approach for Estimating Optimal Cut-Off Point for Classification

O.-A. Ampomah Ⓚ, R. Minkah Ⓚ, G. Kallah-Dagadu Ⓚ, and E. N. N. Nortey Ⓚ

Department of Statistics and Actuarial Science, College of Basic and Applied Sciences, University of Ghana, Accra, Ghana

Correspondence should be addressed to E. N. N. Nortey; ennnortey@ug.edu.gh

Received 25 August 2022; Revised 5 January 2024; Accepted 5 March 2024; Published 15 May 2024

Abstract
Classification is one of the main areas of machine learning, where the target variable is usually categorical with at least two levels. This study focuses on deducing an optimal cut-off point for continuous outcomes (e.g., predicted probabilities) resulting from binary classifiers. To achieve this aim, the study modified univariate discriminant functions by incorporating the error cost of misclassification penalties involved. By doing so, we can systematically shift the cut-off point within its measurement range till the optimal point is obtained. Extensive simulation studies were conducted to investigate the performance of the proposed method in comparison with existing classification methods under the binary logistic and Bayesian quantile regression frameworks. The simulation results indicate that logistic regression models incorporating the proposed method outperform the existing ordinary logistic regression and Bayesian regression models. We illustrate the proposed method with a practical dataset from the finance industry that assesses default status in home equity.

1. Introduction

Classification is one of the main areas of machine learning, where the target variable is qualitative, with at least two groups. If the target variable consists of only two groups, it is called binary. Applicable areas include loan administration, image processing, and survival analysis. Commonly used classification techniques can be categorized into four groups: supervised, unsupervised, semisupervised, and hybrid. The supervised method uses the target variable to classify data points into distinct groups and make predictions. Using the target input and output, the model can measure its accuracy and learn from them. Without a target variable, the unsupervised method is typically recommended to identify uncharacterized patterns in the data set. This method gathers data and distinguishes between data points with expected deviations from the successive data points. As they do not require any target information, the unsupervised method may serve as the first stage in separating data points that do not follow expected patterns, thus classifying them as anomalies. However, semisupervised methods are used when the target information for a particular data set is incomplete. This model first learns the part of the data set containing target scores and uses that to predict the other part without target scores. Lastly, the hybrid methods are just a combination of both the supervised and unsupervised methods.

Any binary classification model aims to classify each data point into one of two distinct groups. However, the results of most binary classification models are usually predicted probabilities [1]. A cut-off point is applied to these predicted probabilities to classify data points into the present (1)-absent (0) maps. Thus, choosing a cut-off point for binary classification is a vital step for decision-making as it may have severe consequences on the model’s accuracy. The default cut-off probability is 0.5. However, this may not result in higher prediction accuracy as data sets are usually imbalanced [1]. Binary classification models are subject to two types of errors: false-positive (FP) and false-negative (FN). These rates, FP and FN, are characterized by error cost functions (i.e., the cost of misclassifying a data point as group 1 when it belongs to group 2 or vice versa). A good
classification model aims to minimize the misclassification function’s expected error cost. However, due to the challenges involved in accurately specifying the error cost of misclassification penalties, in many applicable areas, researchers usually assume an equal cost of misclassification [2–5]. However, this has its drawbacks; for example, Ling and Sheng [2] indicate that the variation between different misclassification costs can be quite large. In addition, Johnson and Wichern [6] state that any classification rule that ignores the error cost of misclassification might be problematic.

As a result, cost-sensitive machine learning has expanded over time due to its ability to integrate financial decision-making considerations such as information acquisition and decision-making error costs [2, 7]. The aim of this type of learning is to minimize the total misclassification cost [2]. Also, cost-sensitive learning plays a significant role in classification model evaluation [8]. Researchers in this field aim at choosing cut-off points to reduce the misclassification rate.

Bayesian methods have recently been used to address binary classification problems [9–11]. Nortey et al. [10] demonstrated that Bayesian quantile regression is a viable classification model for anomaly detection. Often, it is much easier to postulate the error cost ratios than to state their respective component parts [6]. For example, it may be difficult to accurately specify the costs (in appropriate units) of misclassifying a loan application as nonrisky when, in fact, the application is risky and also misclassifying a loan application as risky when, in fact, the application is nonrisky. However, a realistic number of these error cost ratios can be obtained and used to identify an optimal cut-off point for classification. Determining an optimal cut-off point requires simultaneous assessment of the test sensitivity and specificity [12]. The optimal cut-off point is the one that produces the highest sum of test sensitivity and specificity. Thus, it should be chosen as the point that accurately classifies most data points correctly and most minor of them incorrectly [13].

In addition, many studies concentrate on the receiver operating characteristic (ROC) curve and the corresponding area under the curve (AUC), a graph that measures the diagnostic ability of any binary classification model to determine an optimal cut-off point [14, 15]. The ROC curve plots the sensitivity (the true-positive rate) against the complement of specificity (the false-positive rate) for all distinctive cut-off points. Other criteria are also introduced by assuming specific values or defining a linear combination or function of sensitivity and specificity (see, e.g., [12, 13, 16–18]). Moreover, Liu [19] proposed the concordance probability method, which defines the optimal cut-off point as the maximizer of the product of sensitivity and specificity of the model.

In view of the above reasons, this study seeks to develop a cost-sensitive machine learning method that is relatively efficient and consistent based on univariate discriminant functions. The approach modifies the univariate discriminant function to incorporate the cost ratios, thus avoiding the equality of error cost of misclassification assumption.

The remainder of the paper is organized as follows. In Section 2, we set out the framework for estimating the parameter of interest, i.e., the concordance probability, and our approach for modifying the univariate discriminant function to incorporate the cost ratios. In Section 3, we conduct a simulation study to assess the performance of the proposed methods with the existing ones in the literature. Section 4 presents a practical application of the proposed method. Lastly, general conclusions from the simulation results are presented in Section 5.

2. Materials and Methods

This section presents the model framework for estimating the parameters of interest.

2.1. Binary Logistic Regression. Let $p$ be the predicted probability for a binary response variable, $Y = 1$, using an input variable, $x = (x_1, x_2, \ldots, x_k)$'. Then, the logistic response function is modelled for multiple covariates:

$$ p = P(Y = 1|x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k}}. \tag{1} $$

The model in (1) is nonlinear and transformed into linearity using the logit response function. From (1), the logit function is given in the following equation:

$$ \ln \left( \frac{P}{1-P} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k. \tag{2} $$

The logistic regression coefficient in (2) is estimated by the method of maximum likelihood.

2.2. Bayesian Quantile Regression. Let $y_i$ be a response variable and $x_i$ be a $k \times 1$ vector of independent variables for the $i$th observation. Let $q_{r_i}(x_i)$ denote the $r_i$th $(0 < r < 1)$ quantile regression function of $y_i$ given $x_i$. Suppose that the relationship between $q_{r_i}(x_i)$ and $x_i$ can be modelled as $q_{r_i}(x_i) = x_i \beta$, where $\beta$ is a vector of unknown parameters of interest. Then, we consider the quantile regression model given in the following equation:

$$ y_i = x_i \beta + \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{3} $$

where $\varepsilon_i$ is the error term whose distribution (with density say $g(.)$) is restricted to have the $r_i$th quantile equal to zero; that is, $\int_{-\infty}^{0} g(\varepsilon) \, d\varepsilon = \tau$. The error density $g(.)$ is often left unspecified in the classical literature of Kozumi and Kobayashi [20]. Thus, quantile regression estimation of $\beta$ proceeds by minimizing

$$ \sum_{i=1}^{n} \rho_{r_i}(y_i - x_i \beta), \tag{4} $$

where $\rho_{r_i}(.)$ is the loss (or check) function and is defined as

$$ \rho_{r_i}(u) = u[\tau - I(u < 0)], \tag{5} $$

and $I(.)$ denotes the usual indicator function.
Kozumi and Kobayashi [20] considered the linear model from (3) and assumed that \( \epsilon_i \) has a three-parameter asymmetric Laplace distribution with a density function given by

\[
g(\epsilon_i) = \frac{\tau (1 - \tau)}{\sigma} \exp\left\{-\beta (\frac{\epsilon_i}{\sigma})\right\}, \tag{6}
\]

where \( \sigma > 0 \) is the scale parameter. The parameter, \( \tau \), determines the skewness of the distribution, and the \( \tau \)th quantile of this distribution is zero. To develop a Gibbs sampling algorithm for the quantile regression model, Kozumi and Kobayashi [20] utilized a mixture representation based on exponential and normal distribution, as summarised as follows.

Let \( z \) be a standard exponential variable and \( u \) be a standard normal variable. If a random variable \( \epsilon \) follows a three-parameter asymmetric Laplace density with a density as stated in (6), then one can represent \( \epsilon \) as a location-scale mixture of normals given in the following equation:

\[
\epsilon = \sigma (\phi z + \eta \sqrt{z} u), \tag{7}
\]

where \( \phi = (1 - 2\tau/\tau (1 - \tau)) \) and \( \eta^2 = 2\tau (1 - \tau) \).

From these results, the dependent variable \( y_i \) can be rewritten equivalently as follows:

\[
y_i = x_i \beta + \sigma (\phi z_i + \eta \sqrt{z_i} u_i). \tag{8}
\]

Expanding (8) and thereafter reparametrisation, we obtain the following equation:

\[
y_i = x_i \beta + \phi v_i + \eta \sqrt{v_i} u_i, \tag{9}
\]

where \( v_i = \sigma z_i \). The exponential normal mixture distribution of \( \epsilon_i \) shows that \( y_i \) s are normally distributed with mean, \( x_i \beta + \phi v_i \), and variance, \( \eta^2 v_i \) [10, 20]. Therefore, \( y_i \) has a normal density function:

\[
\frac{1}{\eta \sqrt{v_i}} \exp\left[\frac{(y_i - x_i \beta - \phi v_i)^2}{2\eta^2 v_i}\right]. \tag{10}
\]

and the resulting likelihood function is given by

\[
L(y_i) = \left(\prod_{i=1}^{n} \frac{1}{\eta \sqrt{v_i}}\right) \exp\left[-\sum_{i=1}^{n} \frac{(y_i - x_i \beta - \phi v_i)^2}{2\eta^2 v_i}\right]. \tag{11}
\]

The aim then is to estimate the regression coefficients, \( \beta \), scale parameter, \( \sigma \), and the mixture variable, \( v_i \), in (9).

Two sets of prior probability density distributions are selected for the parameters: \( \beta \sim N(U_0, V_0) \), where \( N(.) \) denotes a normal distribution, and \( \beta \sim \exp(U_0, V_0) \), where \( \exp(.) \) denotes a double exponential distribution. Also, we assume that \( v_i \sim \exp(1/\sigma) \), where \( \exp(.) \) denotes an exponential distribution, and \( \sigma \sim IG(n_0/2, s_0/2) \), where \( IG(.) \) denotes an inverse gamma distribution.

In the case of the first Bayesian model (i.e., with \( \beta \sim N(U_0, V_0) \) and \( \beta \sim \exp(U_0, V_0) \)), named Bayesian model 1, we assign

\[
\begin{align*}
&\beta \sim N(U_0, V_0), \\
&v_i \sim \exp\left(\frac{1}{\sigma}\right), \tag{12}
\end{align*}
\]

\[
\sigma \sim IG\left(\frac{n_0}{2}, \frac{s_0}{2}\right).
\]

Likewise, for the second Bayesian model (i.e., with \( v_i \sim \exp(1/\sigma) \) and \( \sigma \sim IG(n_0/2, s_0/2) \)), named Bayesian model 2, we assign

\[
\begin{align*}
&\beta \sim \exp(U_0, V_0), \\
&v_i \sim \exp\left(\frac{1}{\sigma}\right), \tag{13}
\end{align*}
\]

\[
\sigma \sim IG\left(\frac{n_0}{2}, \frac{s_0}{2}\right).
\]

Given the likelihood distribution in (11) and by specifying the prior probabilities of the parameters of interest, the posterior distributions can be derived. Thus, for Bayesian model 1, the posterior distribution is given as

\[
g(\beta, \nu, \sigma \mid y_i) \propto N(x_i \beta + \phi v_i, \eta^2 v_i) N(U_0, V_0) \exp\left(\frac{1}{\sigma}\right) IG\left(\frac{n_0}{2}, \frac{s_0}{2}\right)
\]

\[
\propto \left(\prod_{i=1}^{n} \frac{1}{\eta \sqrt{v_i}}\right) \exp\left[-\sum_{i=1}^{n} \frac{(y_i - x_i \beta - \phi v_i)^2}{2\eta^2 v_i}\right] \times \frac{1}{\sqrt{2\pi V_0}} \exp\left[-\frac{(\beta - U_0)^2}{2V_0}\right] \frac{1}{\sigma} \exp\left(-\frac{v_i}{\sigma}\right) \left(\frac{1}{\sigma}\right)^{(n_0/2)+1} \exp\left(-\frac{s_0}{2\sigma}\right). \tag{14}
\]

Also, for Bayesian model 2, the posterior distribution is given as
\[ g(\beta, \nu, \sigma | y_i) \propto N(x_i^{\prime} \beta + \nu \eta^2 \sigma v_i) \exp(U_0, V_0) \exp\left(\frac{1}{\sigma}\right) \int G\left(\frac{n_0}{2}, \frac{s_0}{2}\right) \]
\[ \propto \left(\prod_{i=1}^{n} \frac{1}{\eta \sqrt{\sigma} v_i}\right) \exp\left[-\sum_{i=1}^{n} \left(\frac{(y_i - x_i^{\prime} \beta - \nu \eta^2 \sigma v_i)^2}{2\eta^2 \sigma v_i}\right)\right] \times \frac{1}{2V_0} \exp\left[-\frac{(\beta - U_0)^2}{2V_0}\right] \times \frac{1}{\sigma} \exp\left(-\frac{\nu^2}{\sigma}\right) \left(\frac{1}{\sigma}\right)^{(n/2)+1} \exp\left(-\frac{s_0}{2\sigma}\right) \] (15)

It can be shown that the marginal conditional density for \( \beta \) in Bayesian model 1 is normally distributed with mean and variance given, respectively, as
\[ \mu = \frac{\sum_{i=1}^{n} \left( x_i (y_i - \nu \eta^2 \sigma v_i) + \frac{U_0}{V_0} \right)}{\left( \sum_{i=1}^{n} \left( \frac{x_i^2}{\eta^2 \sigma v_i} + \frac{1}{V_0} \right) \right)} \quad (16) \]
and
\[ \xi^2 = \left[ \sum_{i=1}^{n} \left( \frac{x_i^2}{\eta^2 \sigma v_i} + \frac{1}{V_0} \right) \right]^{-1} \]

Also, the marginal conditional distribution of \( \beta \) for Bayesian model 2 is normally distributed with mean and variance given as
\[ \mu = \frac{\sum_{i=1}^{n} \left( x_i \phi v_i \right) - \frac{y_i x_i}{\eta^2 \sigma v_i} \frac{U_0}{V_0} \times \sum_{i=1}^{n} \left( \frac{x_i^2}{\eta^2 \sigma v_i} + \frac{1}{V_0} \right) \} \]
and
\[ \eta^2 \sigma v_i \]

\[ g(\sigma | \beta, \nu, y, x) \propto \left(\frac{1}{\sigma}\right)^{(\text{3}n + n_0)/2} \exp\left(-\frac{1}{\sigma} \left[ \sum_{i=1}^{n} \left(\frac{y_i - x_i^{\prime} \beta}{\sigma} - \frac{\nu}{\sigma} \right)^2 - \frac{s_0}{2} \right] \right) \]
\[ \propto \left(\frac{1}{\sigma}\right)^{(3n + n_0)/2} \exp\left(-\frac{1}{\sigma} \left[ \frac{s_0}{2} + \sum_{i=1}^{n} \frac{y_i - x_i^{\prime} \beta}{\eta^2 \sigma v_i} \right] \right) \] (22)

As a result, the marginal distribution of \( \sigma \) follows an inverse gamma with the following parameters:
\[ \frac{3n + n_0}{2} \] (23)
and
\[ \frac{s_0}{2} + \frac{n}{2} \sum_{i=1}^{n} \frac{(y_i - x_i^{\prime} \beta - \nu \eta^2 \sigma v_i)^2}{2\eta^2 \sigma v_i} \] (24)

2.2.1. Estimating the Mixture Component. A mixture distribution for a fixed number of components can be specified as \( \sum_{i=1}^{n} y_i\, \gamma_i \), where \( y_i, \mu_i, \) and \( \sigma_i \) are the proportions of the component distributions, their means, and standard deviations, respectively. To estimate the parameters \( y_i, \mu_i, \) and \( \sigma_i \) associated with \( \nu_i \), the first assumption is that the marginal conditional distribution of \( \nu_i \) is a finite mixture distribution of two normal components. The second assumption is that the latent variable, \( \lambda \), has a value of 0 or 1 associated with absent and present event rates, respectively.

In this context, \( y_i \) and \( \nu_i \) are the proportions of present and absent event rates, i.e., \( p(\lambda = 1) \) and \( p(\lambda = 0) \). The estimation of the mixture variable is done in R using the Bayesian mixture package. The package provides the Gibbs sampling of the posterior distribution, a method to set up the
model, and specifies the priors and initial values required for the Gibbs sampler.

2.2.2. Computing the Probability Score. As the aim is to compute the probability \( P(Y | \lambda = 1) \) for each observation, it can also be noticed from (2.2.5) that \( \lambda_i, i = 0, 1 \) is only dependent on \( Y \) through the estimated values of \( v_i \) and \( \sigma \). Therefore, from the Bayes theorem, the probability that an observation belongs to the present event rate is computed as

\[
p(Y | \lambda = 1) = \frac{p(\lambda = 1 | Y, \sigma, v_i)p(Y, \sigma, v_i)}{\sum_{i=0}^{1} p(\lambda_i | Y, \sigma, v_i)p(Y, \sigma, v_i)}. \tag{25}
\]

2.3. Incorporating the Error Cost Ratios. In this section, we outline our steps in modifying the univariate discriminant function to incorporate the error cost ratios.

Let \( g(y_1) \) and \( g(y_2) \) be the density functions associated with a \( p \times 1 \) random vector variable \( Y \) for \( \pi_1 \) and \( \pi_2 \). An object with related measurement \( y \) must be allocated to either \( \pi_1 \) or \( \pi_2 \). Also, let \( \Omega \) be the sample space of \( Y \) and \( A_1 \) be the values of \( Y \) for which objects are classified as \( \pi_1 \) and \( A_2 = \Omega - A_1 \) as the remaining \( y \) values for which objects are classified as \( \pi_2 \) because they are mutually exclusive and exhaust the sample space.

The probability \( P(2|1) \) of classifying an object as \( \pi_2 \) when it is derived from \( \pi_1 \) is given as

\[
P(2|1) = P(Y \in A_2 | \pi_1) = \int_{A_2 = \Omega - A_1} g_1(y)dy. \tag{26}
\]

Similarly, the probability \( P(1|2) \) of classifying an object as \( \pi_1 \) when it is derived from \( \pi_2 \) is given as

\[
P(1|2) = P(Y \in A_1 | \pi_2) = \int_{A_1} g_2(y)dy. \tag{27}
\]

In addition, let \( p_1 \) and \( p_2 \) be the prior probabilities of an object belonging to \( \Pi_1 \) and \( \Pi_2 \), where \( p_1 + p_2 = 1 \). Then, the total probability of accurately or inaccurately characterizing objects can be deduced as the product of the prior and conditional classification probabilities. For example,

\[
P(\text{objects accurately classified as } \pi_1) = P(Y \in A_1 | \pi_1)p(\pi_1) = P(1|1)p_1. \tag{28}
\]

Classification systems are commonly assessed based on their misclassification probabilities. However, this overlooks the error cost of misclassification. The error cost of misclassification (ECM) can be characterized by a cost matrix given in Table 1.

Thus, we assign a cost of

(i) Zero for accurate classification
(ii) \( c(1|2) \) when an object from \( \Pi_2 \) is inaccurately classified as \( \pi_1 \)
(iii) \( c(2|1) \) when an object from \( \Pi_1 \) is inaccurately classified as \( \pi_2 \)

For any rule, when the off-diagonal entries of the cost matrix are multiplied by their respective probabilities of occurrence, we obtain the expected error cost of misclassification (ECM) as

\[
\text{EECM} = c(2|1)P(2|1)p_1 + c(1|2)P(1|2)p_2. \tag{29}
\]

The areas, \( A_1 \) and \( A_2 \), that reduce the EECM have defined \( y \) values for which the following holds:

\[
\begin{align*}
\frac{(1/\sqrt{2\pi\sigma_1}) \exp\left(-\frac{1}{2\sigma_1^2}(Y - \mu_1)^2\right)}{(1/\sqrt{2\pi\sigma_2}) \exp\left(-\frac{1}{2\sigma_2^2}(Y - \mu_2)^2\right)} & \geq \frac{c(1|2)}{c(2|1)} \times \frac{p_2}{p_1}, \tag{30}
\intertext{and}
\frac{1}{2\sigma_1^2} \left[ (\sigma_1^2 - \sigma_2^2)Y^2 + 2(\sigma_1^2\mu_1 - \sigma_2^2\mu_2)Y + (\sigma_1^2\mu_1 - \sigma_2^2\mu_2)^2 \right] & \geq \ln \left( \frac{\sigma_1^2}{\sigma_2^2} \right)^{(1/2)} \frac{p(1|2)}{p(2|1)} \times \frac{p_2}{p_1}. \tag{31}
\end{align*}
\]

However, if \( \sigma_1 = \sigma_2 = \sigma \), then from (32),
\[(\mu_1 - \mu_2)Y + \frac{1}{2}(\mu_2^2 - \mu_1^2) \geq \sigma^2 \ln \left[ \frac{c(1\mid 2)}{c(2\mid 1)} \times \frac{p_2}{p_1} \right]. \tag{33}\]

We denote the left-hand side of (32) and (33) as the quadratic and linear discriminant functions, say \(q\) and \(\ell\), with their respective right-hand sides as the critical values denoted as \(c_1\) and \(c_2\). The sample estimates for \(q\) and \(c_1\) are given, respectively, as

\[q = \frac{1}{2} \left[ (s_1^2 - s_2^2)Y^2 + 2(s_2Y_1 - s_1Y_2)Y + (s_2^2Y_1^2 - s_1^2Y_2^2) \right], \tag{34}\]

and

\[c_1 = s_1^2s_2^2 \ln \left[ \frac{c(1\mid 2)}{c(2\mid 1)} \times \frac{p_2}{p_1} \right]. \tag{35}\]

respectively. Thus, the ratio of the error cost of misclassification for (28) is obtained from

\[c(1 \mid 2): c(2 \mid 1) = a: k, \tag{36}\]

where \(a \in \mathbb{N}\) and

\[
\left[ \left( \frac{s_1}{s_2} \right)^{(1/2)} \times a \times \frac{p_2}{p_1} \right] e^{-\left( m_x/\sigma^2 \times \sigma^2 \right)} < k < \left[ \left( \frac{s_1}{s_2} \right)^{(1/2)} \times a \times \frac{p_2}{p_1} \right] e^{-\left( m_u/\sigma^2 \times \sigma^2 \right)}. \tag{37}\]

and

\[
\frac{a \times p_2}{p_1} e^{-\left( m_x/\sigma^2 \times \sigma^2 \right)} < k < \left[ \left( \frac{s_1}{s_2} \right)^{(1/2)} \times a \times \frac{p_2}{p_1} \right] e^{-\left( m_u/\sigma^2 \times \sigma^2 \right)}. \tag{40}\]

The metrics of performance evaluation computed from Table 2 includes sensitivity (the ability of the model to correctly classify present event rates as present), specificity (the ability of the model to correctly classify absent event rates as absent), and accuracy (the overall correct classification). Mathematically, the metrics of performance evaluation are computed as follows:

\[\text{Sensitivity} = \frac{TP}{TP + FN} \tag{43}\]

\[\text{Specificity} = \frac{TN}{TN + FP} \tag{44}\]

or

\[
\ell = (\bar{y}_1 - \bar{y}_2)Y + \left( \frac{1}{2}s_2^2 - \frac{1}{2}s_1^2 \right) \geq c_2, \sigma_1 = \sigma_2. \tag{42}\]

The univariate discriminant functions (34) and (35) are proposed for classifying an object into two distinct groups if \(\mu_1\) is statistically different from \(\mu_2\). In the next section, the performance criteria for evaluating these proposals are presented.

2.4. Performance Evaluation. To assess the classification methods, the confusion matrix is of interest. Table 2 shows the confusion matrix.
### 3. Simulation Study

In this section, we present a simulation study to assess the performance of the various models discussed in the previous section. It is organized into two sections, namely, simulation design and results and discussion.

#### 3.1. Simulation Design.

Three sample sizes, \( n = 100, n = 250, \) and \( n = 500 \), are considered to investigate the empirical consistency of the proposed methods. For each sample, our interest is in estimating the average concordance probability (multiplication of the model's test sensitivity and specificity, respectively), bias, and mean square error. The bootstrap sampling technique was used to estimate the standard errors of the estimators of the cut-off point from each model.

In addition, we selected a proportion of event occurrences as low as 0.05 to a high value of 0.5 in each generated sample to study the proposed models’ performance as the proportions vary. To select random samples having a predetermined proportion of event occurrence, we proposed a modified “conditional block bootstrap” in Minkah et al. [21] where the authors implemented it in selecting bootstrap samples for censored data. The conditional block bootstrap is a combination of ideas from the moving block bootstrap [22] and the bootstrap for the censored data [23].

In the proposed modified “conditional block bootstrap” procedure, the absent events are grouped into randomly chosen blocks. However, each block must contain at least one present event observation. The bootstrap observations are obtained by repeatedly sampling with replacement from these blocks and placing them together to form the bootstrap sample. Enough blocks must be sampled to obtain approximately the same sample size as the originally given sample. Given a sample of size, \( n \), and a proportion of present event occurrence, \( p \), the conditional block bootstrap procedure is as follows:

- **A1.** Group the \( n \) observations into two groups, namely, present and absent groups (with their corresponding covariates) with sample sizes \( n_p \) and \( n_a \), respectively. Thus, the proportion of the present event is \( p = n_p/n \).
- **A2.** Let \( n^B_i, i = 1, \ldots, m \ (n^B_i \geq 1) \) represent the number of present observations to be included in a block, \( i \). The block size, \( s \), is obtained as \((n \times n^B_i)/n_p \). If \( s \) is not an integer, let \( s = \lceil (n \times n^B_i)/n_p \rceil \).
- **A3.** The number of blocks, \( m \), is chosen such that \( n = m 	imes s \). In the case \( n = m 	imes s \), the blocks will have the same number of observations. However, if \( n \approx m \times s \), then \( m \) is taken as \( [n/s] \), in which case the first \( m - 1 \) blocks are allocated \( s \) observations each and the remaining \( n - s(m - 1) \) observations are allocated to the \( m \)th block.

- **A5.** Let \( b_j = 1, \ldots, m \) denote the \( j \)th block. Assign observations to each block by random sampling without replacement, \( s - n^B_i \) observations from the absent event group. In addition, randomly sample \( n^B_i \) observations without replacement from the present event group and assign them to each block \( b_j, j = 1, \ldots, m \). Thus, each block would contain \( n^B_i \) and \( s - n^B_i \) observations of present and absent events, respectively.

- **A6.** Sample \( m \) times with replacement from \( b_1, b_2, \ldots, b_m \) and place them together to form the bootstrap sample. These bootstrap samples will have sample sizes equal to or approximately, \( n \).

- **A7.** For the bootstrap samples obtained in A6, perform the analysis using Bayesian model 1 (BM1), Bayesian model 2 (BM2), binary logistic regression using the proposed methodology for obtaining the optimal cut-off point (LM), and binary logistic regression with a cut-off point of 0.5 (LN).

- **A8.** For each model in A7, subsequently, obtain the optimal cut-off point and compute its corresponding \( cp \) denoted as \( c\hat{p}_1, c\hat{p}_2, c\hat{p}_3, \) and \( c\hat{p}_4 \), respectively.

- **A9.** Repeat A1 to A8 a large number of times, \( R \) (\( R \geq 800 \)) (see, e.g., [24] for justification) to obtain the \( cp \) values \( c\hat{p}_{i1}, c\hat{p}_{i2}, \ldots, c\hat{p}_{i4} \) for \( i = 1, 2, 3, 4 \).

- **A10.** Compute the average \( cp \), bias, and MSE for the \( i \)th method in A9, i.e.,

\[
\bar{c}\hat{p}_i = \frac{1}{R} \sum_{j=1}^{R} c\hat{p}_{ij},
\]

\[
\text{bias}(c\hat{p}_i, c.p) = \left( \frac{1}{R} \sum_{j=1}^{R} c\hat{p}_{ij} \right) - c.p,
\]

and

\[
\text{MSE}(c\hat{p}_i, c.p) = \frac{1}{R} \sum_{j=1}^{R} (c\hat{p}_{ij} - \bar{c}\hat{p}_i)^2.
\]

#### 3.2. Results and Discussion.

The results of the simulation study are presented in this section. For brevity and ease of presentation, we display the plots of the average \( cp \), bias, and mean square error (MSE) as a function of the proportion of event occurrences. The criteria for an appropriate model is to have high \( cp \) values (closer to 1) and low values of bias and MSE.

Figure 1 shows the graph of average \( cp \), bias, and MSE as a function of the proportion of event occurrences for the various models and the varying sample sizes.
Clearly, the logistic-based models (LM and LN) have high $cp$ values than their counterparts from the Bayesian frameworks (BM1 and BM2). Also, our proposed logistic regression-based classifier that incorporates the error cost of misclassification, LM, provides a better performance measure than LN as the proportion of event occurrences increases. Also, this observation becomes more apparent as the sample size increases. In the case of the Bayesian framework, BM2 has better $cp$ values, save for smaller proportions of event occurrences. Therefore, in general, our proposed LM model can be considered the most appropriate model with high $cp$ values for classification purposes.

In terms of bias, the results are mixed, but BM1 shows lesser bias in most cases across the sample sizes and proportion of event occurrences. In the case of the MSE, the logistic-based models provide smaller values compared with the Bayesian models, especially for a smaller proportion of event occurrences. In addition, there is a gradual decrease in the MSEs as the sample increases. This is desirable as it indicates the empirical consistency of the estimators of the $cp$ values in each model. In conclusion, the proposed LM model is universally competitive in generating higher $cp$ values regardless of the sample size and the proportion of event occurrences in a data set.
4. Practical Application

This section illustrates the proposed method for estimating the optimal cut-off point on a home equity loan data set. The data comprise 1000 customers in the United States of America. The dependent variable, $Y$, is the loan amount (in thousands of dollars), while the independent variables are the mortgage (amount due on the existing mortgage in thousands of dollars), the value of the current property, the present position, and debt-to-income ratio. In addition, the variable $Bad$ represents the status of the loan repayment. Table 3 shows the structure of the home equity loan dataset. Our interest is in the estimation of the cut-off point for the classification of loan repayment using the methods discussed in the previous sections.

4.1. Estimating Cut-Off Point Using Bayesian Quantile Regression. The quantile regression equation for the home equity loan data is

$$ Y = \beta_1 + \beta_2 \cdot \text{mortgage} + \beta_3 \cdot \text{value} + \beta_4 \cdot \text{reason} + \beta_5 \cdot \text{job} + \beta_6 \cdot \text{years} + \beta_7 \cdot \text{debtinc} + \phi v_i. \quad (49) $$

Estimates of the model’s parameters at $\tau = 0.75$ for BM1 and BM2 are shown in Tables 4 and 5, respectively.

Here, the aim is to identify bad home equity loan through the estimated values of the latent variable $v_i$. The components of the mixture variable, $v_i$, estimated for the home loan equity data are shown in Table 6.

The marginal conditional distribution of $v_i$ is a finite mixture of two normal components. The component with the larger mean is associated with the distribution of bad home equity loans, while the component with the smaller mean is associated with the distribution of good home equity loans. The averages for the bad home equity loan rates are estimated as $\mu_1 = 3.724$ and $\mu_2 = 3.711$ (with corresponding proportions $\gamma_2 = 0.01447$ and $\gamma_2 = 0.01522$), respectively, for BM1 and BM2.

We now compute the probability of each observation belonging to the distribution of bad and good home loan equity rates using (25). Tables 7 and 8 show some data points and their respective computed probabilities.

Furthermore, to ascertain which univariate discriminant function will be most suitable for classification, Levene’s test for equality of variance of the two distributions of present and absent event rates is conducted on the home equity data set. The test results show that there is a significant difference in the variances of the two distributions of bad and good loan repayment events for BM1 ($F = 29.0806, p \text{ value} = 0.010$) and BM2 ($F = 26.6754, p \text{ value} = 0.0001$). Therefore, this implies that a quadratic discriminant function is most appropriate for the classification of this data set.

Moreover, the independent-samples $t$-tests for equal means for the two distributions of present and absent event rates are significant ($t = 33.9785, p \text{ value} < 0.01$) for BM1 and ($t = 34.2974, p \text{ value} < 0.01$) for BM2. Now, using (35) and systematically shifting the $k$ within the bounds (30), we obtain the optimal cut-off points, 0.4902 and 0.4964, at $k = 0.03$ and $k = 0.0005$, for BM1 and BM2, respectively. At these points, the highest concordance probabilities are achieved.

4.2. Estimating Cut-Off Point Using Binary Logistic Regression. The binary logistic regression equation for the home equity loan data is given as follows:

$$ \ln \left( \frac{p}{1 - p} \right) = \beta_0 + \beta_1 \cdot \text{loan} + \beta_2 \cdot \text{mortdue} + \beta_3 \cdot \text{value} + \beta_4 \cdot r2 + \beta_5 \cdot j2 + \beta_6 \cdot j3 + \beta_6 \cdot j3 + \beta_7 \cdot j4 + \beta_8 \cdot j5 + \beta_9 \cdot j6 + \beta_{10} \cdot \text{years} + \beta_{11} \cdot \text{debtinc}. \quad (50) $$

Table 9 presents the parameter estimates obtained through the maximum-likelihood principle for the data set. Also, Table 10 presents the predicted probabilities of bad and good home equity loan rates for the data set.

In addition, Levene’s test for equality of variance between the distribution of good and bad home equity loans shows there are no significant differences ($p \text{ value} = 0.5$). Hence, a linear discriminant function is the most applicable for classification. The sample pooled variances for the two groups of loan statuses are estimated as 0.101581. In addition, the independent-samples $t$-tests for equality of means for a bad and good home equity loan are rejected, with the $p$ values being less than 0.01.

It is similar to the Bayesian quantile regression in the preceding section, but with (39) and (40). We obtain the optimal cut-off point for the logistic regression model as $c_2 = 0.3345$, at $k = 1.7332$. 
Table 3: Home equity loan data.

<table>
<thead>
<tr>
<th>Loan amount</th>
<th>Mortgage</th>
<th>Value</th>
<th>Reason</th>
<th>Job</th>
<th>Years</th>
<th>Debt-to-income ratio</th>
<th>Bad</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.7</td>
<td>30.548</td>
<td>40.32</td>
<td>2</td>
<td>3</td>
<td>9</td>
<td>37.11361</td>
<td>1</td>
</tr>
<tr>
<td>1.8</td>
<td>28.502</td>
<td>43.034</td>
<td>2</td>
<td>3</td>
<td>11</td>
<td>36.88489</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>4.4</td>
<td>53.091</td>
<td>64.074</td>
<td>1</td>
<td>2</td>
<td>14</td>
<td>39.15808</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4: Estimate of the model's parameter for the complete home equity loan data using Bayesian Model 1.

<table>
<thead>
<tr>
<th>Mean</th>
<th>SD</th>
<th>Naive SE</th>
<th>Time-series SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>1.0722</td>
<td>0.0951</td>
<td>1.504e-3</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.0013</td>
<td>0.0011</td>
<td>1.871e-5</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-0.0014</td>
<td>0.0009</td>
<td>1.487e-5</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.0476</td>
<td>0.0331</td>
<td>5.236e-4</td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>-0.0006</td>
<td>0.0156</td>
<td>2.460e-4</td>
</tr>
<tr>
<td>$\beta_6$</td>
<td>-0.0005</td>
<td>0.0022</td>
<td>3.552e-5</td>
</tr>
<tr>
<td>$\beta_7$</td>
<td>-0.0013</td>
<td>0.0016</td>
<td>2.483e-5</td>
</tr>
<tr>
<td>$v_1$</td>
<td>2.7346</td>
<td>1.2414</td>
<td>1.963e-2</td>
</tr>
<tr>
<td>$v_2$</td>
<td>2.8297</td>
<td>1.2576</td>
<td>1.988e-2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>3.2491</td>
<td>0.1182</td>
<td>1.870e-3</td>
</tr>
</tbody>
</table>

Table 5: Estimate of the model's parameter for the complete home equity loan data using Bayesian Model 2.

<table>
<thead>
<tr>
<th>Mean</th>
<th>SD</th>
<th>Naive SE</th>
<th>Time-series SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>1.0662</td>
<td>0.0892</td>
<td>1.411e-3</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.0011</td>
<td>0.0010</td>
<td>1.557e-5</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-0.0012</td>
<td>0.0008</td>
<td>1.251e-5</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.0545</td>
<td>0.0340</td>
<td>5.382e-4</td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>0.0006</td>
<td>0.0160</td>
<td>2.525e-4</td>
</tr>
<tr>
<td>$\beta_6$</td>
<td>-0.0007</td>
<td>0.0024</td>
<td>3.871e-5</td>
</tr>
<tr>
<td>$\beta_7$</td>
<td>-0.0015</td>
<td>0.0016</td>
<td>2.471e-5</td>
</tr>
<tr>
<td>$v_1$</td>
<td>2.7025</td>
<td>1.2289</td>
<td>1.943e-2</td>
</tr>
<tr>
<td>$v_2$</td>
<td>2.7895</td>
<td>1.2292</td>
<td>1.944e-2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>3.2693</td>
<td>0.1141</td>
<td>1.804e-3</td>
</tr>
</tbody>
</table>

Table 6: Estimates of the mixture components for the two Bayesian models.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>BM1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>3.0230</td>
<td>0.0593</td>
<td>(2.9200, 3.1480)</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>3.7240</td>
<td>0.0051</td>
<td>(3.7150, 3.7330)</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.9855</td>
<td>0.0049</td>
<td>(0.9726, 0.9927)</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.0145</td>
<td>0.0049</td>
<td>(0.0072, 0.0274)</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.0274</td>
<td>0.0012</td>
<td>(0.0256, 0.0297)</td>
</tr>
<tr>
<td>BM2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>3.0410</td>
<td>0.0670</td>
<td>(2.9030, 3.1600)</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>3.7110</td>
<td>0.0050</td>
<td>(3.7030, 3.7210)</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.9848</td>
<td>0.0053</td>
<td>(0.9735, 0.9915)</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.0152</td>
<td>0.0053</td>
<td>(0.0085, 0.0265)</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.0269</td>
<td>0.0013</td>
<td>(0.0246, 0.0294)</td>
</tr>
</tbody>
</table>
Table 7: Estimated probabilities for Bayesian model 1.

| \(v_i\) | Loan amount | \(P(Y, \lambda = 0, v_i)\) | \(P(Y, \lambda = 1, v_i)\) | \(P(Y | \lambda = 1)\) | Bad |
|--------|-------------|----------------|----------------|----------------|-----|
| \(v_1\) | 1.7         | 0.0000         | 0.0000         | 0.0000         | 1   |
| \(v_2\) | 1.8         | 0.0000         | 0.0000         | 0.0000         | 1   |
| \(\cdots\) | \(\cdots\) | \(\cdots\) | \(\cdots\) | \(\cdots\) | \(\cdots\) |
| \(v_{1000}\) | 11.7        | 1.0000         | 1.0000         | 0.5000         | 1   |

Table 8: Estimates of the probabilities for Bayesian model 2.

| \(v_i\) | Loan amount | \(p(Y, \lambda = 0, v_i)\) | \(p(Y, \lambda = 1, v_i)\) | \(p(Y | \lambda = 1)\) | Bad |
|--------|-------------|----------------|----------------|----------------|-----|
| \(v_1\) | 1.7         | 0.0000         | 0.0000         | 0.0000         | 1   |
| \(v_2\) | 1.8         | 0.0000         | 0.0000         | 0.0000         | 1   |
| \(\cdots\) | \(\cdots\) | \(\cdots\) | \(\cdots\) | \(\cdots\) | \(\cdots\) |
| \(v_{1000}\) | 11.7        | 1.0000         | 0.9999         | 0.4999         | 1   |

Table 9: Parameter estimates for the home equity loan data.

| Variable   | Estimate | Std. error | \(z\) value | \(\Pr(>|z|)\) |
|------------|----------|------------|-------------|---------------|
| Intercept  | 6.9358   | 0.7013     | 9.8890      | <2e−16        |
| Loan       | −0.4115  | 0.0349     | −11.7860    | <2e−16        |
| Mortdue    | 0.0238   | 0.0084     | 2.8110      | 0.0049        |
| Value      | −0.0072  | 0.0069     | −1.0440     | 0.2966        |
| Reason 2   | −0.0441  | 0.2084     | −0.2120     | 0.8323        |
| Job 2      | 0.8219   | 0.3824     | 2.1490      | 0.0316        |
| Job 3      | 0.2916   | 0.3136     | 0.9300      | 0.3525        |
| Job 4      | 0.5350   | 0.3734     | 1.4330      | 0.1520        |
| Job 5      | −1.0364  | 0.5560     | −1.8510     | 0.0642        |
| Job 6      | −0.8374  | 0.7247     | −1.1550     | 0.2479        |
| Years      | 0.0510   | 0.0157     | 3.2400      | 0.0012        |
| Debtinc    | −0.0866  | 0.01466    | −5.9350     | 2.93e−09      |

Table 10: Predicted probabilities of good and bad for the home equity loan data.

| Loan amount | \(P(Y = 0 | X = x)\) | \(P(Y = 1 | X = x)\) | Bad |
|-------------|----------------|----------------|-----|
| 1.7         | 0.9847         | 0.0153         | 1   |
| 1.8         | 0.9849         | 0.0151         | 1   |
| 2.4         | 0.9756         | 0.0244         | 0   |
| \(\cdots\) | \(\cdots\)     | \(\cdots\)     | \(\cdots\) |
| 11.3        | 0.6448         | 0.3552         | 0   |

Table 11: Performance metric evaluation for the home equity loan data.

<table>
<thead>
<tr>
<th>Measures</th>
<th>Bayesian model 1 (%)</th>
<th>Bayesian model 2 (%)</th>
<th>Logistic regression (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>68.00</td>
<td>60.67</td>
<td>73.33</td>
</tr>
<tr>
<td>Specificity</td>
<td>68.14</td>
<td>78.29</td>
<td>94.00</td>
</tr>
<tr>
<td>Accuracy</td>
<td>64.60</td>
<td>73.00</td>
<td>87.80</td>
</tr>
<tr>
<td>Concordance probability</td>
<td>0.4294</td>
<td>0.4750</td>
<td>0.6893</td>
</tr>
</tbody>
</table>
4.3. Performance Metrics Evaluation. This section presents the performance metrics (specificity, sensitivity, accuracy, and concordance probability) used to assess the various models’ performances on the home equity loan data set. From the results in Sections 4.1 and 4.2, the values of $c_1$ and $c_2$ are used to obtain the performance metrics, and the results are shown in Table 11.

Clearly, the logistic regression model incorporating the proposed methodology produces the highest test sensitivity, specificity, accuracy, and concordance probability values. Also, of the two Bayesian models, Bayesian model 2 has greater test specificity, accuracy, and concordance probability values than Bayesian model 1. However, Bayesian model 1 produces a higher test sensitivity value than Bayesian model 2. Thus, it can be concluded that using logistic regression with the proposed incorporation of the error cost of misclassification produces better performance metrics in classifying loans for home equity.

5. Conclusion

This paper introduced an approach for estimating the optimal cut-off point for classification. The proposed method modifies univariate discriminant functions by incorporating the error cost ratio for classification. Thus, the misclassification cost ratios can be systematically adjusted within some specified measurement range. A corresponding cut-off value is subsequently obtained for each unique cost ratio, and other metrics of performance measures can be computed. Three methods of computing the cut-off point were proposed: a logistic and two Bayesian quantile regressions. A simulation study was conducted to assess the performance of these models in estimating the concordance probability and thus the cut-off point. The results show that incorporating the error cost of misclassification improves the concordance probability and provides smaller values for bias and mean square errors. In particular, the logistic regression with the proposed incorporation of the error cost of misclassification provides the best method as it gives concordance probability values close to 1 and smaller values of bias and mean square error. The proposed method is illustrated using loan data from the finance industry.

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.

Acknowledgments

Ampomah O.-A was supported with funds from the Carnegie Corporation-University of Ghana BANGA-Africa Project. This research work has been supported by funds from the Carnegie Banga-Africa Project, University of Ghana.

References


