• FINANSE I PRAWO FINANSOWE •

• JOURNAL OF FINANCE AND FINANCIAL LAW •

czerwiec / june 2026 • vol. 2(50): 49–66
https://doi.org/10.18778/2391-6478.2.50.03

COPE Member since 2021 JM15102

Agnieszka Rossa https://orcid.org/0000-0002-0444-4181
dr hab. Institute of Statistics and Demography, University of Lodz, e-mail: agnieszka.rossa@uni.lodz.pl

OPTIMAL THRESHOLDING FOR BINARY CLASSIFICATION APPLIED IN CREDIT SCORING

ABSTRACT

The paper concerns a new method of classifying individuals into two subpopulations and demonstrates the application of this method in credit scoring. Individuals are classified into two subpopulations depending on the duration T of a certain phenomenon (e.g., default). The duration may be shorter or longer than a certain fixed value t. It is assumed that the variable T is not known at the time of classification, so the explanatory continuous predictive marker is used instead. The optimal acceptance threshold for a predictive marker is determined by a time-dependent receiver operating curve (ROC) estimated from a random sample. A typical complexity of time-to-event data is that observations in the sample can be right-censored. Therefore, the estimation is based on a sequential random sampling and the Kaplan–Meier estimator.

Keywords: binary classification, sequential random sampling, sensitivity and specificity, time-dependent ROC curves

MSC Classification: 62L10, 62L12, 62N02, 62P05

Received: 05.12.2025 Accepted: 15.05.2026 Available online: 30.06.2026

Funding information: Not applicable. **Conflicts of interests:** None. **Ethical considerations:** The Author assure of no violations of publication ethics and take full responsibility for the content of the publication.

Introduction

Economic, demographic, social or biometric research frequently faces the problem of classifying objects into one of two subpopulations. For example, in credit scoring, such subpopulations may consist of borrowers who default on their loans or not. In clinical trials, the groups may comprise individuals who test positive or negative for a given condition.

The first concept of the classification method was developed by Ronald Fisher (1936), with the aim of finding a certain linear function of the measurements by which the populations are best distinguished. The discrimination function was defined by Fisher as a linear combination of physical measurements that separates two species of irises, i.e., Iris setosa and Iris versicolor. Between 1936 and 1940 Fisher published four articles on statistical discriminant analysis that became a common method used in statistics.

David Durand recognized in 1941 that a similar method can be employed to distinguish between good and bad loans (Thomas, 2000). In the 1980s, good credit card scoring results convinced banks to use statistically derived models in lending decisions. In the 1990s, the rise of direct marketing led to the use of scorecards to improve the response rate to advertising campaigns. Today, classification methods are used in many business decision-making practices, including credit scoring, bankruptcy prediction, fraud detection, or medical diagnostics, among others. A wide range of methods and their improvements have also been proposed. Such methods include, among others, linear, quadratic, and logistic discriminant analysis.

In the classic medical setting, the notions 'sensitivity' and 'specificity' were defined to distinguish between two groups of individuals, i.e., with positive and negative outcomes, based on a binary predictive marker. They are mainly used to measure validity of diagnostic tests (Godyn et al., 1991; Tetrault, 1991; Altman & Bland, 1994; Dickie, 1994). The first one (sensitivity) means the probability of correct classifying an individual with a positive outcome, while the second one (specificity) refers to the probability of correct classifying of a subject with a negative outcome. Both probabilities are estimated as respective empirical proportions. Sensitivity could be estimated by the observed true-positive fraction (TPF) and specificity – by the true-negative fraction (TNF).

When the predictive marker Y is continuous, the threshold criteria $Y \geq c$, $Y < c$ are used to define sensitivity and specificity. When multiple predictors are involved, the Y variable is usually represented by the probability of a positive outcome, either using logistic regression conditioned by the values of the explanatory variables, or using any increasing function of this probability. In medical research, the predictive marker may also be established from a linear combination of biomarkers and other medical measurements (Zhang & Yu, 1998; McNutt et al., 2003; Diaz-Quijano, 2012; Patil & Durairaj, 2015). In credit scoring, Y can be developed from various financial or economic characteristics of the loan applicants (Crook et al., 2007; Hu & Ansell, 2007; Thomas, 2009). In predicting corporate failure, a financial, micro- or macroeconomic measures are employed (Altman, 1968; Ohlson, 1980; Kovacova & Kliestik, 2017; Kovacova et al., 2019) etc.

The correct predictive threshold c of the predictive marker Y plays an important role in the classification problem, since the number of correctly classified units depends on the selected threshold value. There are two types of misclassification errors in any classification method: a type-I error, i.e., the probability of misclassifying a bad unit as a good one, and a type-II error, the probability of misclassifying a good unit as a bad one. Most comparative studies utilize these two types of error rates as accuracy measures of classifiers (Jamain & Hand, 2008). For the continuous predictive marker, it is possible to calculate the Gini index or certain measures of information gain for various cutoff points (Breiman et al., 1984) or to apply a more comprehensive approach based on the Receiver Operating Characteristic curve (ROC) (Green & Swets, 1966; Lloyd, 1998). The ROC curve displays the sensitivity versus 1-specificity for all the possible cutoff points c. It allows both to find the optimal threshold of c and to evaluate basic classification summaries (Baker, 2003).

In many studies, the primary goal is to predict the onset of some phenomenon of interest (e.g., the onset of a default, failure, disease, remission, etc.) in a certain period. For instance, a patient may not be considered as cancer-free when the remission time is less than five years, so cancer-free patients can be recognized when their remission time will be at least five years. In such cases, dynamic sensitivity and specificity should be computed for different time points t. However, in practice, the time to event of interest, T, is often right-censored. It means, that for some individuals, it is not possible to know whether $T < t$ or $T \geq t$ for some fixed points t.

Various approaches have been proposed to deal with the time-dependent classification. Heagerty et al. (2000) used the Kaplan–Meier survival estimator (Kaplan & Meier, 1958) in defining sensitivity and specificity. To ensure the monotonicity of both measures, they considered the nearest neighbor estimator of the bivariate distribution of the random pair (T, Y). Chambless and Diao (2006) proposed a model for estimating sensitivity and specificity as well as the so-called AUC measure conditioning on the observed times-to-event, i.e., in a similar way as in the Kaplan–Meier estimator. Other estimators were derived directly from a fitted conditional survival function, given the decision variable Y. Uno et al. (2007) proposed an inverse probability of censoring weighting method, Pepe (1997, 1998, 2000) used different regression models and adopted the generalized linear model methods in the statistical inferences, Hung and Chiang (2010a, 2010b) used generalized linear models with time-varying coefficients to characterize the conditional distribution of the time-to-event variable and to derive the time-dependent AUC as a function of covariate values.

This study proposes a novel classification method for assigning individuals to two subpopulations based on a time-to-event outcome using a continuous predictive marker. The optimal threshold is determined via time-dependent ROC curves, accounting for right-censored data through sequential sampling and Kaplan–Meier estimation, conditional on the subsets $Y \geq c$ and $Y < c$. Sequential sampling treats observations incrementally, ensuring efficient use of available information and reducing bias due to censoring.

The main research hypothesis is that the proposed method provides an effective and reliable approach to determining the optimal classification threshold in the presence of right-censored time-to-event data.

The general concept of the binary classification

Suppose that there is a population, $\mathcal{G}$, consisting of disjoint subpopulations $\mathcal{G}_0$ and $\mathcal{G}_1$, such that $\mathcal{G} = \mathcal{G}_0 \cup \mathcal{G}_1$, $\mathcal{G}_0 \cap \mathcal{G}_1 = \emptyset$. For simplicity, individuals in the set $\mathcal{G}_1$ will be termed cases, while the others will be termed controls.

Assume that a unit randomly drawn from $\mathcal{G}$ has to be assigned to one or the other set based on a certain variable Y in such a way that when the observed value of Ysatisfies the inequality $Y \geq c$, then an individual goes to $\mathcal{G}_1$, otherwise it goes to $\mathcal{G}_0$. Therefore, Y and c will be termed the predictive marker and the predictive threshold, respectively.

To classify individuals as controls or cases, the following two phases can be defined. First, available information is integrated to generate a numerical value of the predictive marker, Y, e.g., a score. Then, the score is compared with the predictive threshold c.

Let Y_0 and Y_1 denote the continuous marker Y in the subpopulations $\mathcal{G}_0$ and $\mathcal{G}_1$, respectively, i.e.,

$$Y_0 = Y|\mathcal{G}_0,\ Y_1 = Y|\mathcal{G}_1.$$

Let H_0 and H_1 be the distribution functions of variables Y_0 and Y_1, respectively, and let H denote a distribution function of Y; hence

$$H_0(y) = \mathrm{P}(Y_0 < y) = \mathrm{P}(Y < y|\mathcal{G}_0),$$

$$H_1(y) = \mathrm{P}(Y_1 < y) = \mathrm{P}(Y < y|\mathcal{G}_1),$$

and

$$H(y) = \mathrm{P}(Y < y).$$

For a given threshold c, the probabilities

$$\bar{H}_1(c) = 1 - H_1(c) = \mathrm{P}(Y \geq c|\mathcal{G}_1)$$

and

$$H_0(c) = \mathrm{P}(Y < c|\mathcal{G}_0),$$

reflect sensitivity and specificity of the marker Y, respectively.

The main challenge is to choose an appropriate threshold value of c. If we choose the smallest value of c possible, all individuals will be assigned to cases, regardless of which sub-population they actually belong to. In this case, the probability of being classified into $\mathcal{G}_0$ and $\mathcal{G}_1$ will be 0 and 1, respectively. Conversely, if the value of c is the highest possible, all individuals will be classified to controls, so the probability of being classified into $\mathcal{G}_0$ and $\mathcal{G}_1$ will be 1 and 0, respectively. Therefore, c must be carefully selected to ensure that both probabilities of correct classification are relatively high. In selecting the optimal threshold value of c, the Receiver Operating Characteristic Curve (ROC) seems to be a very useful and comprehensive tool. This concept will be discussed in the next section.

ROC curves

The ROC curve analysis is a well-established method, especially in clinical trials, to investigate the discriminatory capacity of diagnostic tests or prognostic clinical markers. In the continuous settings, the ROC curve is a set of points in the (p_0, p_1) plane, where $p_0 = \bar{H}_0(c)$ and $p_1 = \bar{H}_1(c)$, and c runs over the set of real numbers, i.e.,

$$ROC = \{(\bar{H}_0(c), \bar{H}_1(c)),\ c \in \boldsymbol{R}\}.$$

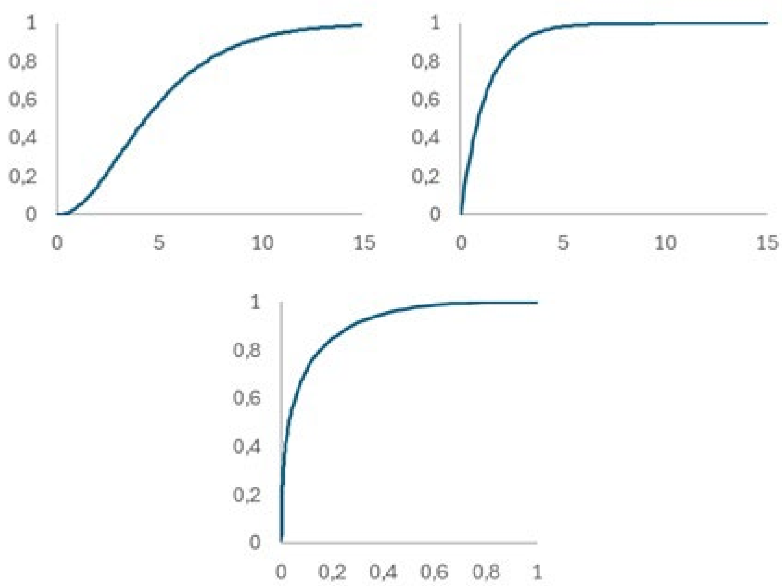
Another way to look at the ROC curve is to view it as a monotone increasing function $[0,1] \to [0,1]$ built using functions $\bar{H}_0$ and $\bar{H}_1$. The general formula for the ROC function is then as follows

$$ROC(c) = \bar{H}_1(\bar{H}_0^{-1}(c)),\ c \in \boldsymbol{R}.$$

Figure 1 illustrates an example of the ROC curve corresponding with some distribution functions.

Figure 1.

The ROC curve for two cumulative distribution functions H_0 (top left) and H_1 (top right)



Source: Original figure prepared by the author for demonstration.

It is an instrumental tool in finding a value of c that minimizes the risk of misclassification of units into controls or cases. The higher is the ROC curve in the unit square $[0,1] \times [0,1]$, the better is the capacity to distinguish both subpopulations $\mathcal{G}_0$ and $\mathcal{G}_1$ (see e.g., Heagerty et al., 2000). However, creating a ROC curve is not possible if the cumulative distribution functions $\bar{H}_0$ and $\bar{H}_1$ are not known. The problem is usually solved by estimating the curve from two random samples of controls and cases. Then, the estimated ROC curve takes the form

$$\widehat{ROC} = \left\{ \left(\hat{\bar{H}}_0(c), \hat{\bar{H}}_1(c) \right),\ c \in \boldsymbol{R} \right\},$$

where $\hat{\bar{H}}_0(c)$ and $\hat{\bar{H}}_1(c)$ are respective estimators of $\bar{H}_0(c)$, $\bar{H}_1(c)$, determined from random samples

$$Y_{01}, Y_{02}, \dots, Y_{0m},$$

and

$$Y_{11}, Y_{12}, \dots, Y_{1n},$$

where Y_{0i}, $i = 1,2, \dots, m$, Y_{1i}, $i = 1,2, \dots, n$ are independent copies of variables Y_0 and Y_1 sampled from subpopulations $\mathcal{G}_0$ and $\mathcal{G}_1$, respectively.

It is worth noting that $\hat{\hat{H}}_0(\cdot)$, $\hat{\hat{H}}_1(\cdot)$ are complement to the well-known empirical distribution functions $\hat{H}_0(\cdot), \hat{H}_1(\cdot)$, respectively, so they can be defined for any $c \in \boldsymbol{R}$ as the following sums

$$\hat{\hat{H}}_0(c) = \frac{1}{m}\sum_{i=1}^{m} \mathbf{1}\,(Y_{0i} \geq c),$$

$$\hat{\hat{H}}_1(c) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\,(Y_{1i} \geq c).$$

where $\mathbf{1}(A)$ is the characteristic function of an event A, taking values 1 if the event A occurs, and 0 otherwise.

In the classical approach, it is assumed that membership status of a unit in $\mathcal{G}_0$ or in $\mathcal{G}_1$ does not change over time. In many problems however, individuals' status changes over time when it is predicted whether a subject will or will not experience an event (e.g., default, death, remission).

Suppose that the time to event of interest is a continuous random variable T with the distribution function F. We will assume that the subpopulation $\mathcal{G}_0$ consists of units with $T \geq t$, and the subpopulation $\mathcal{G}_1$ is formed by units with $T < t$. As the T variable is usually not known at the time of classification, the criterion for classifying individuals into $\mathcal{G}_0$ or $\mathcal{G}_1$ will be a random continuous variable Y related to T. We need to find the optimal predictive threshold c for Y to determine the possibly optimal classification of subjects into both subpopulations depending on whether the time-to-event is at least t or not. Naturally, the value of c will fluctuate with t. Thus, a sequence of ROC curves for possible time points t can be considered rather than a single ROC curve.

The sensitivity and specificity of the marker Y should be now expressed by relating it to the probability distribution of T. To express the dynamic sensitivity and dynamic specificity, Heagerty et al., (2000) used the Bayes formula combining the probability distributions of T and Y, i.e.,

$$\bar{H}(c|T < t) = \mathrm{P}(Y \geq c|T < t) = \frac{\mathrm{P}(Y \geq c, T < t)}{\mathrm{P}(T < t)} = \frac{\mathrm{P}(T < t|Y \geq c)\mathrm{P}(Y \geq c)}{\mathrm{P}(T < t)},$$

$$H(c|T \geq t) = \mathrm{P}(Y < c|T \geq t) = \frac{\mathrm{P}(Y < c, T \geq t)}{\mathrm{P}(T \geq t)} = \frac{\mathrm{P}(T \geq t|Y < c)\mathrm{P}(Y < c)}{\mathrm{P}(T \geq t)}.$$

The expressions can also be written as:

$$\bar{H}(c|T<t)=\frac{F(t|Y\geq c)\bar{H}(c)}{F(t)},$$

$$H(c|T\geq t)=\frac{\bar{F}(t|Y<c)H(c)}{\bar{F}(t)}.$$

In the further part, we will consider two equivalent expressions resulting from the Total Probability Rule applied in the denominators of both formulas. The transformed formulas take the form:

$$\bar{H}(c|T<t)=\frac{F(t|Y\geq c)\bar{H}(c)}{F(t|Y\geq c)\bar{H}(c)+F(t|Y<c)H(c)},$$

$$H(c|T\geq t)=\frac{\bar{F}(t|Y<c)H(c)}{\bar{F}(t|Y\geq c)\bar{H}(c)+\bar{F}(t|Y<c)H(c)}.$$

The reason underlying such a transformation is that the estimators of both expressions, introduced in the next section, ensure to be bounded in $[0,1]$.

The time-dependent ROC curve is now defined as follows

$$ROC_t=\{(1-H(c|T\geq t),\bar{H}(c|T<t)),\ c\in\boldsymbol{R}\},$$

or equivalently

$$ROC_t=\{(\bar{H}(c|T\geq t),\bar{H}(c|T<t)),\ c\in\boldsymbol{R}\}.$$

In practice, the time to event, T, which determines the true membership status of the i-th individual to one of the subpopulations $\mathcal{G}_0$ or $\mathcal{G}_1$, is often right-censored, therefore estimators of $F(\cdot\,|Y\geq c),\bar{F}(\cdot\,|Y\geq c)$ and $F(\cdot\,|Y<c),\bar{F}(\cdot\,|Y<c)$ adapted to right-censored data are required. Such estimators will be proposed in the next section. They are based on the sequential right-censored random samples and the well-known Kaplan-Meier estimation (Kaplan & Meier, 1958).

Sequential random sampling and the Kaplan-Meier estimator

It is usually expected that random copies T_i of time-to-event variable T are fully observable. However, this assumption becomes problematic when it cannot be satisfied due to right-censoring. Further, we will assume that censored survival times are observed, i.e. $X_i=\min(T_i,Z_i)$ instead of T_i, where Z_i are the so-called censoring times.

To estimate $F(\cdot\,|Y \geq c), \bar{F}(\cdot\,|Y \geq c)$ as well as $F(\cdot\,|Y < c), \bar{F}(\cdot\,|Y < c)$, we propose adopting a sequential sampling scheme which allows for an unbiased estimation of the afore-mentioned functions for any $t < t_0$ where t_0 is a positive real number fixed in advance, as opposed to the negatively biased standard Kaplan-Meier estimator $KM(t)$ which is undefined for some t if the largest observation in the sample is right-censored.

Suppose that there is given a fixed period t_0 of observation of individuals in a study. Individuals enter the study at random times, and the observation finishes at the end of that period. If the interesting event occurs for an i-th individual, than it yields the true time-to-event T_i, otherwise one observes a partial time to event, say Z_i, such that $Z_i < T_i$. In such a case, we say that Z_i right-censors the true time T_i. Variables Z_i are usually called censoring times. From a formal point of view, one observes the so-called censored survival times, which can be expressed as pairs (X_i, δ_i), where

$$X_i = \min(T_i, Z_i),$$

and

$$\delta_i = \mathbf{1}(T_i \leq Z_i) = \begin{cases} 1, & \text{if } T_i \leq Z_i, \\ 0, & \text{otherwise.} \end{cases}$$

is the censoring indicator.

If the censoring times Z_i are random and independent of survival times T_i then the model (X_i, δ_i) is well known as the random and independent right-censorship model.

We will assume that the times-to-event, T_i, as well as the censoring times, Z_i, are continuous, mutually independent random variables with common distribution functions F and G, respectively. These assumptions imply that variables $X_i = \min(T_i, Z_i)$ are independent and have a distribution function, say K, satisfying the relation $K = 1 - \bar{F}\bar{G}$.

The sequential random sampling scheme will be defined as follows. Let t_0, c_0 be real fixed values, such that $0 < t_0 < \sup\{t: \bar{K}(t) > 0\}$, $c_0 < \sup\{c: \bar{H}(c) > 0\}$, and let $k \geq h \geq 2$ be some predefined integers. Assume that individuals arrive at random during the study and we record (X_i, δ_i) as well as Y_i until for k individuals the following inequalities hold

$$X_{i_j} \geq t_0, \qquad j = 1,2, \dots, k,$$

and for h of them, the following inequalities are satisfied at the same time

$$Y_{i_j} \geq c_0, \qquad j = 1,2, \dots, h.$$

It is worth noting that the size N of the random sample

$$(X_1, \delta_1, Y_1), (X_2, \delta_2, Y_2), \dots, (X_N, \delta_N, Y_N),$$

obtained in such a sampling scheme is a random variable.

Since we need a sorted sample, let us also define the $\leq_l$ relation as follows

$$(X_i, \delta_i, Y_i) \leq_l (X_j, \delta_j, Y_i) \Leftrightarrow (X_i < X_j) \text{ or } \big(X_i = X_j \text{ and } \delta_i \geq \delta_j\big).$$

An i-th observation in the sample reordered according to this relation will be denoted by $(X_{i:N}, \delta_{[i]}, Y_{[i]})$, where $\delta_{[i]} = \delta_j$ and $Y_{[i]} = Y_j$, if $X_{i:N} = X_j$.

Then, the sequential Kaplan–Meier-type estimator $KM(t)$ of the probability $\bar{F}(t)$ takes the form

$$KM(t) = \prod_{i:X_{i:N} \leq t} \left(1 - \frac{\delta_{[i]}}{N - i}\right), \qquad t \leq t_0,$$

with the initial assumption that $KM(t) = 1$ whenever the set $\{i: X_{i:N} \leq t\}$ is empty. It is worth noting that such a sequential estimator is an unbiased estimator of the cumulative distribution function $\bar{F}$, even in the presence of right-censoring (see Rossa, 2008, 2009 for more details).

By analogy, the sequential Kaplan–Meier estimators $KM(t|Y < c)$, $KM(t|Y \geq c)$ of the respective probabilities $\bar{F}(t|Y < c)$ and $\bar{F}(t|Y \geq c)$ are as follows

$$KM(t|Y < c) = \prod_{i:X_{i:N} \leq t, Y_{[i]} < c} \left(1 - \frac{\delta_{[i]}}{N_1 - i}\right), \ t \leq t_0, \ c \leq c_0,$$

and

$$KM(t|Y \geq c) = \prod_{i:X_{i:N} \leq t, Y_{[i]} \geq c} \left(1 - \frac{\delta_{[i]}}{N_2 - i}\right), \ t \leq t_0, \ c \leq c_0,$$

where

$$N_1 = \#\{i: Y_{[i]} < c\}, \ N_2 = \#\{i: Y_{[i]} \geq c\}.$$

For a fixed value of $t \leq t_0$, the empirical ROC curve takes the form

$$\widehat{ROC}(c, t) = \left\{\left(\widehat{\widehat{H}}(c|T < t), \widehat{\widehat{H}}(c|T \geq t)\right), \ c \leq c_0\right\},$$

where $\widehat{\widehat{H}}(c|T < t)$ and $\widehat{\widehat{H}}(c|T \geq t)$ are defined as follows

$$\widehat{\widehat{H}}(c|T < t) = \frac{[1 - KM(t|Y \geq c)]\widehat{\widehat{H}}(c)}{1 - [KM(t|Y \geq c)\widehat{\widehat{H}}(c) + KM(t|Y < c)\widehat{H}(c)]},$$

$$\hat{\hat{H}}(c|T \geq t) = 1 - \frac{KM(t|Y < c)\hat{H}(c)}{KM(t|Y \geq c)\hat{\hat{H}}(c) + KM(t|Y < c)\hat{H}(c)},$$

with the convention that $0/0 \equiv 0$.

The functions $\hat{\hat{H}}(c)$ and $\hat{H}(c)$ appearing in the numerators and denominators of both estimators, are defined as follows

$$\hat{\hat{H}}(c) = \frac{1}{N-1}\sum_{i=1}^{N-1} \mathbf{1}\,(Y_i \geq c), \qquad \hat{H}(c) = 1 - \hat{\hat{H}}(c).$$

Note that estimators $\hat{\hat{H}}(c|T < t)$ and $\hat{\hat{H}}(c|T \geq t)$ are bounded in $[0,1]$.

Now, based on the estimator of the ROC curve it is possible to find an optimal predictive threshold c of the marker Y that ensures as high classifications accuracy as possible. This process will be illustrated in more detail using a credit-scoring application.

Application of the new class of ROC curve estimators in credit scoring

Credit scoring is the term used to describe formal statistical methods used for classifying applicants for credit into ‘good’ and ‘bad’ risk classes. Such methods have become important with the significant growth in consumer credit in recent years. A wide range of statistical methods has been applied to address this issue (see e.g., Hand & Henley, 1997; Abellan & Castellano, 2017, among others). One of them is based on time-dependent ROC curves. In the practice of credit-risk management, the models based on the receiver operating characteristic (ROC) curves are helpful in assessing e.g. the discriminatory power of the so-called scorecards.

Among consumers borrowing money from banks, there are almost always those who do not repay their loans according to the scheduled payments; such loans are called defaults. In this example, we consider the simple case of a fictitious bank that classifies 12-month loans as ‘good’ or ‘bad’ based on the probability that they will be fully repaid or will not be repaid before a fixed number of t months have passed. Typically, the classification should be performed before a decision to grant the loan is made, i.e., at the application stage. Since the bank cannot know in advance whether the loan will be repaid, creditworthiness has to be established using other available information, and typical sources for this are the loan application form, the applicants’ financial history and other characteristics.

To keep our presentation simple, let us consider 12-month loans and assume that the predictive marker Y_i is the monthly income of the i-th applicant. The classification rule is as follows: an i-th loan will be classified as ’good’ when the value of Y_i is at least c, where c

is some decision threshold. The main problem with this approach is that we first need to find the optimal value of c that is based on the most recent data and has the highest probabilities of correctly classifying loans as repaid before t months have elapsed. To solve this problem, the time-dependent ROC curve will be employed. However, a random sample of the most recent loans approved by the bank is needed in order to establish a correct classification rule.

Let T_i represent the time between the date an i-th loan was approved and the date of its repayment, and let Z_i denote the time between the approval date and the current date (censoring time). The repayment time T_i of the loan longer than censoring time Z_i are not observable, which means that we can observe the smaller one, i.e. $X_i = \min(T_i, Z_i)$. We will call X_i the censored repayment time for the i-th loan. Observations are collected sequentially, and after each new loan is included, the sample is evaluated against a predefined stopping rule. The data collection continues until at least k observations satisfy the condition $X_i \geq t_0$, ensuring sufficient information about loans with relatively long durations, and simultaneously at least h observations satisfy the condition $Y_i \geq c_0$, ensuring adequate representation of individuals with higher values of the predictive marker.

This stopping rule ensures that the sample contains enough information to reliably estimate time-dependent classification measures. In the illustrative example, the parameters are set to $k = h = 2$, $c_0 = 2.2$ (Euro $1{,}000$) and $t_0 = 27$ (months). The sample censored repayment times X_i, censoring indicators δ_i, and monthly incomes Y_i yielded by the experiment are shown in Table 1.

Table 1.

Input data

Censored time X_i	Censoring indicator δ_i	Income per month Y_i (Euro 1,000)
4	1	1.4
6	0	1.8
7.4	1	1.5
8	0	1.7
9.5	1	1.5
10.1	1	1.3
10.6	0	1.9
11	0	2

Censored time X_i	Censoring indicator δ_i	Income per month Y_i (Euro 1,000)
11.5	1	1.6
12.3	1	1.6
12.5	0	1.8
13	1	1.8
13.2	1	1.6
13.5	1	1.7
13.9	0	1.8
14	0	1.8
14.2	0	1.9
14.5	0	1.9
15.1	1	1.9
16	0	1.8
17.8	0	1.9
18	0	1.9
18.5	0	2
20.5	1	2
20.8	1	1.9
21	0	2
22	1	1.7
24	0	1.9
27.1	1	2
27.5	1	2
28	0	2.2
29	1	2.5

Source: author's own elaboration (sample data).

Table 2 reveals specificity and sensitivity values obtained using the data in Table 1 for some selected values of $t \leq t_0$.

Table 2.

Classification accuracy via sensitivity and 1-specificity measures

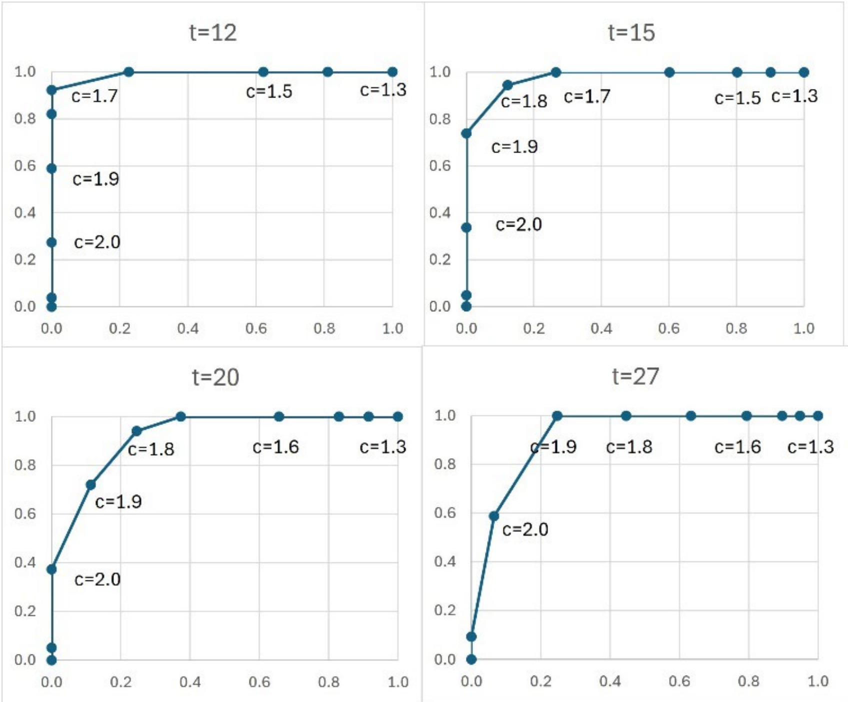
c	1-spec.	sens.	c	1-spec.	sens.
t=12 months			t=20 months		
1.3	1	1	**1.3**	1	1
1.4	1	0.8106	**1.4**	1	0.9148
1.5	1	0.6212	**1.5**	1	0.8296
1.6	1	0.2269	**1.6**	1	0.6569
1.7	0.9231	0	**1.7**	1	0.3738
1.8	0.8212	0	**1.8**	0.9407	0.2455
1.9	0.5895	0	**1.9**	0.7199	0.1131
2	0.2735	0	**2**	0.3735	0
2.1	0.039	0	**2.1**	0.0523	0
2.2	0	0	**2.2**	0	0
t=15 months			t=27 months		
1.3	1	1	**1.3**	1	1
1.4	1	0.9013	**1.4**	1	0.9486
1.5	1	0.8026	**1.5**	1	0.8971
1.6	1	0.6018	**1.6**	1	0.7938
1.7	1	0.2652	**1.7**	1	0.6336
1.8	0.9453	0.1224	**1.8**	1	0.4463
1.9	0.7387	0	**1.9**	1	0.2474
2	0.3382	0	**2**	0.5886	0.0652
2.1	0.0482	0	**2.1**	0.0934	0
2.2	0	0	**2.2**	0	0

Source: Author's calculations based on sample data from the Table 1.

Based on data in columns 2 and 3 of Table 2, the coordinates of the ROC curve corresponding to some selected values of t can be plotted using the linear interpolation method (Figure 2). The optimal value of the predictive threshold c of the income per month is given by the point on the ROC curve, which is the nearest to the left upper corner of the unit square, i.e., whose coordinates are $(0,1)$. In our example, the optimal value of c is 1.7 (Euro 1,000) if $t = 12$ (months), 1.8 (Euro 1,000) if $t = 15$ or $t = 20$ (months) and $c = 1.9$ (Euro 1,000) if $t = 27$ (months), as these values ensure relatively high probabilities that loans have been correctly classified into good or bad ones.

Figure 2.

Estimated ROC curves for four selected values of t



Source: author’s own elaboration.

Final remarks

The proposed approach offers several advantages over traditional classification methods. First, it incorporates the time-to-event dimension, allowing classification to depend on whether an event occurs before or after a specified time t, rather than treating outcomes as static. Second, it properly handles right-censored data using the Kaplan–Meier estimator, avoiding information loss and potential bias. Third, time-dependent ROC curves enable a dynamic evaluation of classification performance across different time horizons. Fourth, the sequential sampling scheme ensures that the sample contains sufficiently informative observations, improving estimation efficiency. Fifth, the method allows for data-driven selection of the optimal threshold c, maximizing classification accuracy. From a credit scoring perspective, this enables more accurate assessment of default risk over a specified time horizon, supporting better-informed lending decisions and risk management.

However, the approach has some limitations. It relies on the assumption of independent right-censoring. Violation of this assumption may lead to biased estimates. Moreover, the procedure can be computationally demanding when multiple thresholds or time points are considered.

In summary, the proposed classification method, incorporating time-dependent ROC curves, sequential sampling, and Kaplan–Meier estimation, provides a more accurate and reliable determination of the optimal threshold for a continuous predictive marker in the presence of right-censored time-to-event data than traditional static classification approaches.

Disclosure statement

No potential conflict of interest is reported by the author.

References

Abellan, J. & Castellano, J.G. (2017). A comparative study on base classifiers in ensemble methods for credit scoring. *Expert Systems with Applications* 73(1): 1–10. https://doi.org/10.1016/j.eswa.2016.12.020.

Altman, D.G. & Bland, J.M. (1994). Diagnostic tests. 1: Sensitivity and specificity. *British Medical Journal* 308(6943): 1552. https://doi.org/10.1136/bmj.308.6943.1552.

Altman, E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance* 23(4): 589–609. https://doi.org/10.1111/j.1540-6261.1968.tb00843.x.

Baker, S.G. (2003), The central role of receiver operating characteristic (roc) curves in evaluating tests for the early detection of cancer. *Journal of the National Cancer Institute* 95(7): 511–515. https://doi.org/10.1093/jnci/95.7.511.

Breiman, L., Friedman, J.H., Olshen, R.A., et al. (1984), *Classification and Regression Trees*. Chapman and Hall/CRC, New York.

Chambless, L. & Diao, G. (2006), Estimation of time-dependent area under the roc curve for long-term risk. *Statistics in Medicine* 25(20): 3474–3486. https://doi.org/10.1002/sim.2299.

Crook, J.N., Edelman, D.B. & Thomas, L.C. (2007), Recent developments in consumer credit risk assessment. *European Journal of Operational Research* 183(3): 1447–1465. https://doi.org/10.1016/j.ejor.2006.09.100.

Diaz-Quijano, F.A. (2012), A simple method for estimating relative risk using logistic regression. *BMC Medical Research Methodology* 12(14): 1471–2288. https://doi.org/10.1186/1471-2288-12-14.

Dickie, G.L. (1994), Statistical notes. defining sensitivity and specificity. *British Medical Journal* 309(6953): 539. https://doi.org/10.1136/bmj.309.6953.539a.

Fisher, R.A. (1936), The use of multiple measurement in taxonomic problems. *Annals of Eugenics* 7(2): 179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x.

Godyn, J.J., Tomaszewski, J.E. & Zmijewski, C.M. (1991), Specificity of a reproducible qualitative urine examination is not a constant test characteristic. *American Journal of Clinical Pathology* 95(2): 265–266. https://doi.org/10.1093/ajcp/95.2.265.

Green, D.M. & Swets, J.A. (1966), *Signal detection theory and psychophysics*. Wiley, New York.

Hand, D.J. & Henley, W.E. (1997), Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society A: Statistics in Society* 160(3): 523–541. https://doi.org/10.1111/j.1467-985X.1997.00078.x.

Heagerty, P.J., Lumley, T. & Pepe, M.S. (2000), Time-dependent roc curves for censored survival data and a diagnostic marker. *Biometrics* 56(2): 337–344. https://doi.org/10.1111/j.0006-341x.2000.00337.x.

Hu, Y.C. & Ansell, J. (2007), Measuring retail company performance using credit scoring techniques. *European Journal of Operational Research* 183(3): 1595–1606. https://doi.org/10.1016/j.ejor.2006.09.101.

Hung, H. & Chiang, C. (2010a), Estimation methods for time-dependent AUC models with survival data. *Canadian Journal of Statistics* 38(1): 8–26. https://doi.org/10.1002/cjs.10046.

Hung, H. & Chiang, C. (2010b), Optimal composite markers for time-dependent receiver operating characteristic curves with censored survival data. *Scandinavian Journal of Statistics* 37(4): 664–679. https://doi.org/10.1111/j.1467-9469.2009.00683.x.

Jamain, A. & Hand, D.J. (2008), Mining supervised classification performance studies: A meta-analytic investigation. *Journal of Classification* 25(1): 87–112. https://doi.org/10.1007/s00357-008-9003-y.

Kaplan, E.L. & Meier, P. (1958), Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 53(282): 457–481. https://doi.org/10.2307/2281868.

Kovacova, M. & Kliestik, T. (2017), Logit and probit application for the prediction of bankruptcy in Slovak companies. *Equilibrium-Quarterly Journal of Economics and Economic Policy* 12(4): 775–791. https://doi.org/10.24136/eq.v12i4.40.

Kovacova, M., Kliestik, T., Valaskova, K., et al. (2019), Systematic review of variables applied in bankruptcy prediction models of visegrad group countries. *Oeconomia Copernicana* 10(4): 743–772. https://doi.org/10.24136/oc.2019.034.

Lloyd, C.J. (1998), Using smoothed receiver operating characteristic curves to summarize and compare diagnostic systems. *Journal of the American Statistical Association* 93(444): 1356–1364. https://doi.org/10.1080/01621459.1998.10473797.

McNutt, L.A., Wu, C., Xue, X., et al. (2003), Estimating the relative risk in cohort studies and clinical trials of common outcomes. *American Journal of Epidemiology* 157(10): 940–943. https://doi.org/10.1093/aje/kwg074.

Ohlson, J.A. (1980), Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research* 18(1): 109–131. https://doi.org/10.2307/2490395.

Patil, M. & Durairaj, M. (2015), Risk factors of hypertension among adult men: Evidence from a real world outcomes investigation in a western Indian population. *International Journal of Advanced Research* 3(7): 274–282. https://doi.org/10.2307/2490395.

Pepe, M.S. (1997), A regression modelling framework for receiver operating characteristic curves in medical diagnostic testing. *Biometrika* 84(3): 595–608. https://doi.org/10.1093/biomet/84.3.595.

Pepe, M.S. (1998), Three approaches to regression analysis of receiver operating characteristic curves for continuous test results. *Biometrics* 54(1): 124–135. https://doi.org/10.2307/2534001.

Pepe, M.S. (2000), Receiver operating characteristic methodology. *Journal of the American Statistical Association* 95(449): 308–311. https://doi.org/10.1080/01621459.2000.10473930.

Rossa, A. (2008), Estimation of survival distributions under right-censoring when sample size is random. *Sequential Analysis* 27(2): 174–184. https://doi.org/10.1080/0747494080198915.

Rossa, A. (2009), The Nelson-Aalen and Kaplan-Meier estimators under a sequential sampling scheme. *Communications in Statistics – Theory and Methods* 38(16–17): 3077–3098. https://doi.org/10.1080/03610920902947568.

Tetrault, G. (1991), Sensitivity and specificity of clinical tests. *American Journal of Clinical Pathology* 96(4): 556. https://doi.org/doi:10.1093/ajcp/96.4.556.

Thomas, L.C. (2000), A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *International Journal of Forecasting* 16(2): 149–172. https://doi.org/10.1016/S0169-2070(00)00034-0.

Thomas, L.C. (2009), Modelling the credit risk for portfolios of consumer loans: Analogies with corporate loan models. *Mathematics and Computers in Simulation* 79(8): 2525–2534. https://doi.org/10.1016/j.matcom.2008.12.006.

Uno, H., Cai, T., Tian, L., et al. (2007), Evaluating prediction rules for t-year survivors with censored regression models. *Journal of the American Statistical Association* 102(478): 527–537. https://doi.org/10.1198/016214507000000149.

Zhang, J. & Yu, K.F. (1998), What's the relative risk? a method of correcting the odds ratio in cohort studies of common outcomes. *Journal of the American Medical Association* 280(19): 1690–1691. https://doi.org/10.1001/jama.280.19.1690.