InfoSci®-ONDemand

Chapter Download

Chapter 16
Survival Analysis and ROC Analysis in Analyzing Credit Risks:
Assessing Default Risks Over Time

Nan Hu
University of Utah, USA

Haojie Cheng
University of Utah, USA

ABSTRACT

As the aim of large banks has been changing to select customers of highest benefits, it is important for banks to know not only if but also when a customer will default. Survival analyses have been used to estimate over time risk of default or early payoff, two major risks for banks. The major benefit of this method is that it can easily handle censoring and competing risks. An ROC curve, as a statistical tool, was applied to evaluate credit scoring systems. Traditional ROC analyses allow banks to evaluate if a credit-scoring system can correctly classify customers based on their cross-sectional default status, but will fail when assessing a credit-scoring system at a series of future time points, especially when there are censorings or competing risks. The time-dependent ROC analysis was introduced by Hu and Zhou to evaluate credit-scoring systems in a time-varying fashion and it allows us to assess credit scoring systems for predicting default by any time within study periods.

INTRODUCTION

Risk managements, in general, are the identification, assessment and prioritization of risks followed by application of resources to monitor the probability and impact of unfortunate events (Hubbard 2009). Among different risks, financial risk is the high-priority risk for all businesses. The financial risk can be classified into various types such as market risk, credit risk, liquidity risk, operational risk and legal risk (Ray 2011). In finance, the risk management usually focuses on market risks and credit risks. Mar-

DOI: 10.4018/978-1-4666-9458-3.ch016

ket risks present in the market, and usually arise from market movements. The equity risk, interest risk, currency risk and commodity risk are different types of market risks. The credit risk refers to risk that borrowers will default on their debts by failing to make payments. The credit risk management deals with the probability of non-payment from the debtors and is the fundamental work of financial institutions. With the recent global financial crisis and the following credit crunch, more attentions have been paid to analyzing and managing credit risks. In this chapter, we focus on analyzing credit risks using survival analyses and on evaluating the accuracies of credit scoring systems for predicting default risks using a receiver operating characteristic (ROC) curve. In the Background section, we give the motivation for applying survival models and ROC curve methods for analyzing credit risks. Then, we review the applications of survival analysis in analyzing default risks and illustrate, in detail, how survival analysis methods can be implemented to estimate hazard of default and survival functions (proportion of default-free), to compare survival functions among different groups, and to make statistical inference of covariate effect on the risk of default. In addition, we demonstrate how ROC analysis can be applied to assess accuracies of credit scoring systems for predicting the cross-sectional status of default using a traditional ROC curve. We give mathematical definitions for different classification accuracy parameters, such as sensitivity and specificity, and describe the procedure to establish an ROC curve. Furthermore, we describe the extension from traditional ROC curves to time-dependent ROC curves, which can be used to evaluate the classification accuracies for predicting defaults by a specific future time in a time-varying fashion. Audiences of this book chapter should realize that survival analyses and ROC analyses, in general are two separate topics. However, the time-dependent ROC method is actually combing traditional ROC curve approach with survival analysis methods. We conclude this chapter by summarizing the application of survival analyses and ROC methods in analyzing credit risks and providing future research directions on these two topics.

BACKGROUND

The primary goals of risk analysis for the credit industry are to monitor the credit risks and to decide whether it should grant credit to an applicant. Traditionally, it is done by estimating the probability that the applicant will eventually default (Stepanova & Thomas 2002). More recently, however, the aim of banks and credit companies has been changing to select customers of highest benefits. This changes implies that it is important not only if but also when a customer will default in a future time (Banasik *et al.* 1999). It is possible that if the time to default is long, the acquired interest from the customer will compensate or even exceed the losses resulting from default. On the other hand, an early pay-off by the customer can also impact the profitability of financial businesses. Depending on when the actual repayment occurs, the lender will lose a proportion of interest in the loan. It has been shown that survival analysis can be applied to estimate time to default or early repayment (Narian 1992, Banasik *et al*. 1999).

The survival analysis (also known as the time-to-event analysis), is a branch of statistics dealing with analysis of time until one or more events happens. The major benefit of using the survival analysis lies in the fact that it can handle the censored data when modeling the hazard or risk for the default or the early pay-off. Survival analysis has been intensively applied in medical research and reliability studies. It allows researchers to incorporate information from both censored and uncensored observations in estimating model parameters. In analyzing credit risks for the financial industry, profits realized on loans, credit cards or their related products depend heavily on whether borrowers pay interest regularly, missing payments, or default. Survival analysis can directly relate predictor variables - variables used to predict the survival outcomes - to the risk of default or events related to default, such as early pay-off. By establishing a predictive regression model, analysts in the financial industry are able to develop credit scoring systems to predict default and other risky events in the future Credit risk models have been discussed in previous works such as Crouhy *et al*. (2000) and Duffie *et al*. (2003). Usually, a credit scoring system was utilized to estimate the probability that a loan will default before the end of the pre-specified loan term. Based on the probability, banks can decide whether loans should be granted to borrowers. The ROC curve analysis, as a statistical tool, has been applied to evaluate the performance of credit scoring systems (Chen *et al*. 2010, Gupta *et al*. 2014, Miura *et al*. 2010). Traditional ROC curve allows financial businesses to evaluate whether a credit scoring system can correctly classify customers based on cross-sectional status of default. For example, it can evaluate accuracies for predicting defaults at the end of a loan period, or at a fixed time point (e.g, 24 months since the loan starts). However, the traditional ROC analysis can only deal with the binary decision space. Hence, it will fail when evaluating performance of a credit scoring system at a series of time points within the entire loan term (in which situation there are infinite number of time points and therefore the decision space is infinitely dimensional), especially when censored data exists. Hu & Zhou (2014) proposed to use time-dependent ROC analysis for evaluating credit-scoring systems under the time-dependent infrastructure, which allows us to evaluate scoring systems for predicting default by any of the future time points.

In the following sections, we provide detailed descriptions and explanations regarding survival analysis and ROC methods in analyzing credit risks.

ANALAYZING CREDIT RISKS USING SURVIVAL MODELS

A number of techniques have been applied to model credit risks. Durand (1941) pioneered the use of discriminant analysis for credit scoring; Orgler (1970) applied regression analysis in a model for commercial loans; Wiginton (1980) was among the first studies to obtain credit scoring results by using logistic regression (LR), and compared the LR method with discriminant analysis; Leonard (1993) applied LR model to evaluate commercial loans. In addition, decision tree approaches was adopted by, e.g., Metha (1968) and Makowski (1985) (Wekesa *et al*. 2012).

Lane *et al.* (1986) and Luoma & Laitinen (1991) are the first two studies that applied survival analysis for failure predictions. Narain (1992) and Banasik *et al*. (1999) showed that the survival analysis can be used to predict the time to default or the time to early repayment. Specifically, Narain (1992) applied Exponential accelerated failure time (AFT) model to 24 months of loan data and showed that the proposed model reliably estimated the number of failures at each of the event times. The author then built a scorecard based the multivariable regression and found that a better credit-granting decision could be made if the score is supported by the survival time. Narain (1992) concluded that survival analysis add an extra dimension on the standard analysis of credit risk at that time and can be applied to any area of credit operation in which the associations between predictor variables and certain kind of event are of interest. Banasik *et al*. (1999) mentioned that, for financial institution, it is important not only if but also *when* a customer will default. They believed that it is possible that if the time to default is long, financial institutions may acquire more interest that may compensate or even excess losses induced by default. They claimed that early payoff and account closing can also affect the profitability of the credit industry. Banks and credit card companies will lose a proportion of interest on the loan depending on when the customers close their account or pay off the loan (note: we believe that refinancing could be another risk for interest loss). Banasik *et al*. (1999) compared the accuracy performance of predicting defaults by 12 months and 24 months among four different models: parametric survival model under Exponential distribution, parametric model under Weibull distribution, semi parametric Cox proportional hazards (PH) model, and the LR model. They found that survival-analysis methods are competitive with, and sometimes superior to, the traditional LR approach. In addition, they believed that default and early payoff can be considered as competing risks. Furthermore, they suggested that early payoff events could be treated as censorings when estimating default risks, and, similarly, defaults should be treated as censoring events when calculating risk of early payoff. Thomas (2000) demonstrated that the survival analysis approach provides more detailed and relevant information for credit management than traditional approaches. Thomas and Stepanova (2002) showed how survival-analysis tools allow one to build credit-scoring models that assess aspects of profit as well as default. The paper looked at three extensions of the Cox's PH model and applied them to a personal loan data.

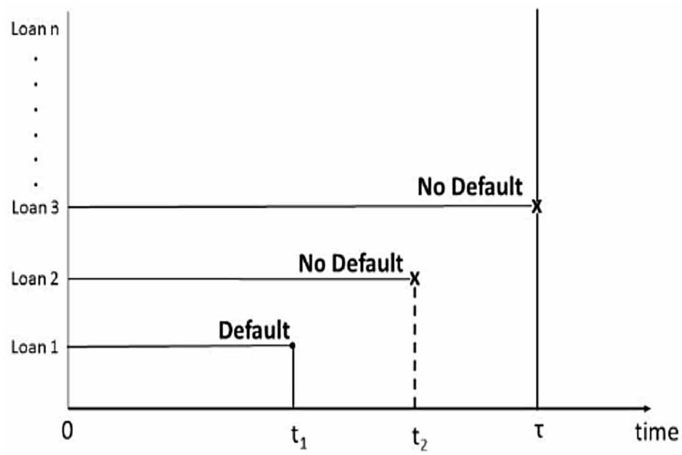
Noh *et al*. (2005) used information from credit card centers in South Korea to develop a prediction model for personal credit risk using time-dependent predictor variables. Mavri *et al*. (2008) proposed a two-stage dynamic credit scoring model which estimates the credit performance of an applicant using generalized linear models, and accommodated the changes of a borrower's characteristics after the issuance of the credit card to forecast the default time using survival analysis. Tsai *et al.* (2009) constructed a consumer loan default prediction model through conducting the empirical analysis on consumers of unsecured loan through four different prediction methods, including survival models for neural network analysis. By implementing a discrete-time survival model, Francesca (2012) showed that "when default occurs" is an important element to predict the probability of default. Im *et al.* (2012) introduced the time-dependent PH survival model to capture temporal phenomena of credit risks, and concluded that incorporating the time dependency can improve accuracy of risk scoring systems. Witzany *et al*. (2012) applied a survival analysis methodology to estimate the Loss Given Default (LGD) parameters and showed that Cox PH model applied to LGD performed better than the linear and logistic regressions. Edward *et al.* (2012) proposed a mixture cure models for predicting time-to-default and compared the performance of their model with Cox PH and standard LR models using data from a UK personal loan portfolio.

Setting of Survival Analysis for Credit Risks

Figure 1 demonstrates three common situations that may occur during the lifetime of a loan or credit (please note this is not a regular x-y plot in which variable of y-axis is a function of variable presented on x-axis). Assume that the study period is within the time interval $[0, \tau]$, where time 0 is the time that a loan or a credit starts and τ is time at the end of the study (which can be either less than or equal to time at the end of the loan). In this demonstrative figure, the horizontal axis is the time to defaulting events and the lines denote the observed time for each of the corresponding loans. A solid point represents a defaulting event and a cross stands for a censoring. In the first case (loan 1), the loan defaults before the end of study at time τ, so the exact defaulting time can be recorded. In the second case (loan 2), the loan ends before end of the study with no default. In the last case (loan 3), the loan lasts until the end of study without a default. In the last two cases, we cannot observe the exact time at default.

Assume that the study data are collected in a typical setting of time-to-default with covariates, variables that is possibly predictive of the survival outcome. In this book chapter, we will use predictor

Figure 1. Demonstration of Time-to-default



variables and covariates interchangeably. Suppose that there are n loans in the dataset and let T_i and C_i be the failure and censoring time, respectively for loan $i\,(i = 1, 2, \ldots, n)$. The censoring time is usually the time to leaving the loan due to reasons other than default, such as the end of the study period (e.g., when we study the risk of default by 24 months for loans with terms longer than 24 months). In reality, what we record are the minimum of default time and censoring time together with an indicator variable specifying whether a default (versus a censoring) occurred. We use X_i to denote the observed time and Δ_i to denote the event indicator for loan *i*. Mathematically, $X_i = \min(T_i, C_i)$ and $\Delta_i = 1(T_i \le C_i)$. The indicator function 1(.) gives 1 if condition in its argument is satisfied and 0 otherwise. Therefore, if $\Delta_i = 1$, the observed time X_i is the default time, i.e., $X_i = T_i$. Otherwise, when $\Delta_i = 0$, we record the censoring time, i.e., $X_i = C_i$. We use $\mathbf{Z}_i$ to denote a *vector* of covariates that can affect the default time for loan *i*. Assume there are p predictor variables, then the covariate vector, $\mathbf{Z}_i = (Z_{i1}, Z_{i2}, \ldots, Z_{ip})$. The default time T is a random variable with a distribution and can usually be modeled through the following five functions:

Probability density function (PDF): $f\left(t\right) = \lim\limits_{\Delta t \to 0} \dfrac{Pr(t \le T \le t + \Delta t)}{\Delta t}$,

Cumulative distribution function (CDF): $F(t) = \int_0^t f(u)du = Pr(T \le t)$,

Survival function: $\mathrm{S}(t) = 1 - F(t) = \int_t^\infty f(u)du = P\,\mathrm{r}(T > t)$,

Hazard function: $h(t) = \lim\limits_{\Delta t \to 0} \dfrac{Pr(t \le T \le t + \Delta t \mid T \ge t)}{\Delta t} = \dfrac{f(t)}{s(t)} = d\dfrac{-\ln[S(t)]}{dt}$, and

Cumulative hazard function: $H(t) = \int_0^t h(u)du = -\ln[S(t)]$,

where $\ln[S(t)]$ is the natural logarithm of $S(t)$. The above functions are evaluated at time t.

In the following subsections, we provide statistical methods for estimating and comparing the survival functions, and for making inference of covariate effects on the default time T.

Kaplan-Meier Estimators

Assume there are k unique event times among the *n* loans (where *k* may not be equal to *n* due to ties) and t_i is the *j*th unique event time in the study such that $j = 1, 2, \ldots, k, t_1 < t_2 < \ldots < t_k$, and $t_{k+1} = \infty$. Let n_j be the number of loans at risk of default immediately prior to time t_j and d_j be the number of loans defaulting at t_j. Then the non-parametric Kaplan-Meier estimator (Kaplan and Meier 1958) for survival function of random variable T at time *t* is given by:

$$\hat{S}(t) = \prod_{t_j \le t}(1 - \frac{d_j}{n_j}) \approx \prod_{t_j \le t}\exp(-\frac{d_j}{n_j}) = \exp\left[-\sum_{t_j \le t}(\frac{d_j}{n_j})\right] \tag{1}$$

where $\exp(.)$ denotes the exponential function. Kaplan-Meier approach (1) provides a way to estimate the proportion of default-free loans by the time *t*, $\hat{\mathrm{S}}(t)$.The CDF of default time can be easily calculated as $\hat{\mathrm{F}}\left(t\right) = 1 - \hat{\mathrm{S}}(t)$. However, it provides information on neither comparing survival function among different groups nor evaluating effects of predictor variables on risk of default. Thus, it can only give a summary statistic but provide information on statistical inferences and predictions.

Log Rank Test for Comparing Survival Functions

Sometimes, banks and credit companies may be interested in testing the equality of survival functions of default among different groups of customers (e.g., males versus females, older versus younger borrowers, or loan applicants with different income levels). Under this situation, the log rank test (Mantel 1966), a non-parametric rank statistic based method, can be used for comparison functions between and among groups. Table 1 shows the data structure for log rank statistic comparing two groups (or, two-sample log rank statistic). Again, we assume that there are k unique time points, $t_1 < t_2 < \ldots < t_k$. Let $j = 1, 2, \ldots, k$ be the index of the unique default time, t_j. Let Z be the grouping variable such that $Z = 1$ indicates the first group and $Z = 2$ indicates the other group. We assume that Y_j loans are subject to default immediately before time t_j, and D_j loans default at t_j. Additionally, we use D_{1j} to denote the number of defaulting loans in group 1 at t_j. Let $E(D_{1j})$ and $Var(D_{1j})$, respectively, be the expectation and variance of D_{1j}. Given D_j and Y_j, the random variable D_{1j} has a hypergeometric distribution from a total of Y_j at-risk loans. Thus, $E(D_{1j})$ and $Var(D_{1j})$ can be obtained using the following equations:

Table 1. Two-sample log rant statistic data structure

t	Z	Failures	No Failures	N
$t = t_1$	$Z = 1$	D_{11}	$Y_{11} - D_{11}$	Y_{11}
	$Z = 2$	D_{21}	$Y_{21} - D_{21}$	Y_{21}
	subtotal	D_1	$Y_1 - D_1$	Y_1
$t = t_k$	$Z = 1$	D_{1k}	$Y_{1k} - D_{1k}$	Y_{1k}
	$Z = 2$	D_{2k}	$Y_{2k} - D_{2k}$	Y_{2k}
	subtotal	D_k	$Y_k - D_k$	Y_k
Total		$\sum_{j=1}^{k} \mathbf{D_j}$	$\sum_{j=1}^{k} \mathbf{Y_j} - \sum_{j=1}^{k} \mathbf{D_j}$	$\sum_{j=1}^{k} \mathbf{Y_j}$

The log rank statistic for comparing survival functions of the two groups is given by:

$$E(D_{1j}) = \frac{D_k Y_{1k}}{Y_k}$$

and

$$Var(D_{1j}) = \frac{Y_{1k} Y_{2k} D_k (Y_k - D_k)}{Y_k^2 (1 - Y_k)}.$$

$$\frac{[\sum_{j=1}^{k} (D_{1j} - E_{1j})]^2}{\sum_{j=1}^{k} \mathrm{var}(D_{1j})} \qquad (2)$$

Mantel (1966) showed that the two-sample log rank statistic follows a Chi-squared distribution with 1 degree of freedom, x_1^2. In order to test the equality of default-free proportions between two groups, we simply compare the log rank statistic (2) with distribution of x_1^2 to obtain significance level (p-value) of the test.

The two-sample log rank statistic (2) can be extended to (*m*+1) samples, which tests whether the (m+1)-level factor is independent of the survival from default. Table 2 shows the data structure, at the time t_j, for the (*m*+1)-sample log rank statistic. Let $\mathbf{D}_j - \mathbf{E}_j = (D_{1j} - E_{1j}, \ldots\ldots, D_{kj} - E_{kj},)$ be the length-*m* vector of $D_{lj} - E_{lj}$ with $E_{lj} = E(D_{lj}) = D_j Y_{lj} / Y_j$, and $Var(\mathbf{D})$ be the $m \times m$ variance-covariance matrix for $\mathbf{D}$. The main diagonal items are variances of $D_{lj}(l = 1,2,\ldots,m)$ given by $Var(D_{lj}) = Y_{lj}(Y_j - Y_{lj})D_j(Y_j - D_j) / [Y_j^2(Y_j - 1)]$, and the off-diagonal items are the covariance for D_{lj} and $D_{l'j}$ $(l = 1,2,\ldots,m; l' = 1,2,\ldots,m;\ l \neq l')$ given by

$Cov(D_{lj}, D_{l'j}) = -Y_{lj}Y_{l'j}D_j(Y_j - D_j) / [Y_j^2(Y_j - 1)]$.

Then, the (m+1)-sample log rank statistic can be expressed by:

$$\left[\sum_{j=1}^{k}(\mathbf{D_j} - \mathbf{E_j})\right] Var^{-1}(\mathbf{D}) \left[\sum_{j=1}^{k}(\mathbf{D_j} - \mathbf{E_j})\right]^T \tag{3}$$

The (m+1) sample log rank statistic has a Chi-squared distribution with m degree of freedom. Hence, we can compare the log rank statistic (3) with the m degree of freedom Chi-squared distribution to obtain p-value of the test.

Regression Methods

Regression models for time-to-default include parametric, non-parametric and semi-parametric models. Parametric models assume a certain distribution function for default time T. Parametric models include, for example, Exponential, Weibull, and log-Normal models. Non-parametric models base on rank of the default time among covariate-defined groups to test effect of the covariate. Semi-parametric model include the famous Cox PH model, transformation models and the proportional odds model. In Cox PH model, the relationship between baseline and covariate-specific hazard is specified parametrically, while the baseline hazard function is unspecified. In this book chapter, we will focus on the commonly used Cox PH model and AFT models.

Table 2. (m+1)-sample log rant statistic data structure at time t_j

t	Z	Failures	No Failures	N
$t = t_j$	$Z = 0$	D_{0k}	$Y_{0k} - D_{0k}$	Y_{0k}
	$Z = 1$	D_{1k}	$Y_{1k} - D_{1k}$	Y_{1k}
	$Z = m$	D_{mk}	$Y_{mk} - D_{mk}$	Y_{mk}
Total		$\mathbf{D_j}$	$\mathbf{Y_j} - \mathbf{D_j}$	$\mathbf{Y_j}$

First, we need to define the covariate-specific hazard and survival functions since regression models always contain covariates. Let $S(t \mid \mathbf{z})$, $h(t \mid \mathbf{z})$ and $H(t \mid \mathbf{z})$ denote the covariate-specific survival function, hazard function and cumulative hazard function, respectively. Statistically, these quantities can be defined by:

$$S(t \mid \mathbf{z}) = \Pr(T > t \mid \mathbf{Z} = \mathbf{z}) = \int_t^{\infty} f(u \mid \mathbf{z}) du \ ,$$

$$h(t \mid \mathbf{z}) = \frac{f(t \mid z)}{S(t \mid z)} = d\,\frac{-\ln S(t \mid \mathbf{z})}{dt},$$

and

$$H(t \mid \mathbf{z}) = \int_0^t h(u \mid \mathbf{z}) du = -\ln[S(t \mid \mathbf{z})],$$

where $f(. \mid \mathbf{z})$ is the probability density function of default time T conditional on the covariates $\mathbf{z}$. Here, we assume that the survival and hazard functions are evaluated at $\mathbf{z}$.

When predictor variables, $\mathbf{Z}$, are not time-dependent, i.e., $\mathbf{Z}(t) = \mathbf{Z}$ for all t, the Cox PH model can be mathematically addressed by (Cox 1972):

$$h(t \mid \mathbf{z}) = h_0(t) \exp(\beta_1 z_1 + ... + \beta_p z_p) = h_0(t) \exp(\boldsymbol{\beta}^T \boldsymbol{Z}), \tag{4}$$

where $h_0(t)$ is an estimable unknown baseline function at time *t*, and $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)$ are the *p* regression coefficients of covariates $\mathbf{Z}$. In addition, $\mathbf{z} = (z_1, ..., z_p)$ is a realization of $\mathbf{Z}$. The linear combination $\beta_1 z_1 + ... + \beta_p z_p$ can be expressed as $\boldsymbol{\beta}^T \boldsymbol{Z}$ in vector format, where $\boldsymbol{\beta}^T$ is the transpose of vector $\boldsymbol{\beta}$. Regression coefficient β_j ($j = 1, \ldots, p$) can be interpreted as the logarithm of the hazard ratio (HR) for one unit increase in the value of the predictor variable Z_j when Z_j is continuous. If Z_j is a categorical variable, coefficient β_j would be interpreted as the log of HR comparing the level z_j with the reference level of Z_j. Please notice that the subscript index for a specific loan, *i*, is depressed here since a regression model should not be for an individual loan. Statistically, the regression coefficients $\boldsymbol{\beta}$ can be estimated by maximizing the partial likelihood (Cox 1975) in which the hazard function $h_0(t)$ was excluded. Breslow (1972) provide a full likelihood method to estimate the baseline hazard function $h_0(t)$ and the regression coefficients. Coefficients of Cox PH models can be obtained using popular commercial software such as SAS (procedure *PHREG*, SAS Institute, Inc., Cary, NC, USA) and STATA (*stcox* command, Stata Inc., College Station, TX, USA), and the free statistical package R (*coxph* command within *survival* library, www.r-project.org).

The key assumption for the Cox PH model is that hazard for levels defined by covariates are proportional for the entire study period. This is, sometime, a strong assumption, especially when the number of levels are large. The proportionality assumption can be checked by either plotting the survival function among levels of the covariate on a log-log scale, or by performing the proportionality test based on Schoenfeld residuals (Schoenfeld 1982). The SAS procedure PHREG has the option, namely *proportionality_test*, for testing proportionality among groups by creating interaction terms between survival time and covariates.

AFT model is an alternative to Cox PH regression. AFT models directly relate predictor variables to default time. Mathematically, the model is given by:

$$\log(t \mid \mathbf{z}) = -(\beta_1 z_1 + ... + \beta_p z_p) + \varepsilon = -\boldsymbol{\beta}^T \boldsymbol{Z} + \varepsilon\,, \tag{5}$$

where $\boldsymbol{\beta} = (\beta_1, ..., \beta_{p)})$ is the vector of regression parameters, and ϵ is the random error with unspecified distribution. The regression parameter $\beta_j (j = 1, ..., p)$ can be interpreted as the ratio of default time per unit change in covariate Z_j. Assume that the error term ϵ has a distribution function F_0 or a survival function S_0. The relationship between baseline survival functions S_0 and covariate-specific survival function, $S(t \mid \mathbf{z})$, is given by

$$S(t \mid \mathbf{z}) = S_0[t \cdot \exp(\boldsymbol{\beta}^T \boldsymbol{Z})] \tag{6}$$

Model (6) implies that covariates affect survival functions by changing its time scale. If the total covariate effect $\boldsymbol{\beta}^T \boldsymbol{Z} > 0$, then the failure rate is accelerated; on the other hand, when $\boldsymbol{\beta}^T \boldsymbol{Z} < 0$, the failure rate is decelerated; when $\boldsymbol{\beta}^T \boldsymbol{Z} = 0$, the covariate specific failure rates are the same as the baseline rate. The relationship between hazard and covariates can be easily derived, and is given by:

$$h(t \mid z) = h_0[t \cdot \exp(\boldsymbol{\beta}^T \boldsymbol{Z})] \cdot \exp(\boldsymbol{\beta}^T \boldsymbol{Z}) \tag{7}$$

Model (7) is analogous to the Cox PH model (4) but it has a time scale change from the baseline hazard h_0 to the covariate-specific hazard, h.

Comparison between Regression Methods

Different regression methods are based upon different assumptions of the relationship of hazard or survival functions among groups being compared. The Cox PH model assumes that the hazard rates of default among those groups are proportional, while the AFT model assumes that the default times among groups are linear on the log scale, or equivalently, the time to default is being accelerated. Thus, choice of regression models for default time should depend on natural of the data. Generally speaking, the Cox PH model is the standard method for survival analysis and can be used for time-to-event analysis unless the model assumption is not seriously violated. An AFT model can be selected if we have the evidence or prior knowledge that predictor variables accelerate the default time.

Difference between Survival Analysis and LR Modeling

Before the survival analyses model was introduced in analyzing credit risks, the most common approach is the LR method (Stepanova and Thomas, 2002). We can specifies a period of time (e.g., 24 months), and then fits a logistic model to historical customer data for predicting the probability that an applicant will default within the specified time period, as a function of the predictor variables for credit applicants. In survival analysis, we are interested in analyzing the default time and the hazard or survival function of it. In short, survival analysis aims to model the distribution of the time-to-default (or of some other event associated with default) while LR usually predict the probability of default within a single, specified period of time.

Competing Risks for the Time-to-Default Outcome

Although problem regarding competing risks in credit risk analysis were not intensively discussed in literature, it is always an issue when analyzing risks for loans. Thus, in this book chapter, we present more thoroughly on this issue.

Competing risks are said to be present when a loan is at risk of ending from more than one mutually exclusive events. In analyzing risk of default, several possible reasons may cause the end of a loan before its original intended term: (a) default, (b) early pay-off from borrower's own pocket, and (c) refinancing. One way to handle competing risks is to obtain the *cause-specific hazard*, which is defined as the instantaneous risk of ending a loan (before the originally specified loan terms) from a particular cause K ($K = 1,2,3$ for default, early pay-off from customers own pocket and refinancing, respectively. However, under the general setting of *k* competing risks, $K = 1,2,...,k$). Mathematically, the K^{th} cause-specific hazard at time *t*, $h_K(t)$, and survival function, $S_K(t)$, are given by (Prentice *et al*. 1978):

$$h_K(t) = \lim_{n\to\infty} \frac{Pr(t \le T \le t + \Delta t, K = k \mid T \ge t)}{\Delta t},$$

and

$$S_K(t) = \exp(-\int_0^t h_K(u)du)\,.$$

Under the cause-specific hazard, we are specifically interested in the failure due to the K^{th} reason. Hence, failures due to other reasons are considered as censored. Banasik *et al*. (1999) suggested that hazard for default and early pay-off can be estimated using this method. However, the cause-specific hazard does not have a direct interpretation in terms of survival probabilities for a specific type of failure (Fine & Gray1999). In another word, the cause-specific hazard, by itself, is interpretable but the corresponding survival function (derived from these hazard functions) is difficult to be described as a probability.

Another method to deal with competing risks is to estimate the *sub-distribution hazard* for cause *K*, which is defined as the instantaneous risk of ending a loan due to the K^{th} reason given that the loan has not ended due to cause K . We denote the sub-distribution hazard for cause *K* by $h_{KS}(t)$. Mathematically, the sub-distribution hazard for the K^{th} reason is given by:

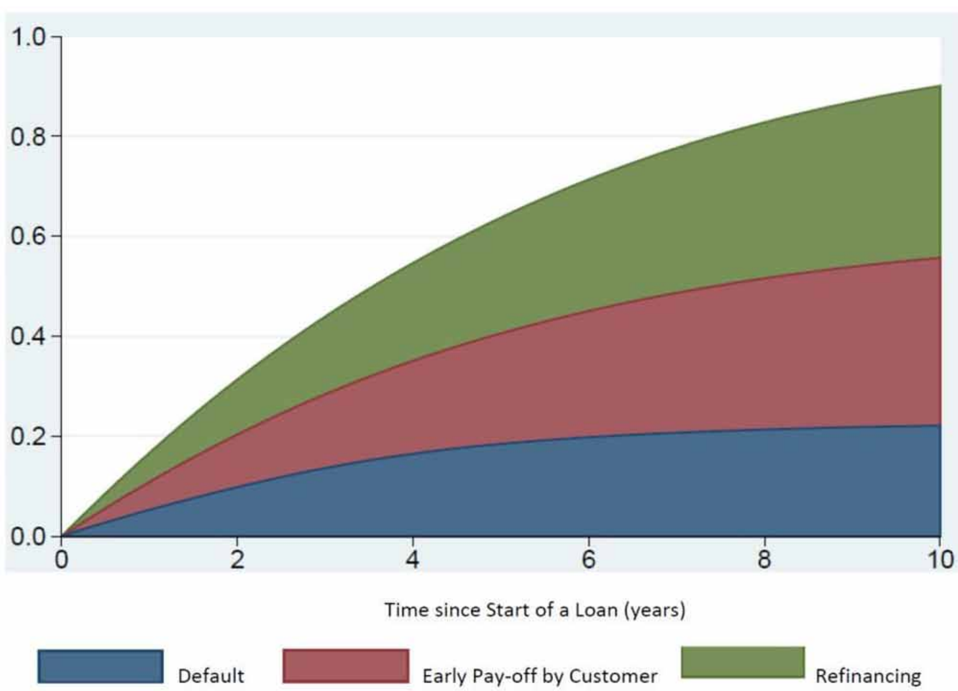
$$h_{KS}(t) = \lim_{\Delta t \to \infty} \frac{\Pr[t \le T < t + \Delta t \mid T > t \, or \, (T \le t \,\&\, Cause \ne K)]}{\Delta t}.$$

The cumulative incidence function (CIF) under this approach can be derived form h_{KS} and is interpreted as the proportion of loans that end early at time t due to the K^{th} reason accounting for the fact that loans can ends due to other reasons. The sub-distribution method provides a way of breaking down probabilities of different cause of ending a loan before the pre-specified term, and give the financial industry a clear indication of risks that they are facing. Using the sub-distribution method to deal with competing risks, we can estimate the sub-distribution CIFs for ending a loan due to default, early pay-off by borrowers, and refinancing, and then stack the CIFs together to estimate the overall probability of lending the loan. Figure 2 illustrates the way to stack the sub-distribution CIFs for the above 3 causes. The blue area denotes the CIF for loan ending due to default. The red and green areas represent CIFs due to early pay-off by customers and refinancing, respectively. Sum of these sub-CIFs is the probability of ending the loan, regardless of causes.

Based on the relationship between hazard and survival functions, Fine and Gray (1999) defined the sub-distributions for failure due to each of the k reasons (Putter *et al*. 2007). Motivated by the idea of the sub-distribution hazards, they further proposed a regression model that directly relates predictor variables, or covariates, to the cumulative incidence function. This model makes it possible to directly assess covariate effects on sub-distribution CIFs.

Competing risks analysis can be implemented in SAS, Stata and R. In SAS, the PHREG procedure can provide users with ratios of both cause-specific and sub-distribution hazards between groups, but

Figure 2. Stacked Cumulative Incidence Function for Early End of Loans Due to Different Reasons



PHREG cannot give cumulative hazard functions (or CIFs) for both event of interest and competing events. SAS macro *%CIF* can be implemented to estimate and compare CIFs. In Stata, the command *stcrreg* applies Fine & Gray's regression method to compute the CIFs. The *stcompadj* package calculates adjusted CIFs in the presence of competing risks by applying Cox PH model, and the *stpm2ci*f package can be utilized to obtain CIFs based on flexible parametric model (model that uses restricted cubic splines for baseline hazard function) for up to 10 causes of a failure. In R, the command *crr* within the *cmprsk* package can be utilized to compute the cumulative incidence function based on Fine & Gray's method.

The major difference between the cause-specific and the sub-distribution hazard is the risk set being used. The cause-specific hazard based on a risk set that decreases each time when the loan ends due to causes other than the one of interest. That is, a loan ending due to other causes will be considered as censored. On the other hand, with the sub-distribution hazard, loans ending due to other causes remain in the risk set and each loan has a censoring time that is greater than the time of failure due to all reasons.

Readers may ask this question: whether the cause-specific or the sub-distribution hazard/CIFs should be used for analyzing risk of loans? Our opinion is that it depends on analysis goals and specific settings of the dataset. For example, when default and refinancing are competing risks for ending a loan before its pre-specified term, we suggest that sub-distribution method be used because of the fact that customers who refinanced their loan could have defaulted if they did not refinance (they may refinance the loan with a better interest rate to relieve their financial hardship or just want to avoid the loaner). Hence, it is reasonable to assume that these customers stay in the risk set of default even after refinancing. However, when competing risks are default and early pay-off by customers, we prefer to use cause-specific hazard because customers who pay off early are less likely to default on their loans and would not stay in the risk set of default. To the best of our knowledge, model for sub-distribution hazards and CIFs has yet been used in analyzing credit risks. Due to the difficulty in interpreting the survival probability derived from cause-specific hazard, we suggest that the sub-distribution method be used to estimate credit risk due to default whenever occurrence of the competing events does not rule out the existence of loan with the events in the risk set of default.

EVALUATING CLASSIFICATION ACCURACY OF CREDIT SCORING SYSTEMS USING ROC ANALYSIS

In this section, we first define accuracy parameters of binary classification tools, and then extend to the evaluation of continuously-valued credit scoring systems using the ROC analysis. By evaluating the classification accuracies of these tools, banks and credit card companies can easily examine the losses and benefits for positive and negative test results based on their scoring tools, and can directly link the classification accuracy to important decision makings (Cornell *et al*. 2008).

Accuracy Parameters for Classification and Decision-Makings

Considering the evaluation of classification tools with dichotomous values. First of all, we define two important classification accuracy parameters, sensitivity and specificity, and two misclassification measures, false positive rate and false negative rate. Let DS denote the binary true default status such that $DS = 0$ indicates "no default" and $DS = 1$ denotes "default". Let Y be random variable of the classifier, such that $Y = 1$ denotes the positive test result for default, and $Y = 0$ indicates the negative test result. Here, a "positive test result" refers to as customer who is classified as high risk for default and a "negative test results" indicates a customer with low risk for default defined by *Y*. In reality, customers with positive test results are often refused for a loan. The *sensitivity* of the binary classifier Y is defined as the probability that a customer who defaults on a loan ($DS = 1$) has the positive test result ($Y = 1$). Mathematically, this probability can be expressed as:

$$Sensitivity = \Pr(Y = 1 \mid DS = 1),$$

where the symbol | denotes "conditional on", the statistical meaning of which can be found in introductory statistics books such as Wasserman (2004), Chap. 1. Sensitivity of a test is also known as the *true positive rate* (*TPR*) of the classification tool.

Another important accuracy parameter of Y is the *specificity*, which is defined as the probability of negative test results $(Y = 0)$ among loans with no default $(DS = 0)$. This probability is mathematically given by:

$$Specificity = \Pr(Y = 0 \mid DS = 0).$$

Specificity is often used interchangeably with *true negative rate* (*TNR*) in literature.

The binary classifier *Y* may also misclassify customers and leads to wrong credit-granting decisions. There are two types of misclassification rates. The first is the *false positive rate* (*FPR*), which is defined as the probability of a positive test result when default is absent. Mathematically,

$$FPR = \Pr(Y = 1 \mid DS = 0).$$

A false positive occurs when a "refusal of credit" decision is made to a potential borrower who would never default. By examining the definitions of *FPR* and *specificity*, we note that, mathematically, *FPR* = 1- *specificity*.

Another misclassification rate is the *false negative rate* (*FNR*), which is the probability of a negative test when default is present. This probability can be expressed by:

$$FNR = \Pr(Y = 0 \mid DS = 1).$$

A false negative occurs when a loan is granted to a borrower that later defaults on the loan. Also, we note that *FNR* = 1 - *sensitivity*.

Table 3 summarizes the aforementioned accuracy and misclassification rates using equations. The rows of this table are split by the true status of default $(DS = 1$ versus $DS = 0)$, and columns are classified by test results $(Y = 1$ versus $Y = 0)$. In each of the four cells defined by DS and Y, top rows display the frequency of loans within that cell, and bottom rows give the mathematical equation for calculating the corresponding accuracy or misclassification rate.

Table 3. Frequencies and mathematical equations of accuracies and misclassification rates for a binary classifier (Y)

Classifier True Status	Positive Test Result *(Y*=1*)*	Negative Test Result (*Y*=0)	Total
Default Present (*DS*=0)	a_0 $Sensitivity = a_0 \,/\, (a_0 + b_0)$	b_0 $FPR = b_0 \,/\, (a_0 + b_0)$	$a_0 + b_0$
Default Absent (*DS*=1)	a_1 $FNR = a_1 \,/\, (a_1 + b_1)$	b_1 $Specificity = b_1 \,/\, (a_1 + b_1)$	$a_1 + b_1$

ROC Curves

In this sub-section, we describe how to evaluate the accuracy of a continuously-valued classification tool for distinguishing customers who default on loans from those who never default. The credit scoring system is one example of those classification tools. For a credit scoring system, we cannot define its sensitivity and specificity without dichotomizing its probability distribution. A standard and popular tool for evaluating accuracy of the scoring system is the ROC curve, which has been studied or described in detail in literature such as Swets & Pickett (1982), Swets (1988), Hanley (1989), Begg (1991), Chock *et al*. (1997), Zhou, Obuchowski and McClish (2002), and Pepe (2003). To dichotomize the distribution of a continuously-valued classifier, we usually need to specify a *testing threshold* or *cut-point* upon which to base our decision. In this section, we illustrate how ROC curves are established for assessing performance of a credit scoring system. There are two dimensions of an ROC curve, and the first aspect is the true default status, also known as the gold standard (GS). The GS of default can usually be determined based on the bank's guidelines (for example, no payment for greater than 90 days). The GS for default should always be binary ("default" and "no default"). Value of the credit scoring system is another aspect of ROC curves. When a threshold value for the classifier needed to make a decision is not known *a priori*, an ROC curve can provide a useful description of its accuracies. The ROC curve plots sensitivities versus FPRs (or 1-specificities) of the classifier at each cut-point *c*. As the cut-point defining the positivity changes across all possible values of the scoring system, the ROC curve is plotted. An ROC curve tracks pairs of sensitivity and FPR at all possible threshold values, and theoretically, there are infinite number of cut-points from which we can choose to dichotomize. In order to mathematically define an ROC curve, we use Y to denote the continuous random variable of a credit scoring system with the convention that higher values of Y are correlated with higher chances of default. We further denote Y for "default" and "no default" by Y_1 and Y_0 respectively. Let DS denote the true default status such that $DS = 1$ indicates "default" and $DS = 0$ indicates "no default". When a cut-point c is used to dichotomize the distribution of Y_1 and Y_0, the sensitivity and FPR can be expressed given by:

$$Sensitivity = \Pr(Y > c \mid DS = 1) = \Pr(Y_1 > c) = S_1(c) = 1 - F_1(c),$$

and

$$FPR = \Pr(Y > c \mid DS = 0) = \Pr(Y_0 > c) = S_0(c) = 1 - F_0(c),$$

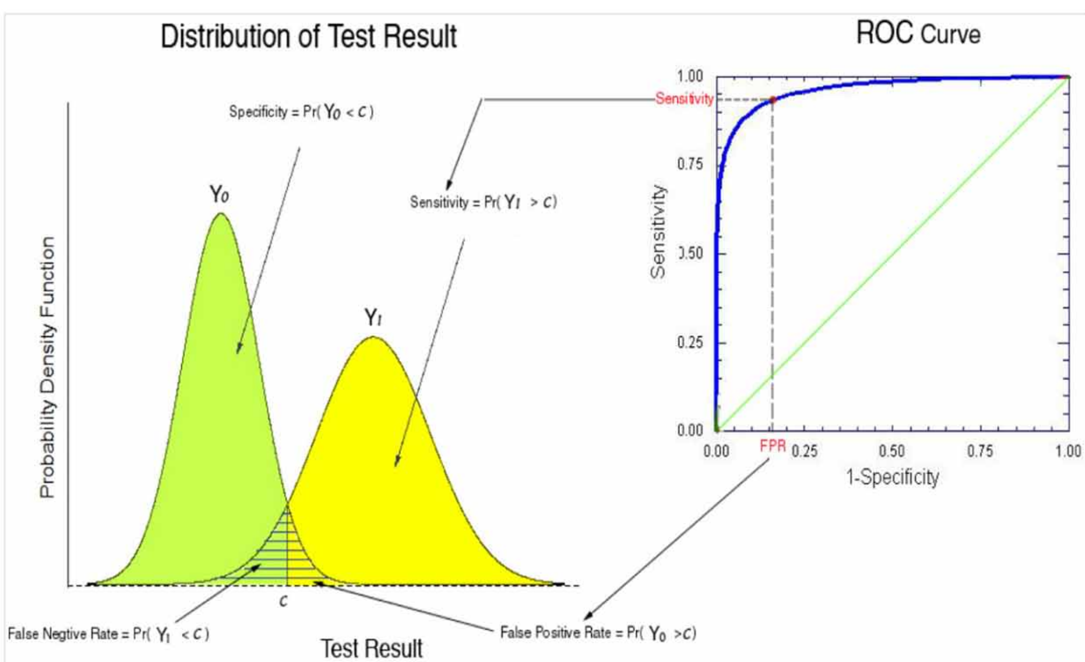
where $F_1(c)$ and $F_0(c)$ denote the distribution functions of Y_1 and Y_0, respectively. If we denote the *FPR* at cut-point c as $FPR(c) = S_0(c) = v$, then the ROC function is defined as the corresponding sensitivity at $FPR(c) = v$. Mathematically, this is a composite function of S_1 and S_0 in the following form:

$$ROC(v) = S_1(c) = S_1[S_0^{-1}(v)],$$

where $S_0^{-1}(.)$ denotes the inverse function of S_0 such that $S_0^{-1}(v) = c$ is equivalent to $S_0(c) = v$.

Figure 3 shows the process of making an ROC curve, and the relationship between the ROC curve and distributions of Y_1 and Y_0. In order to follow the convention that higher values of Y 's are correlated with a higher probability of default, we assume that Y is risk score for default (instead of the general credit score of which a higher value correlated with a lower risk of default). In this figure, the distribution for "non- default" (controls), Y_0, is in green and located to the left of the distribution for "default" (cases), Y_1, in yellow. There is an overlapping of the two distributions, represented by the shaded area. At a given threshold *c*, the sensitivity of the test can be visually represented by the area under the distribution of Y_1 from *c* to the maximal value of Y_1. This sensitivity is shown on the *y*-axis of the resultant ROC curve (the right panel of Figure 3). In addition, shaded area on the right side of *c* represents the probability that Y_0 is greater than *c*, which is the FPR of at cut-point *c* and shown on the *x*-axis of the ROC curve. As the cut-points change over an infinite number of all possible cut-points, a continuous ROC curve from point (0, 0) to (1, 1) within the unit square is obtained.

Figure 3. Distribution of test result for cases/controls and the corresponding ROC curve



ROC curves are simple and straightforward graphical tools that convey comprehensive information about classification accuracies. Because an ROC curve is a plot across all possible decision thresholds, it does not require the specification of threshold values to generate the curve (Zweig and Campbell, 1993). It is independent of the number of cases that occur among the whole population, and it is invariant with respect to monotone transformations that keep the rank of the test score for each individual. ROC curves are monotonic and non-decreasing. An ROC curve that is located closer to the upper and left boundary of the unit square indicates better performance of the test.

The area under the ROC curve (AUC) is an indicator of the overall accuracy performance for Y. Zhou, Obuchowski & McClish (2002) suggested that AUCs can be interpreted in the following three ways: (1) the average sensitivity over all possible values of specificity; (2) the average specificity across all possible values of sensitivity; and (3) the overall concordance of the test values with the gold standard. Specifically, the AUC equals to the probability that Y will assign a greater probability of being default than that of not being default. In this sense, the AUC is equivalent to the concordance *C-statistic* proposed by Harrell (2001). When evaluating the risks of loans, banks usually employ LR models to estimate risk scores. Then, the accuracy of the models is often measured by the AUC as a summary measure for the overall performance of predictive models (Miura *et al*. 2010). AUCs are often used to compare the performance among different classifiers. For example, when new predictor variables are identified for use in risk predictions, we often want to know whether there is a significant improvement in classification accuracies by adding these new risk factors into the current model. In this case, it is more appropriate to consider the increment in the AUC since the demonstration of a statistically significant association between new predictor variables and the gold standard (as many investigators presented) is usually inadequate.

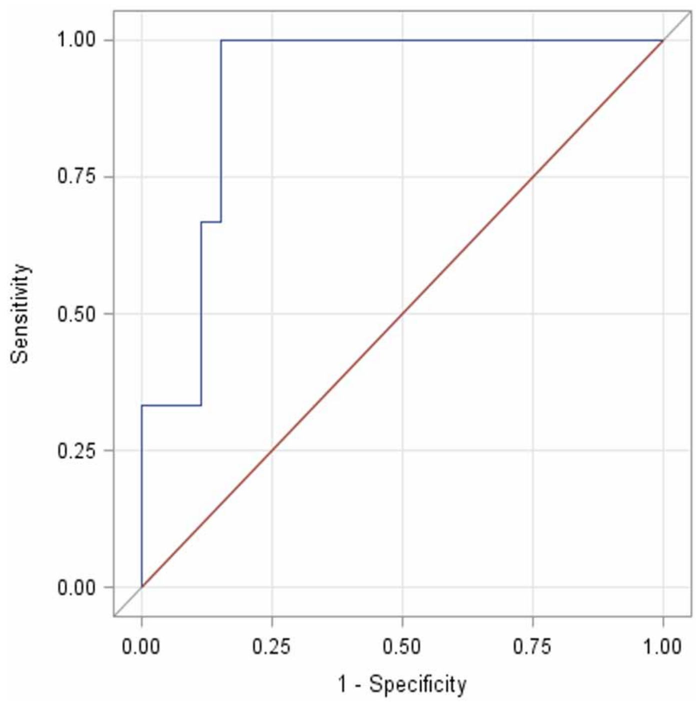
AUCs can be calculated either non-parametrically using an empirical distribution function or parametrically by assuming a bi-normal distribution of the test modality among cases and controls. The AUC for test modalities can be easily calculated using SAS and Stata. They can also be used to calculate the AUC for predicted values from fitting a LR model (PROC LOGISTIC in SAS and the *logistic* command in Stata). In SAS PROC LOGISTIC, the ROC statement can also generate ROC plots, estimated AUCs, and their standard errors (SEs). In addition, ROC curves from different predictive models that are fit to the same set of observations can be compared using ROCCONTRAST statements in the same SAS procedure.

Although the ROC analysis is an appealing method for assessing the accuracy of credit scoring models to predict defaults, it will fail when using logistic prediction models for rare default events. When events are rare, the corresponding ROC curve will be insensitive to the change of TPRs, thus providing litter information for evaluating the logistic prediction model (King and Zeng 2001). Figure 4 provides an example of ROC curves when defaults are rare. Since defaults are rare, ROC curve shown in Figure 4 is not smooth enough and the estimated sensitivities could have heavy jumps at some thresholds. Therefore, the estimates of both sensitivities and AUCs will not be accurate.

Extension to Time-Dependent ROC Analysis

When the binary true decision space is a function of time, the traditional ROC analysis introduced in the previously is not adequate. In evaluating a credit scoring system, customers' default status may change over the whole loan period. In order to accommodate this situation, Hu & Zhou (2014) suggested to use the time-dependent ROC method to evaluate the accuracies of credit scoring systems for predicting

Figure 4. An example of ROC curve when defaults are rare



the default over time. The time-dependent ROC approach gives a way to evaluate the performance of a credit scoring system (or a prediction model) to predict the customers' defaulting by a specific time, *t*, since the beginning of the loan. For instance, in order to assess the performance of a credit scoring system to predict default by 24 months, we can use the time-dependent ROC curves to obtain the overall accuracy (AUC), sensitivity and specificity at a certain cut-point of the scoring, or the sensitivity at a pre-specified specificity.

Time-dependent ROC analyses combine the traditional ROC methods for the binary decision space and the survival analysis. In time-dependent ROC analyses, the TPR and FPR are both dependent on time. The time-dependent TPR measures the probability of the estimated positives among the default cases up to a certain time point *t*, and the time-dependent TPR gives the probability of the estimated positives among those without defaulting by time *t*. Using the same notation as in previous sections, we can mathematically defined the time-dependent TPR and FPR for prediction defaults up to time *t* based on the prediction score *Y* at threshold value *c*:

$$TPR(c;t) = \Pr(Y > c \mid DS(t) = 1) = \Pr(Y > c \mid T \leq t)\,, \tag{8}$$

and

$$FPR(c;t) = \Pr(Y > c \mid DS(t) = 0) = \Pr(Y > c \mid T > t)\,. \tag{9}$$

Using the Bayes' Rule, the time dependent TPR (8) and FPR (9) can be further expressed as:

$$TPR(c;t) = \frac{\Pr(Y > c, T \le t)}{\Pr(T \le t)} = \frac{\Pr(T \le t \mid Y > c)\Pr(Y > c)}{\Pr(T \le t)}, \tag{10}$$

and

$$FPR(c;t) = \frac{\Pr(Y > c, T > t)}{\Pr(T > t)} = \frac{\Pr(T > t \mid Y > c)\Pr(Y > c)}{\Pr(T > t)}. \tag{11}$$

In equations (10) and (11), the conditional probability of defaulting time $\Pr(T \mid Y)$ can be estimated using methods provided in the survival analysis section and the probability of Y, $\Pr(Y)$, can be estimated using similar methods for the traditional ROC methods for the binary decision space. When estimating the covariate-specific time-dependent ROC curves or examining the covariate effects on time-dependent TRPs and FPRs, we can extend equations (10) and (11) by adding the covariates $\mathbf{Z}$. The corresponding TPRs and FPRs can be estimated based on the conditional probabilities, $\Pr(T \mid Y, \mathbf{Z})$ and $\Pr(Y \mid \mathbf{Z})$.

Hu & Zhou (2010) summarized the up-to-date statistical methods in time-dependent ROC analysis, such as Heagerty & Zheng (2005), Cai *et al*. (2006), and Song & Zhou (2008). Readers who are particularly interested in this issue are referred to their paper for an overview. To the best of our knowledge, calculation of time-dependent ROC curve is not available in commercial statistical software such as SAS and Stata. An R function for estimating the time-dependent ROC curves and AUCs using Song & Zhou (2008) method is available from authors of this chapter upon request.

CONCLUSION

Logistic regression models were traditionally used for examining and predict risk of default. With financial industry realizing that not only if but also when loans default is important for its profitability, survival (time-to-event) models was introduced to deal with the time-to-default data. Since then, several studies implement survival analysis on loan data and concluded that survival modeling performed better than logistic models when predicting defaults by a future time. The Cox PH model was the most commonly used regression method for analyzing time-to-default. Censoring and competing risks for default are common in loan data, but can be handled using survival analysis methods. Credit scorings and predictive models are quantitative tools used to predict and control credit risks. ROC curves are usually applied to evaluate the classification accuracy of these tools by plotting sensitivity versus FPR (1-specificity) across all possible thresholds over the range of the predictive scores. Area under the ROC curve (AUC) summarizes the overall performance of scoring systems. In presenting survival and ROC analysis to operating analysts of the financial industry, we suggest that these statistical method should be considered as fundamental tools in risk analyzing and controlling credit risk, especially the risk of default.

FUTURE RESEARCH DIRECTIONS

As we reviewed the literature regarding the topic of this book chapter, we found that few studies used or discussed competing risks analysis although defaults in loan data are often subject to competing risks. We hope that more envision that more studies competing risks of default will be published in the near future, especially research that compare the cause-specific and sub-distribution hazard of default using real world data or through simulation studies. With the rapid increase in credit scoring systems and risk predictive models, evaluation of these classification tools becomes a key element in analyzing and evaluating risks for the credit industry. However, yet, we have not seen studies assessing credit risk scorings using the time-dependent ROC analysis. As the timing of default is considered more and more important for the credit industry, future research may apply statistical methods of time-dependent ROC analysis (discussed in Hu & Zhou 2010) for assessing the credit scoring systems and predictive models for predicting defaults and the related events.

REFERENCES

Banasik, J., Crook, J. N., & Thomas, L. C. (1999). Not if but when loans default. *The Journal of the Operational Research Society*, *50*(12), 1185–1190. doi:10.1057/palgrave.jors.2600851

Basel Committee on Banking Supervision. (2000). *Principles for the Management of Credit Risk - final document*. Retrieved September 2000 from http://www.bis.org/publ/bcbs75.html

Begg, C. B. (1991). Advances in statistical methodology for diagnostic medecine ni the 1980's. *Statistics in Medicine*, *10*(12), 1887–1895. doi:10.1002/sim.4780101205 PMID:1805316

Breslow, N. E. (1972). Discussion following "Regression models and life tables" by D. R. Cox. *Journal of the Royal Statistical Society. Series B. Methodological*, *34*, 187–220.

Breslow, N. E., & Crowley, J. (1974). A large-sample study of the life table and product limit estimates under random censorship. *Annals of Statistics*, *2*(3), 437–454. doi:10.1214/aos/1176342705

Cai, T., Pepe, M. S., Lumley, T., & Jenny, N. S. (2006). The sensitivity and specificity of markers for event times. *Biostatistics (Oxford, England)*, *7*(2), 182–197. doi:10.1093/biostatistics/kxi047 PMID:16079162

Charles, T. (2004). *Risk and Financial Management: Mathematical and Computational Methods*. John Wiley & Son.

Chen, X., Wang, X., & Wu, D. D. (2010). Credit risk measurement and early warning of SMEs: An empirical study of listed SMEs in China. *Decision Support Systems*, *49*(3), 301–310. doi:10.1016/j.dss.2010.03.005

Chock, C., Irwig, L., Berry, G., & Glasziou, P. (1997). Comparing dichotomous screening tests when individuals negative on both tests are not verified. *Journal of Clinical Epidemiology*, *50*(11), 1211–1217. doi:10.1016/S0895-4356(97)00122-4 PMID:9393377

Cornell, J., Mulrow, C. D., & Localio, A. R. (2008). Diagnostic test accuracy and clinical decision making. *Annals of Internal Medicine*, *149*(12), 904–906. doi:10.7326/0003-4819-149-12-200812160-00011 PMID:19075211

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B. Methodological*, *34*, 187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, *62*(2), 269–276. doi:10.1093/biomet/62.2.269

Crouhy, M., Galai, D., & Mark, R. (2000). A comparative analysis of current credit risk models. *Journal of Banking & Finance*, *24*(1-2), 59–117. doi:10.1016/S0378-4266(99)00053-9

Duffie, D., & Singleton, K. J. (2003). *Credit Risk: Pricing, Measurement and Management* (1st ed.). Princeton, New Jersey: Princeton University Press.

Durand, D. (1941). *Risk Elements in Consumer Instalment Financing*. New York: National Bureau of Economic Research.

Fine, J. P., & Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, *94*(446), 496–509. doi:10.1080/01621459.1999.10474144

Francesca, G. (2012). A Discrete-Time Hazard Model for Loans: Some Evidence from Italian Banking System. *American Journal of Applied Sciences*, *9*(9), 1337–1346. doi:10.3844/ajassp.2012.1337.1346

Gupta, J., Wilson, N., Gregoriou, A., & Healy, J. (2014). The effect of internationalization on modeling credit risk for SMEs: Evidence from UK market. *Int. Fin. Markets*. *Inst. and Money*, *31*, 397–413. doi:10.1016/j.intfin.2014.05.001

Hanley, J. A. (1989). Receiver operating characteristic (ROC) methodology: The state of the art. *Critical Reviews in Diagnostic Imaging*, *29*(3), 307–335. PMID:2667567

Harrell, F. E. (2001). *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. Springer. doi:10.1007/978-1-4757-3462-1

Heagerty, P. J., & Zheng, Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics*, *98*, 409–417. PMID:15737082

Hu, N., & Zhou, X.-H. (2010). A review of time-dependent ROC curve for evaluating the prognosis capacity of biomarkers and semiparametric regression methods. *Proceeding of Joint Statistical Meeting*, Vancouver, Canada.

Hu, N., & Zhou, X.-H. (2014). *ROC analysis in business decision-making. Encyclopedia of Business Analytics and Optimization*. Hershey, PA: IGI Global.

Hubbard, D. W. (2009). *The Failure of Risk Management: Why It's Broken and How to Fix It*. Hoboken, New Jersey: John Wiley & Sons.

Im, J., Apley, D., Qi, C., & Shan, X. (2012). A time-dependent proportional hazards survival model for credit risk analysis. *The Journal of the Operational Research Society*, *63*(3), 306–321. doi:10.1057/jors.2011.34

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assn.*, *53*(282), 457–481. doi:10.1080/01621459.1958.10501452

King, G., & Zeng, L. (2001). Logistic Regression in Rare Events Data. *Political Analysis*, *9*(2), 137–163. doi:10.1093/oxfordjournals.pan.a004868

Lane, W. R., Looney, S. W., & Wansley, J. W. (1986). An application of the Cox's proportional hazards model to bank failure. *Journal of Banking & Finance*, *10*(4), 511–531. doi:10.1016/S0378-4266(86)80003-6

Leonard, K. J. (1993). Empirical Bayes analysis of the commercial loan evaluation process. *Statistics & Probability Letters*, *18*(4), 289–296. doi:10.1016/0167-7152(93)90017-D

Luoma, M., & Laitinen, E. (1991). Survival analysis as a tool for company failure prediction. *Omega-International Journal of Management Science*, *19*(6), 673–678. doi:10.1016/0305-0483(91)90015-L

Makowski, P. (1985). Credit scoring branches out. *Credit World*, *75*, 30–37.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its considerations. *Cancer Chemotherapy Reports. Part 1*, *50*(3), 163–170. PMID:5910392

Mavri, M., Angelis, V., Ioannou, G., Gaki, E., & Koufodontis, I. (2008). A two-stage dynamic credit scoring model, based on customers' profile and time horizon. *Journal of Financial Services Marketing*, *13*(1), 17–27. doi:10.1057/fsm.2008.2

McNeil, A. J., Frey, R., & Embrechts, P. (2005). *Quantitative Risk Management. Concepts, Techniques and Tools, Princeton Series in Finance*. Princeton, NJ: Princeton University Press.

Mehta, D. (1968). The formulation of credit policy models. *Management Science*, *15*(2), 30–50. doi:10.1287/mnsc.15.2.B30

Miura, K., Yamashita, S., Yamashita, S., & Eguchi, S. (2010). Area under the curve maximization method in credit scoring. *The Journal of Risk Model Validation*, *4*(2), 3–25.

Narain, B. (1992). Survival analysis and the credit granting decision. In Credit Scoring and Credit Control. Oxford.

Noh, P. J., Rohb, T. H., & Hana, I. (2005). Prognostic personal credit risk model considering censored information. *Expert Systems with Applications*, *28*(4), 753–762. doi:10.1016/j.eswa.2004.12.032

Orgler, Y. E. (1970). A credit scoring model for commercial loans. *Journal of Money, Credit and Banking*, *2*(Nov), 435–445. doi:10.2307/1991095

Pepe, M. S. (2003). *The statistical evaluation of medical tests for classification and prediction*. New York: Oxford University Press.

Prentice, R. L., Kalbfleisch, L. D., Peterson, A. V. Jr, Flournoy, N., Farewll, V. T., & Breslow, N. E. (1978). The analysis of failure time in the presence of competing risk. *Biometrics*, *34*(4), 541–554. doi:10.2307/2530374 PMID:373811

Putter, H., Fiocco, M., & Geskus, R. B. (2007). Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine*, *26*(11), 2389–2430. doi:10.1002/sim.2712 PMID:17031868

Ray, R. (2011). *Managing Financial Risk via Prediction Mar*kets. doi:10.2139/ssrn.1905097

Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika*, *69*(1), 239–241. doi:10.1093/biomet/69.1.239

Song, X., & Zhou, X.-H. (2008). A semiparametric approach for the covariate specific ROC curve with survival outcome. *Statistica Sinica*, *18*, 947–965.

Stepanova, M., & Thomas, L. C. (2002). Survival Analysis Methods for Personal Loan Data. *Operations Research*, *50*(2), 277–289. doi:10.1287/opre.50.2.277.426

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*(4857), 1285–1293. doi:10.1126/science.3287615 PMID:3287615

Swets, J. A., & Pickett, R. M. (1982). *Evaluation of diagnostic systems: Methods from signal detection theory*. New York: Academic Press.

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

Tong, E. N. C., Mues, C., & Thomas, L. C. (2012). Mixture cure models in credit scoring: If and when borrowers default. *European Journal of Operational Research*, *218*(1), 132–139. doi:10.1016/j.ejor.2011.10.007

Tsai, M., Lin, S., Cheng, C., & Lin, Y. (2009). The consumer loan default predicting model – An application of DEA–DA and neural network. *Expert Systems with Applications*, *36*(9), 11682–11690. doi:10.1016/j.eswa.2009.03.009

van Deventer, D. R., Imai, K., & Mesler, M. (2004). *Advanced Financial Risk Management: Tools and Techniques for Integrated Credit Risk and Interest Rate Risk Management*. Singapore: John Wiley.

Wasserman, L. (2004). *All of statistics: A concise course in statistical inference*. New York: Springer. doi:10.1007/978-0-387-21736-9

Wekesa, O., Samuel, M., & Mwita, P. (2012). Modelling Credit Risk for Personal Loans Using Product-Limit Estimator. *International Journal of Financial Research*, *3*(1), 22–22. doi:10.5430/ijfr.v3n1p22

Wiginton, J. C. (1980). A Note on the Comparison of Logit and Discriminant Models of Consumer Credit Behavior. *Journal of Financial and Quantitative Analysis*, *15*(3), 757–770. doi:10.2307/2330408

Witzany, J., Rychnovský, M., & Charamza, P. (2012). Survival Analysis in LGD Modeling. *European Financial and Accounting Journal*, *7*(1), 6–27.

Zhou, X.-H., Obuchowski, N. A., & McClish, D. K. (2002). *Statistical methods in diagnostic medicine*. New York: Wiley. doi:10.1002/9780470317082

Zweig, M., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, *39*(4), 561–577. PMID:8472349

KEY TERMS AND DEFINITIONS

AUC (C-Statistic): Is the area under an ROC curve that summarize the overall probability for correct classification.

Competing Risks: Are risks caused by competing events for a common survival outcome.

Covariate: Is a variable that is possibly predictive of the outcome under study.

Cumulative Incidence: Is the incidence rate (or risk rate) of an event cumulated from time zero to the time of interest.

Hazard: Is equivalent to the risk rate divided by the survival function of an event time.

Predictor Variable: Is a variable that can be used to predict the value of another variable.

ROC Curve: A curve that plots a diagnostic test's sensitivity versus its false positive rate across all possible threshold values for defining positivity.

Sensitivity: The probability that a diagnostic test can correctly identify a true case.

Specificity: The probability that a diagnostic test can correctly specify a non-case.

Survival Function: Is event-free probability at a given time. Mathematically, it is equivalent to 1 minus the cumulative incidence function.