XVII CONGRESSO anpcont
Universidade Presbiteriana Mackenzie
29 de Novembro a 01 de Dezembro de 2023
São Paulo-SP

200

Probability of Default for Lifetime Credit Loss for IFRS 9 Using MachineLearning Competing Risks Survival Analysis Models

Doutor/Ph.D. Cayan Atreio Portela Bárcena Saavedra ORCID iD[1], Doutor/Ph.D. Juliana Betini Fachini Gomes ORCID iD[2], Doutor/Ph.D. Eduardo Monteiro de Castro Gomes ORCID iD[2], Doutor/Ph.D. Herbert Kimura ORCID iD[1]

[1]University of Brasilia (UnB), Laboratory of Machine Learning in Finance and Organizations (LAMFO/UnB), Brasília, DF, Brazil. [2]University of Brasília, Statistics Department, Brasília, DF, Brazil

Doutor/Ph.D. Cayan Atreio Portela Bárcena Saavedra
0000-0002-6908-6733
Doutor/Ph.D. Juliana Betini Fachini Gomes
0000-0002-6677-4768
Doutor/Ph.D. Eduardo Monteiro de Castro Gomes
0000-0002-8948-9855
Doutor/Ph.D. Herbert Kimura
0000-0001-6772-1863

Resumo/Abstract

This study introduces a machine learning competing risks survival analysis model aiming at exploring the Probability of Default component of credit risk. Due to modelling of a cumulative probability of default over time, the model is applicable to assess Lifetime Expected Credit Loss under the International Financial Reporting Standard (IFRS) 9 regulation for financial institutions. Whilst most credit models focus on the default event itself, in many loan transactions, there is a competing event affecting risk: the possibility of the borrower prepay their debt before maturity. In this case, credit risk ceases to exist. We derive a statistical model that supports handling competing risks (credit risk and prepayment risk) in a machine learning survival analysis setup. As there is no available implemented computer package or library, we build the computational algorithm with subdistribution hazards using boosting as an ensemble method. Results of the model are generated using a dataset of credit card refinancing operations of a US financial institution. We observe, comparing different survival analysis techniques, that ComponentWise Gradient Boosting (CWGB) models showed better performance on both scenarios (subdistribution hazards and cause-specific models), closely followed by cause specific Cox Proportional Hazards, and that Gradient Boosting Survival was outperformed in all comparisons. The derived model is useful to address the guidelines of the IFRS 9 for credit risk, taking into account the context of lifetime credit exposure.

Modalidade/Type
Artigo Científico / Scientific Paper
Área Temática/Research Area
Contabilidade Financeira e Finanças (CFF) / Financial Accounting and Finance

Probability of Default for Lifetime Credit Loss for IFRS 9 Using Machine Learning Competing Risks Survival Analysis Models

ABSTRACT

This study introduces a machine learning competing risks survival analysis model aiming at exploring the Probability of Default component of credit risk. Due to modelling of a cumulative probability of default over time, the model is applicable to assess Lifetime Expected Credit Loss under the International Financial Reporting Standard (IFRS) 9 regulation for financial institutions. Whilst most credit models focus on the default event itself, in many loan transactions, there is a competing event affecting risk: the possibility of the borrower prepay their debt before maturity. In this case, credit risk ceases to exist. We derive a statistical model that supports handling competing risks (credit risk and prepayment risk) in a machine learning survival analysis setup. As there is no available implemented computer package or library, we build the computational algorithm with subdistribution hazards using boosting as an ensemble method. Results of the model are generated using a dataset of credit card refinancing operations of a US financial institution. We observe, comparing different survival analysis techniques, that ComponentWise Gradient Boosting (CWGB) models showed better performance on both scenarios (subdistribution hazards and cause-specific models), closely followed by cause specific Cox Proportional Hazards, and that Gradient Boosting Survival was outperformed in all comparisons. The derived model is useful to address the guidelines of the IFRS 9 for credit risk, taking into account the context of lifetime credit exposure.

KEYWORDS

accounting standards, IFRS 9, probability of default, credit risk, machine learning, survival analysis, competing risks, prepayment risk, lifetime expected credit loss

1. Introduction

The study of the occurrence of a specific credit event in a lifetime context has become more important over the past years. The lifetime expected credit loss (Lifetime ECL), introduced by the International Financial Reporting Standard 9 (IFRS 9), implied the development of new credit models. More particularly, the models should measure the present value of potential losses that could arise from the default on an obligation throughout the life of the loan (BIS, 2017). In our study, we focus on the Probability of Default (PD) component of credit risk, aiming at developing a model that could be applicable to IFRS 9.

In this context, Survival Analysis (SA) techniques naturally arise as first-to-go method, where the objective is to study the occurrence of a certain event during a period of time. From a credit risk perspective, the event of interest is default during the lifetime of the loan. Narain (1992) first introduced the use of survival analysis in credit scoring by estimating the probability of default in a 24-month loan dataset using an accelerated life exponential model. The author states that the use of estimated survival times supporting score ratings can improve credit-granting decision. The study of Narain (1992) paved the way to many others, with more advanced survival methods applied to credit scoring (Dirick, Claeskens, & Baesens, 2017).

Technological advancements and data availability have provided new tools and information to tackle the issue of estimation of the Lifetime ECL. Some of these developments relate to machine learning (ML) algorithms embedded in survival analysis models. These machine learning survival analysis mechanisms can be used in many applications, in several fields, for instance, to increase user retention (decreasing churn rate), to predict cross-selling opportunities (Harrison & Ansell, 2002), to leverage business strategy (Kauffman & Wang, 2001), to estimate the prediction of purchase of online games (Yang et al., 2019). The SA framework can also help financial institutions comply with regulators' guidelines in credit risk management, such as the IFRS 9 (BIS, 2017).

In addition to the advantage of allowing the estimation of a curve representing the risk of default through time, survival analysis can be used to model more than one event. By modelling mutually exclusive events as competing risks, survival analysis can be enhanced. From a loan credit risk context, the survival analysis adjusted by competing risks can, for instance, assess both default and prepayment events. Loans therefore can be analyzed by the event of default, which is the main concern in risk management, and also by the event of prepayment of the loan, situation in which credit risk ceases to exist. Both default and prepayment can be associated with losses.

Default is likely to be more severe due to potential losses in interest and principal value of the loan. However, prepayment also may bring losses, due to unearned interest on the remaining installments (Li, Li, Bellotti, & Yao, 2023). In addition, since prepayment may be unexpected, the cash surplus may be invested at lower interest rates. Hence, the modelling of competing risks can enhance the understanding of potential risk-adjusted performance of credit portfolios, enabling better expected profit strategies for financial institutions.

Competing risks models have already been investigated in studies concerning credit risk, for instance, with personal loan (Banasik, Crook, & Thomas, 1999; Stepanova & Thomas, 2002) and mortgage applications (Agarwal, Ambrose, & Liu, 2006; Deng, Quigley, & Van Order, 2000; Steinbuks, 2015; Thackham & Ma, 2022). In this study, we use a dataset of refinancing operations, which brings an interesting aspect due to borrower profile. Operations consist of borrowers who had already defaulted. Therefore, the study does not configure a traditional application of scoring model, but instead seek to contribute on understanding the context of debt renegotiation.

The approach adopted in past studies mainly focused on CoxPH (Cox, 1972) and its adaptation to a competing risk framework (Lunn & McNeil, 1995). However, we follow another approach by embedding a machine learning technique based on boosting into a competing survival analysis framework. Therefore, our study has two contributions: i) we focus our analysis on a dataset of renegotiated transactions of defaulted loans, and ii) we develop a novel ensemble model combining machine learning, more specifically, the boosting algorithm, within the context of competing risks in survival analysis. The competing risks relate to credit risk and to prepayment risk.

The study is structured as follows. In the next section, we discuss credit risk and regulatory standards that imply the use of survival analysis techniques. Then, we analyze the adjustments to embed a machine learning technique within the context of competing risks in survival analysis. We apply the derived algorithm in a dataset of renegotiated loans of credit card transactions of a US financial institution to empirically assess how the proposed model behaves, compared with other techniques. Finally, we conclude the study indicating implications and limitations and suggesting future research.

2. Related works

The financial industry is one field where survival analysis modelling approaches are specially useful, as they can provide additional information to support credit scoring decisions. The approach allows the analysis of time-to-event data, when there is an interest in the time to the occurrence of an event. In credit risk, the event of interest is the default, and the time-to-event would represent when a particular default will likely to happen.

Standard credit scoring methodologies express the probability of default in terms of a binary classification problem (Li et al., 2023). The borrower is classified as “good” or “bad” depending on the estimated probability of default and a given threshold. In survival analysis framework, credit analysis can be translated not just to “if” a borrower will default, but “when” a borrower will default (Banasik et al., 1999). The possibility of building a predictive model that takes into account the “if” and “when” questions naturally complies with international regulations, such as the IFRS 9.

International Financial Reporting Standard 9 (IFRS 9) was released in 2014 and became effective since 2018,substituting the International Accounting Standard 39 (IAS 39). IFRS 9 incorporated a forward-looking approach for loss allowances calculation. It requires financial institutions to adhere to this forward-looking perspective for expected loss impairment models.

The IFRS 9 indicates mechanisms to calculate provisions, by assessing ECL considering the entire time horizon of the financial instrument, and making adjustments in the Profit and Loss (P&L) account (Gornjak, 2020). The method involves checking whether there has been a significant increase in risk since the initial recognition.

Considering the relevance of identify not only if but also when a default can occur, according to Dirick et al. (2017) early studies explored survival analysis techniques in credit risk investigating parametric accelerated failure time (AFT) survival methods or non-parametric baseline approach based on Cox proportional hazards (CoxPH) model (e.g. Banasik et al. (1999); Bellotti and Crook (2009); Cao (2009); Narain (1992); Stepanova and Thomas (2002, 2001); Zhang and Thomas (2012)). Other studies introduced mixture cure mechanisms (e.g. Dirick, Claeskens, and Baesens (2015); Tong, Mues, and Thomas (2012) in survival analysis.

In particular, (Dirick et al., 2017)) using credit datasets from European banks, compare results of different configurations of traditional survival techniques based on AFT and CoxPH and mixture cure models. The study identify that models with single event mixture cure and spline adjustment in the hazard function lead to better credit scoring.

More recently, machine learning algorithms begin to be incorporated in survival analysis. For instance, adapting a boosting algorithm to Cox models, Binder, Allignol, Schumacher, and Beyersmann (2009) develop CoxBoost and Y. Chen, Jia, Mercola, and Xie (2013) build the GBMCI (gradient boosting machine for concordance index) (Bai, Zheng, & Shen, 2021). Although many applications aimed at applications in medicine and health, credit risk emerges naturally as an area to explore machine learning with survival analysis, due the characteristics of the probability of default within a given period of time.

One example of study on credit is from Bai et al. (2021) that propose a nonparametric ensemble tree model (GBST) coupling survival tree models with a gradient boosting algorithm. Using two different large datasets, the results suggest that the GBST leads to better classification metrics when compared with other machine learning survival models such as random survival forest (Ishwaran, Kogalur, Blackstone, & Lauer, 2008), CoxBoost (Y. Chen et al., 2013), conditional inference survival forest (CIF) model (Wright, Dankowski, & Ziegler, 2017), and DeepHit based on deep neural networks (Lee, Zame, Yoon, & der Schaar, 2018).

Finally, in survival analysis, competing risks are relevant, since there may be events that preclude the event of interest from happening (Geskus, 2015). For instance, in medicine the focus of the study could be related to the risk of an individual getting cancer with a specific treatment, death is a competing risk. In credit analysis, prepayment can be a competing risk of default (Li et al., 2023). As Schuster, Hoogendijk, Kok, Twisk, and Heymans (2020) suggest, biased results can emerge when survival data is analyzed without taking into account competing risks.

However, even though competing events should are relevant, is a less well-known element of survival analysis (Schuster et al., 2020). In this context, although some papers explore survival analysis and competing risks in credit (e.g., Agarwal et al. (2006); Banasik et al. (1999); Li et al. (2023); Stepanova and Thomas (2002)), there are fewer studies that embed machine learning techniques (e.g. XXX)

Considering refinancing operations, many studies focus on mortgages or home equity line of credit (HELOC). Tracy and Wright (2016) applied Cox competing risk models to investigate how mortgage payment reduction from the Home Affordable Refinance Program (HARP) affects the probability that the borrower defaults after having refinanced. The authors suggests that refinancing can have a positive impact in loss mitigation.

J. Chen, Xiang, Yang, et al. (2018) analyze the risk of re-default on Federal Housing Administration (FHA) modified loans. Authors finds suggests that modified loans are more likely to default compared to identical loans with no modifications. There has been few studies on the assessment of expected losses on refinancing operations of usual credit lines. In our study, we focus on credit card refinancing exploring a boosting approach embedded in survival analysis techniques with competing risks.

3. Machine Learning Survival Analysis for Competing Risks

In this study, we apply machine learning survival analysis models that takes into account competing risks. We incorporate a boosting mechanism to assess competing risks in a survival analysis setting. Although in this study we apply the method to credit risk and prepayment risk, the proposed model is suitable for any two competing risks.

In our study, we aim at analysing credit risk loans that have two relevant elements: (i) the borrower default can occur in any moment until maturity and (ii) the borrower can prepay the loan in any moment until maturity. In the occurrence of any of the two events, potential credit risk ceases to exist, as the risk of not complying with the loan has been realized with the default or there is no credit risk anymore, as the loan was fully paid in advance.

3.1. Survival Analysis

Changes resulting from the new regulation implied adaptations in the estimation of PD, which is one of the most important risk component in credit risk analysis (Vaněk & Hampel, 2017). More specifically, the need to calculate Lifetime ECL requires a method to analyze credit risk not only on a given time, for instance, at maturity or after a year, but also throughout all the period of the loan.

In this context, Survival Analysis methodology can be considered a feasible and appealing approach, since it allows to tackle the default problem from a different perspective. SA models allows to assess whether as well as when a default will occur (Banasik et al., 1999).

The focus of SA methods is on the time T until an event occurs (e.g., default). Observations that did not experience the specified event are called censored observations. Usually SA data are represented by a pair of random variables (T, C). In the absence of competing risks, the censoring variable C takes the value 1 if the event of interest was observed or 0 if it is a censored observation. When $C = 1$, T refers to the time of occurrence of the event of interest and when $C = 0$, T refers to the time at which the observation was censored.

The function $S(t)$ represents the probability of not having experienced a given event until time t (i.e., the probability to survive until time t) and is given by:

$$S(t) = P(T > t) \quad (1)$$

Therefore, the cumulative distribution is defined as the probability of an observation do not survive until time T, that is, $F(t) = 1 - S(t)$ (Colosimo & Giolo, 2006) and the probability density function is $f(u) = -\frac{d}{du}S(u)$ (Dirick et al., 2017).

Additionally, the hazard function, which represents the instantaneous risk, is expressed as:

$$h(t) = \lim_{\delta_t \to 0} \frac{P(t \leq T < t + \delta_t \mid T \geq t)}{\delta_t} = \frac{f(t)}{S(t)} \quad (2)$$

which can also be written in terms of the survival function (**??**) and the probability density function (**??**). From these equations, the cumulative hazard function can be defined as:

$$H(t) = \int_0^t h(u)du = \int_0^t \frac{f(u)}{S(u)} du = \int_0^t \frac{d\{1 - S(u)\}}{S(u)} dt = -\log\{S(t)\} \quad (3)$$

Since $S(t) = exp-H(t)$, there is an one-to-one correspondence between the hazard rate $h(t)$ and the cumulative risk distribution $F(t)$ (Andersen & Keiding, 2012).

3.2. Cox Proportional Hazard

The Cox Proportional-Hazards model (Cox, 1972) allows to incorporate covariables information into a censored regression model and is one of the most traditional approaches based on time-to-event techniques. It consists of a semi-parametric model for the model is composed by two components: a non-parametric base hazard λ_0 and a parametric component $g(X\beta)$. The parametric component is usually used as $g(X\beta) = \exp(X\beta)$ (Colosimo & Giolo, 2006). Therefore, the model is given by:

$$\lambda(t) = \lambda_0(t) \exp(X\beta) \quad (4)$$

where $\boldsymbol{x}_i$ is a vector of observed data and $\boldsymbol{\beta}$ is a $p \times 1$ vector of parameters for each covariable. Proportional-hazards comes from the assumption that the ratio of failure rates among two individuals is constant over time. For instance, considering $\lambda_i(t)$ and $\lambda_j(t)$ representing the failure rate of two individuals at time t, it follows that:

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \frac{\lambda_0(t) \exp(x_i\beta)}{\lambda_0(t) \exp(x_j\beta)} = \exp(x_i\beta - x_j\beta), \quad (5)$$

where the ratio of failure rate is constant and independent of time. For parameter estimation, Cox (1972, 1975) proposed a partial likelihood without the semi-parametric component. The partial likelihood is a product of all terms associated to different failure times, i.e.:

$$L(\beta) = \prod_{i=1}^{n} \left(\frac{\exp(X\beta)}{\sum_{j \in R(t_i)} \exp(X\beta)} \right)^{\delta_i} \quad (6)$$

, where δ_i is the censoring indicator, taking value $\delta_i = 1$, if an event is observed and $\delta_i = 0$, in case of censoring. The risk set $R(t_i)$ is composed by individuals who have not yet failed until time t_i. Values of β that maximize the partial likelihood function are obtained by solving the system defined by $U(\beta) = 0$, where $U(\beta)$ is the score vector of first-order derivatives of $l(\beta) = \log(L(\beta))$.

3.3. Competing Risks

The approach based on competing risks is adequate when there are two mutually exclusive events, i.e., the occurrence of one event implies the non-occurrence of the other. The (T, C) can be extended to $C = \{0, 1, 2, ..., k\}$ where $k \geq 2$ types of events are possible. When competing risks are present, the Cumulative Incidence Function (CIF) represents the probability of occurrence of a specific type of event before time t. Considering j competing events, the CIF for cause j is defined as (Frydman & Matuszyk, 2022):

$$F_j(t) = P(T \leq t, C = j) = \int_0^t P(T = t, C = j)du = \int_0^t f_j(u)du \quad (7)$$

Thus, the probability that any event takes place before time t, is the sum of all j CIF:

$$F(t) = P(T \leq t) = \sum_{j=1}^{k} P(T \leq t, C = j) = \sum_{j=1}^{k} F_j(t) \quad (8)$$

When dealing with competing risks in Cox-PH regression models, there are two main methods for estimating CIF (Austin, Steyerberg, & Putter, 2021): (i) modeling the cause-specific hazard by considering each event separately and combining the models to estimate CIF (Kalbfleisch & Prentice, 2011), and (ii) modeling the Fine-Gray (Fine & Gray, 1999) subdistribution hazard function, which enables a direct way for modelling the effect of covariates. Each method have a defined hazard function for a specific event type: the cause-specific hazard function (9) and the subdistribution hazard function (9) (Austin & Fine, 2017):

$$h_j^{cs}(t) = \lim_{\Delta_t \to 0} \frac{P(t \leq T < t + \Delta_t, C = j \mid T \geq t)}{\Delta_t} \quad (9)$$

$$h_j^{sd}(t) = \lim_{\Delta_t \to 0} \frac{P(t \leq T < t + \Delta_t, C = j \mid T \geq t \cup (T < t \cap C \neq j))}{\Delta_t} \quad (10)$$

The cause-specific hazard function is the instantaneous risk of event j in individuals who have not experienced any type of event until time t. The subdistribution hazard function, is the risk of event j considering individuals who have not experienced the specific j event until time t (Austin & Fine, 2017). In this sense, the subdistribution hazard function proposed by Fine and Gray (1999) take into account individuals who have not experienced the primary event of interest, but, have experienced a competing event.

Austin, Lee, and Fine (2016) suggest that, on one hand, subdistribution hazard models are better suited for clinical prediction models and risk-scoring systems, where there is a natural interest in estimating the absolute incidence of the primary event. On the other hand, cause-specific hazard models are more suitable when the objective is to assess epidemiological questions of etiology (Austin et al., 2016). Furthermore, the former estimates cause-specific hazard functions for each competing event to derives de CIF from there, while the latter allows directly estimate the CIF Frydman and Matuszyk (2022).

Fine and Gray (1999) proposes an adaptation for Cox partial likelihood, by changing the risk set R_j and adding weights w_j. The adapted likelihood is given by:

$$L(\beta) = \prod_{i=1} \frac{\exp(x_i\beta)}{\sum_{j\in R_i} w_{ij}\exp(x_j\beta)} \tag{11}$$

where R_i consists of observations that did not experience the primary event, even if they have experienced a competing risk event. The risk set is defined as (Pintilie, 2006):

$$R_j(m) = [j; T_j \geq m \quad \text{or} \quad (T_j \leq m \text{ and the subject has experienced a competing risk event})]. \tag{12}$$

Additionally, the observations on the risk set are weighted by:

$$w_{ij} = \frac{\hat{G}(t_i)}{\hat{G}(\min(t_i, t_j))} \tag{13}$$

where $\hat{G}$ is the Kaplan-Meier estimate of the survivor function of the censoring distribution (Pintilie, 2006). The weight goes to zero as the distance between the time point t_i and the time recorded for the competing risk event increases. Thus, observations that experienced a competing risk event do not participate fully in the likelihood (Pintilie, 2006).

3.4. Boosting algorithm

Boosting is an ensemble method that sequentially fits models to the data, in which each subsequent model places more emphasis on the observations that were misclassified by the previous models (Freund & Schapire, 1997). J. H. Friedman (2001) proposes Gradient Boosting Machine (GBM), a boosting framework that generalizes loss functions for regression problems. The GBM algorithm is depicted in *Enum*1 (Ridgeway, 1999):

J. Friedman, Hastie, and Tibshirani (2000) connects boosting with well-known statistical principles (e.g. additive modeling and maximum likelihood) and demonstrates how the method relates to algorithms used for fitting linear models, such as IRLS (Iteratively Reweighted Least Squares).

Ridgeway (1999) builds a generalization of boosting algorithms for the exponential family and proportional hazards regression models. The proposed generalization is based on Fisher scoring (Fine & Gray, 1999), a variant of the Newton-Raphson optimizer. The author illustrates

Initialize $\hat{F}(x) = \min_\rho \sum_{i=1}^{n} \Psi(y_i, \rho)$
For m in $1, \cdots, M$ do

1. Compute the negative gradient as the working response

$$z_i = -\frac{\partial}{\partial F(x_i)} \Psi(y_i, \rho) \Big|_{F(x_i)=\hat{F}(x_i)}$$

2. Fit a regression model on z_i given covariates x_i

3. Choose a gradient descent step

$$\rho = \min \Psi(y_i, \hat{F}(x_i) + \rho f(x_i))$$

4. Update $F(x)$ estimate as

$$\hat{F}(x) \leftarrow \hat{F}(x) + \rho f(x)$$

Enum 1: J. H. Friedman (2001) Gradient Boost algorithm

adaptations of the algorithm for generalized linear model under a framework proposed by Nelder and Wedderburn (1972).

For proportional hazards regression models, the illustration is made by allowing likelihood based loss functions in Friedman's gradient boosting machine. Thus, we can make use of Cox partial likelihood for fitting censored data. Considering the ideas described above, and that boosting fits nonlinear regression models (Ridgeway, 1999), *Enum*2 can be adapted for censored data by searching $F(x)$ to maximize Cox's log-partial likelihood (Ridgeway, 1999), by replacing $\Psi(y, F)$ with the $-\log PL(F \mid t, \delta, x)$, where:

$$\log PL(F \mid t, \delta, x) = \sum_{i=1}^{N} \delta_i \left(F(x_i) - \log \left(\sum_{j=1}^{N} I(t_j \geq t_i) e^{F(x_j)} \right) \right) \tag{14}$$

Therefore, the negative gradient is given by:

$$z_i = \delta_i - \sum_{j=1}^{N} \delta_j I(t_i \geq t_j) \frac{e^{\hat{F}(x_j)}}{\sum_{k=1}^{N} I(t_k \geq t_j) e^{\hat{F}(x_k)}} \tag{15}$$

Following the same steps proposed in *Enum*1, the algorithm for boosting CoxPH for censored data model is illustrated in *Enum*2 (Ridgeway, 1999):

Gradient Boosting methods can also operate as a regularization framework (Bühlmann & Hothorn, 2007). The core idea relies on a stepwise optimization of a function $F(.)$ in function space, by minimizing a loss function (Binder & Schumacher, 2008). This approach has been user for survival context, using Cox negative partial log-likelihood as loss function (Binder & Schumacher, 2008).

Initialize $\hat{F}(x) = \min_\rho \sum_{i=1}^{n} \Psi(y_i, \rho)$

For m in $1, \cdots, M$ do

1. Compute the negative gradient as the working response

$$z_i = \delta_i - \sum_{j=1}^{N} \delta_j I(t_i \geq t_j) \frac{e^{\hat{F}(x_j)}}{\sum_{k=1}^{N} I(t_k \geq t_j) e^{\hat{F}(x_k)}}$$

2. Fit a regression model on z_i given covariates x_i

3. Choose a gradient descent step

$$\rho = \min \Psi(y_i, \hat{F}(x_i) + \rho f(x_i))$$

4. Update $F(x)$ estimate as

$$\hat{F}(x) \leftarrow \hat{F}(x) + \rho f(x)$$

Enum 2: J. H. Friedman (2001) Gradient Boost algorithm

With componentwise least squares as base learner, in each m step, the negative gradient of the loss function is evaluated for the current estimate $F_m(x; \hat{\beta}_m)$ (Binder & Schumacher, 2008). For each predictor variable, a simple linear regression is fitted to the gradient. Then, the coefficient of the predictor variable with the smallest sum of squares is updated. This can lead to many of the estimated coefficients being zero, resulting in sparse fits resembling Lasso-like approaches (Binder & Schumacher, 2008).

3.5. Boosting algorithm with competing risks

Considering Ridgeway (1999), a natural way to incorporate competing risks into a boosting framework is to replace $\Psi(y, F)$ with an adapted log-partial likelihood derived from (11). Binder et al. (2009) proposed a competing risk boosting framework for high-dimensional data for fitting proportional sub-distribution hazards models. The study involves a context in which the number of covariates is greater than the number of observations, and a sparse vector of estimated parameters is desirable. With this, the authors implement a componentwise boosting approach using penalized maximum partial likelihood, incorporating previous boosting steps as an offset. Another adaptation occurs in a definition of sets of mandatory and optional covariates. Before each boosting step, parameters referring to the mandatory covariates are updated simultaneously by one maximum partial likelihood Newton–Raphson. In each boosting step, only one parameter corresponding to the optional covariates is updated.

Applications in credit risk scoring generally involve models with greater degrees of freedom on parameters, with the number of covariates being smaller than the number of observations. In this way, we proceed without the penalty term and restrictions on the covariates. Therefore, we can incorporate information of competing risks by considering the adapted likelihood proposed by Fine and Gray (1999). Hence, we wish to maximize the following log-partial likelihood:

$$\log FG(F \mid t, \delta, x) = \sum_{i=1} \delta_i \left(F(x_i) - \log \left(\sum_{j \in R_i} w_{ij} e^{F(x_j)} \right) \right) \quad (16)$$

where $j \in R_i$ if ($t_j \geq t_i$) or ($t_j \leq t_i$ and individual j experienced a competing risk event). Taking the derivative with respect to $F(x_i)$, leads to:

$$\frac{\partial}{\partial F(x_i)} = \sum_{i=1} \delta_i \left[1 - \frac{w_{ij} e^{F(x_i)}}{\sum_{j \in R_i} w_{ji} e^{F(x_j)}} \right] \quad (17)$$

Similar to (15), the the negative gradient computed as the working is responses is given by:

$$z_i = \delta_i - \sum_{j=1}^{N} \delta_j I(i \in R_j) \frac{w_{ij} e^{\hat{F}(x_i)}}{\sum_{k \in R_j} w_{kj} e^{\hat{F}(x_k)}} \quad (18)$$

The final boosting algorithm that allows to incorporate information on secondary events is given in *Enum*3 by:

Initialize $\hat{F}(x) = \min_\rho \sum_{i=1}^{n} \Psi(y_i, \rho)$

For m in $1, \cdots, M$ do

1. Compute the negative gradient as the working response

$$z_i = \delta_i - \sum_{j=1}^{N} \delta_j I(i \in R_i) \frac{w_{ij} e^{\hat{F}(x_i)}}{\sum_{k \in R_j} w_{kj} e^{\hat{F}(x_k)}}$$

where $w_{ij} = \frac{\hat{G}(t_i)}{\hat{G}(\min(t_i, t_j))}$

2. Fit a regression model on z_i given covariates x_i

3. Choose a gradient descent step

$$\rho = \min \Psi(y_i, \hat{F}(x_i) + \rho f(x_i))$$

4. Update $F(x)$ estimate as

$$\hat{F}(x) \leftarrow \hat{F}(x) + \rho f(x)$$

Enum 3: Boosting algorithm for Fine-Gray adapted likelihood

With a similar approach, Binder et al. (2009) fits subdistribution hazards models under a boosting framework for sparse data. The authors adapted a componentwise likelihood-based boosting (Binder & Schumacher, 2008) for competing risks.

Boosting has been shown to improve the prediction accuracy of survival analysis models, particularly for high-dimensional data (Mayr, Binder, Gefeller, & Schmid, 2014). However, the interpretability of the GBM can be challenging, as it combines many weak learners to make the final prediction. Therefore, it is important to carefully consider the trade-off between model accuracy and interpretability when using boosting in survival analysis.

4. Data and Method

This study analyzes data consisting of credit card refinancing operations of a US financial institution. Therefore, in contrast to traditional credit scoring applications, this research explores a different profile of borrowers. Instead of measuring the likelihood of a borrower default on a new credit operation, we model the probability of default in refinancing borrowers who had already been delinquent in their credit card debt some time in the past. As a result, our research advances knowledge of the credit risk phenomena in the context of a different borrower profile, i.e., a previous defaulter.

Due to confidentiality and strategic issues, we have access to refinancing operations that span from January 2014 to December 2015. Therefore, the outdated database precludes the disclose of recent information, such as default rate, but allows the identification of outcomes, based on real-world data, of the use of machine learning models embedded in survival analysis techniques.

Dataset consists of 118,967 operations with a time-maturitie of 36 months. We selected operations starting until 2014 and 2015 (102,538) to train the model and separate those starting in 2016 (10,230) for an out-of-time evaluation. In order to reduce computational time we selected a 10% sample of operations that started in 2014 and 2015, leaving the final train dataset with 10,230 observations.

Dataset used during modeling considered the following information variables:

(1) Loan amount: the listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.
(2) Interest rate: interest rate of the loan.
(3) Installment: the monthly payment owed by the borrower.
(4) Employment length: employment length in years, ranging from zero to ten, where zero means less than one year and ten means ten or more years.
(5) Home ownership: the home ownership status provided by the borrower during registration or obtained from the credit report (rent, own or mortgage).
(6) Annual income: the self-reported annual income provided by the borrower during registration.
(7) Verification status: indicates whether income was verified or not.
(8) Dti: A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.
(9) Total acc: The total number of credit lines currently in the borrower's credit file.
(10) Earliest credit line: time since borrower's earliest reported credit line was opened.
(11) Loan percentage to income: a ratio computed as the loan amount on the annual income, reflecting the share of commitment of income with the loan.
(12) Time to default or repayment (in months).

(13) Default status: binary variable with 1 (default) or 0 (non default).
(14) Repayment status: binary variable with 1 (default) or 0 (non default).

5. Results

In this section we compare results provided by fitted models. We consider loss functions based on cause-specific (CS) and subdistribution hazard (SH) models. For CS models the secondary risk (early payment event) is assumed to be censored. For SH models, individuals with prepayment event before time t remain in the risk set with an associated weight.

In credit risk context, ignoring the competing risk event of prepayment results in upwardly-biased estimate of the cumulative probability of default (Frydman & Matuszyk, 2022). In this way, we first evidence the importance of competing risk modeling by comparing each survival model as if early repayment was not considered as secondary risk.

Figure 1 shows, for each model, the estimated curve of cumulative probability of default from a hypothetical renegotiation with 12% interest rate, assigned "rent" regarding ownership status and taking median values on all other covariates. For cause-specific models this is the Cumulative Incidence Function, as for the other models, it is represented by 1-predicted survival function. This shows that cause-specific models leads to a lower curve of cumulative probability of default with the same predictive power.

We evaluate predictive performance on both a test dataset and an out-of-date dataset. For performance comparison we compute three metrics commonly used to assess goodness of fit on survival models. The Concordance Index (Harrell, Califf, Pryor, Lee, & Rosati, 1982), which measure a rank correlation between estimated risks and observed times. The Integrated Brier Score (IBS), showing accuracy risk predictions over time. The Dynamic AUC providing a measure of calibration over time, by distinguishing observations who fail by time $t_i \leq t$ from those failing after time $t_i > t$.

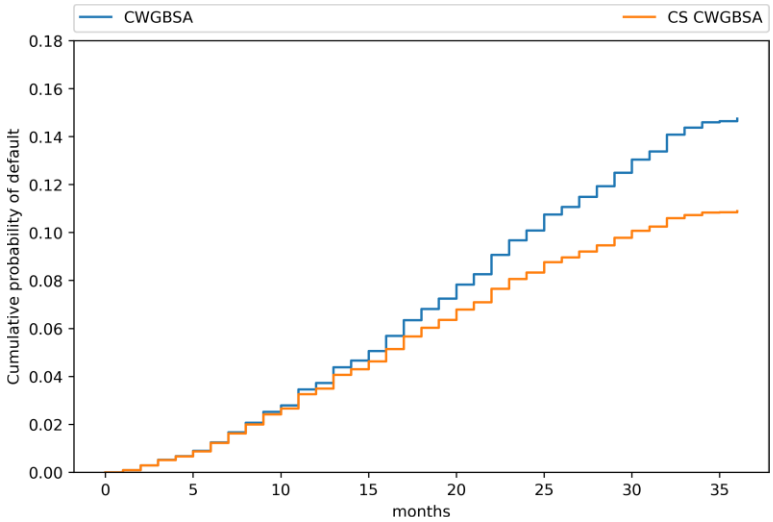
Table 1 displays out-of-sample and out-of-time results. Models with a ComponentWise Gradient Boosting approach showed the best results, closely followed by Cox-PH. Gradient Boosting Survival Analysis was outperformed and every comparison.

Table 1. Results

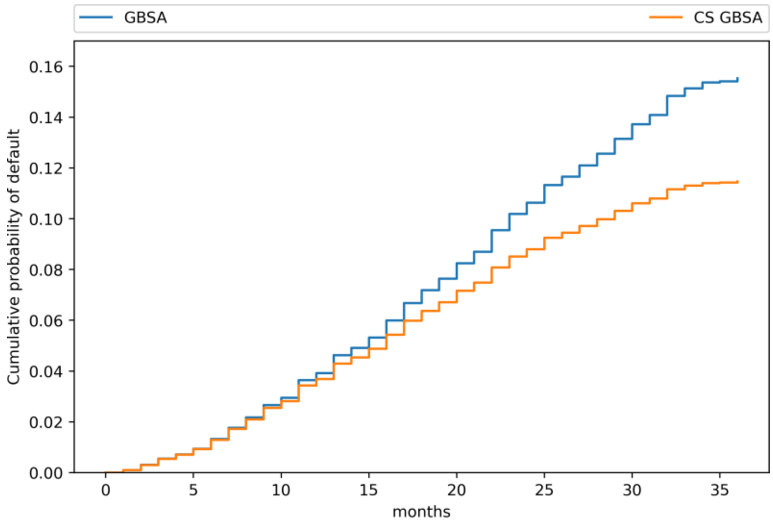
Model	Out of sample			Out of time		
	C-Index	IBS	AUC	C-Index	IBS	AUC
SH CWGB	0.6462	0.0665	0.6607	0.6873	0.0961	0.6975
CS CWGB	0.6447	0.0664	0.6602	0.6887	0.0968	0.7009
CS CoxPH	0.6441	0.0661	0.6574	0.6711	0.0958	0.6789
CS GBSA	0.6313	0.0666	0.6412	0.6536	0.0964	0.6615

In an out-of-sample test, adaptation for subdistribution hazards showed slightly better results than cause-specific, with a C-index of 0.6462 (HS) over 0.6447 (CS), closely followed by CoxPH (0.6441) and GBSA (0.6313). IBS showed the same behavior, demonstrating good calibration performance for all models. Out-of-sample dynamic AUC (Figure 2) for CWGB also show competitive numbers with similar values for SH (0.6607) and CS (0.6602), both approaches presenting better results than CoxPH (0.6574) and GBSA (0.6412).

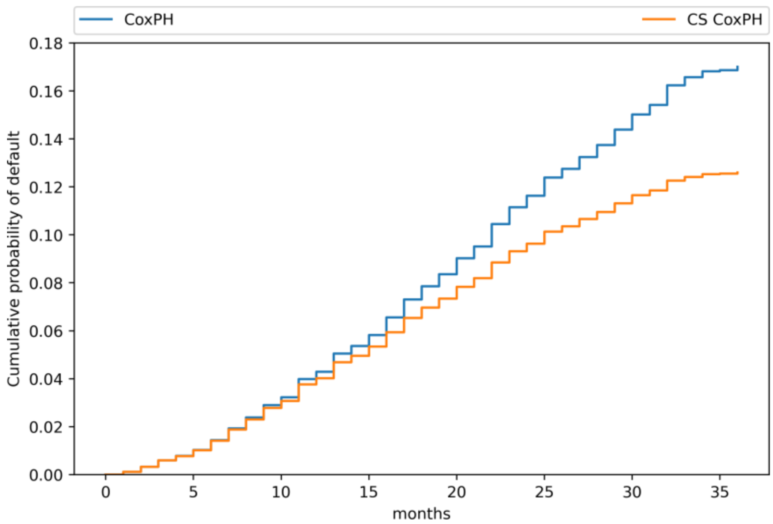
In operations starting in 2015 a reversed behavior is seen among CWGB approaches, with CS with slightly higher values than SH. However, both models outperformed Cox-PH and GBSA



(a)



(b)



(c)

Figure 1. Cumulative probability of default comparison when ignoring prepayment event for (a) CWGBSA (b) GBSA and (c) Cox-PH.

on all metrics considered. Out-of-time dynamic AUC (Figure 3) shoes greater disparitie with CWGB (0.709 CS and 0.6795 HS) over CoxPH (0.6789) and GBSA (0.6615).

In general, we observe that ComponentWise Gradient Boosting models showed better performance on both scenarios, closely followed cause specific Cox Proportional Hazards, and Gradient Boosting Survival was outperformed in all comparisons. The loss function adaptation to subdistribution hazards on CWGB showed comparative performance.

Aside from predictive power, it is also interesting to analyze default prediction. This can be achieved by comparing the cumulative incidence functions (CIFs), which provides an idea of the probability of failure (Pintilie, 2006). We analyze predicted curves for loans with 7.5% (Figure 4), 10% (Figure 5) and 12% (Figure 6) interest rate, with home ownership status assigned as “rent” ad ‘mortgage” and taking the median value on all other covariates. A higher curve is estimated by SH CWGBSA in all scenarios. While SH CWGBSA presents the same cumulative curve for both status of home ownership (keeping interest rate constant), it appears to be sensitive to interest rate level, with significant increase on the cumulative probability as higher rates are considered. This could be a reflection of the penalized approach lasso-like, leading to a prediction made by few covariates. Cause-specific models provides lower estimated curves and more mixed behavior of different interest rates and home ownership status. CS CWGBSA and CS GBSA presents a similar behaviuor, providing higher curves for “mortgage” when interest rate is 7% , and with a decreasing impact of home ownerhsip as higher rates are considered (with CS GBSA curve higher then CS CWGBSA in all scenarios). For CS CoxPH higher curves are observed for “rent” than for “mortgage” status, specially for higher interest rates. For instance, with 12% interest CS Cox-PH has the lowest curve for “mortgage” and the second highest for “rent”.

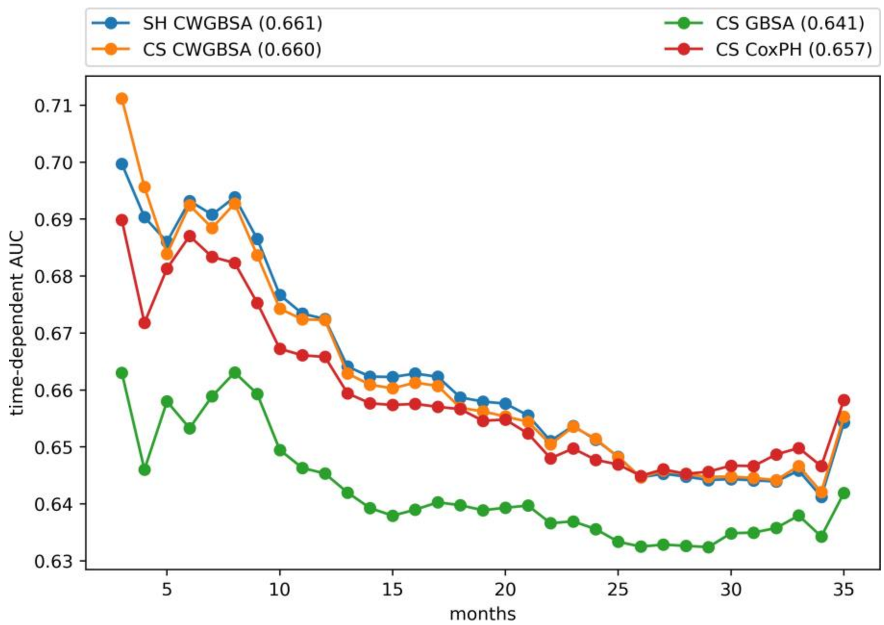


Figure 2. Out-of-sample cumulative Dynamic AUC

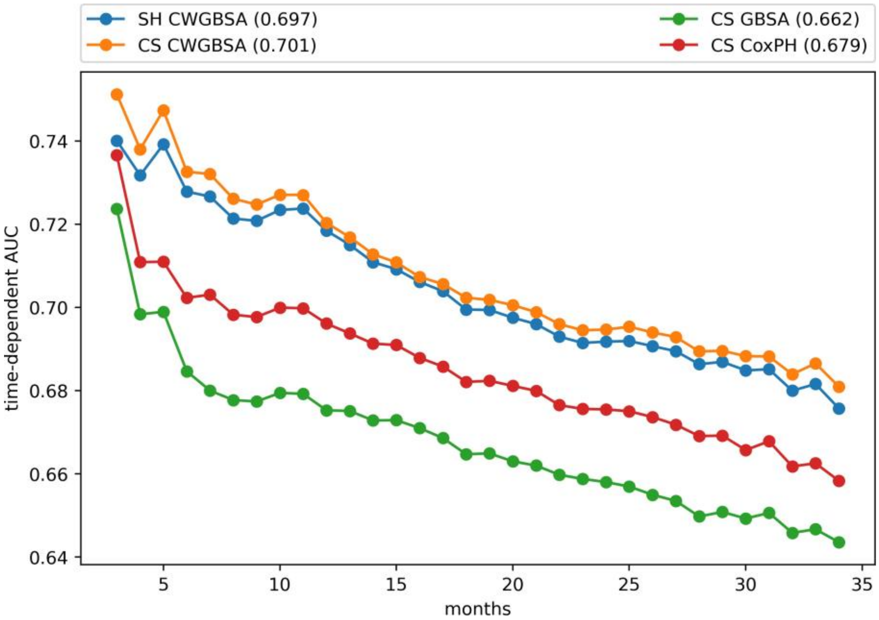
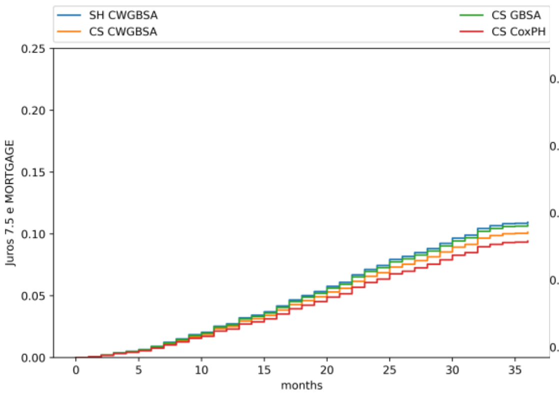
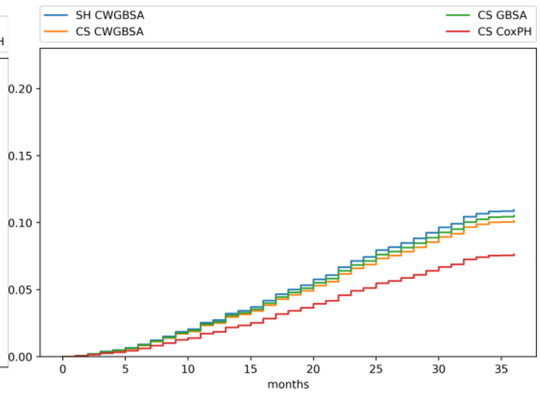


Figure 3. Out-of-time cumulative Dynamic AUC

Figure 4. Cumulative Probability of default for an operation with interest rate of 7% and home ownership assigned as (a) Rent and (b) Mortgage



(a)



(b)

Figure 5. Cumulative Probability of default for an operation with interest rate of 10% with home ownership assigned as (a) Rent and (b) Mortgage

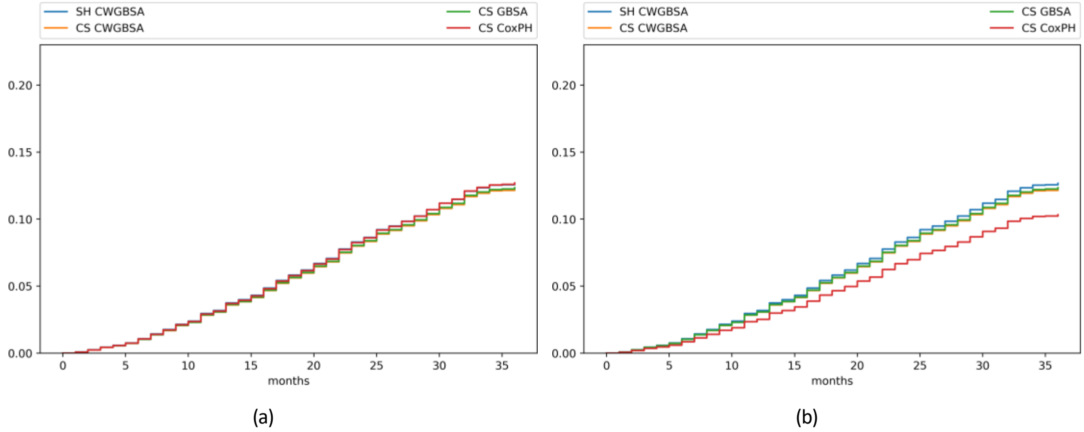
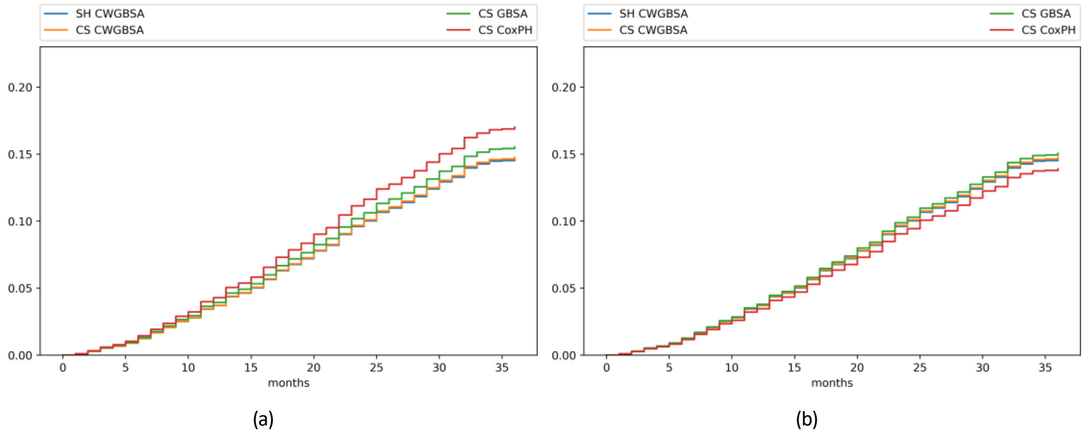


Figure 6. Cumulative Probability of default for an operation with interest rate of 12% and home ownership assigned as (a) Rent and (b) Mortgage



6. Conclusion

In this paper, we introduced a method to assess competing risks in credit portfolios, embedding machine learning techniques within a survival analysis framework. More specifically, the paper contributes to the discussion of lifetime expected credit loss required by the IFRS 9.

We explore a competing risk approach with subdistribution hazards under a boosting framework using component wise least squares as base estimators. Focusing on credit analysis, differently from studies applying a similar framework in other areas (Binder et al., 2009), we used non penalized loss functions, since it is uncommon to have more covariables than observations in loans context.

We derive the statistical properties and define a computational algorithm of the ensemble boosting mechanism adapted to incorporate computing risks (credit and prepayment risks). We also computationally implement the machine learning competing risks survival analysis model, as there is no available preprogrammed package or library.

We showed that, in a dataset of refinancing operations of credit card loans of a US financial institution, adapting the loss function to include competing risks during estimation on the CIF leads to comparable results in relation to outputs of cause-specific models.

Future studies could test different base learners in the boosting framework as well as investigate comparative results in other credit datasets. In our study, we focused on one component of credit risk: Probability of Default (PD). However, the implementation of techniques, compatible with IFRS 9, aiming at modelling the other components (Loss Given Default - LGD and Exposure at Default EAD) over time, could also be relevant. Finally, other studies could explore bias and fairness in models based on machine learning techniques applied to competing risks in credit assessment.

References

Agarwal, S., Ambrose, B. W., & Liu, C. (2006). Credit lines and credit utilization. *Journal of Money, Credit and Banking*, 1–22.

Andersen, P. K., & Keiding, N. (2012). Interpretability and importance of functionals in competing risks and multistate models. *Statistics in medicine*, *31*(11-12), 1074–1088.

Austin, P. C., & Fine, J. P. (2017). Practical recommendations for reporting f ine-g ray model analyses for competing risk data. *Statistics in medicine*, *36*(27), 4391–4400.

Austin, P. C., Lee, D. S., & Fine, J. P. (2016). Introduction to the analysis of survival data in the presence of competing risks. *Circulation*, *133*(6), 601–609.

Austin, P. C., Steyerberg, E. W., & Putter, H. (2021). Fine-gray subdistribution hazard models to simultaneously estimate the absolute risk of different event types: cumulative total failure probability may exceed 1. *Statistics in Medicine*, *40*(19), 4200–4212.

Bai, M., Zheng, Y., & Shen, Y. (2021, June). Gradient boosting survival tree with applications in credit scoring. *Journal of the Operational Research Society*, *73*(1), 39–55. Retrieved from https://doi.org/10.1080/01605682.2021.1919035

Banasik, J., Crook, J. N., & Thomas, L. C. (1999). Not if but when will borrowers default. *Journal of the Operational Research Society*, *50*(12), 1185–1190.

Bellotti, T., & Crook, J. (2009, December). Credit scoring with macroeconomic variables using survival analysis. *Journal of the Operational Research Society*, *60*(12), 1699–1707. Retrieved from https://doi.org/10.1057/jors.2008.130

Binder, H., Allignol, A., Schumacher, M., & Beyersmann, J. (2009). Boosting for high-dimensional time-to-event data with competing risks. *Bioinformatics*, *25*(7), 890–896.

Binder, H., & Schumacher, M. (2008). Allowing for mandatory covariates in boosting estimation of sparse high-dimensional survival models. *BMC bioinformatics*, *9*, 1–10.

BIS. (2017). *Ifrs 9 and expected loss provisioning - executive summary* (Report No. FSI Connect). Bank of International Settlements.
Bühlmann, P., & Hothorn, T. (2007). Boosting algorithms: Regularization, prediction and model fitting.
Cao, V. J. M. D. A., Ricardo. (2009). Modelling consumer credit risk via survival analysis. *SORT*, *33*(1), 3-30. Retrieved from http://eudml.org/doc/43039
Chen, J., Xiang, J., Yang, T. T., et al. (2018). Re-default risk of modified mortgages. *International Real Estate Review*, *21*(1), 1–40.
Chen, Y., Jia, Z., Mercola, D., & Xie, X. (2013). A gradient boosting algorithm for survival analysis via direct optimization of concordance index. *Computational and Mathematical Methods in Medicine*, *2013*, 1–8. Retrieved from https://doi.org/10.1155/2013/873595
Colosimo, E. A., & Giolo, S. R. (2006). *Análise de sobrevivência aplicada*. Editora Blucher.
Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187–202.
Cox, D. R. (1975). Partial likelihood. *Biometrika*, *62*(2), 269–276.
Deng, Y., Quigley, J. M., & Van Order, R. (2000). Mortgage terminations, heterogeneity and the exercise of mortgage options. *Econometrica*, *68*(2), 275–307.
Dirick, L., Claeskens, G., & Baesens, B. (2015, March). An akaike information criterion for multiple event mixture cure models. *European Journal of Operational Research*, *241*(2), 449–457. Retrieved from https://doi.org/10.1016/j.ejor.2014.08.038
Dirick, L., Claeskens, G., & Baesens, B. (2017). Time to default in credit scoring using survival analysis: a benchmark study. *Journal of the Operational Research Society*, *68*(6), 652–665.
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.
Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, *55*(1), 119–139.
Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics*, *28*(2), 337–407.
Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232.
Frydman, H., & Matuszyk, A. (2022). Random survival forest for competing credit risks. *Journal of the Operational Research Society*, *73*(1), 15–25.
Geskus, R. B. (2015). *Data analysis with competing risks and intermediate states* (Vol. 82). CRC Press.
Gornjak, M. (2020). Literature review of ifrs 9 and its key parameters. *Management*, *20*, 22.
Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L., & Rosati, R. A. (1982). Evaluating the yield of medical tests. *Jama*, *247*(18), 2543–2546.
Harrison, T., & Ansell, J. (2002). Customer retention in the insurance industry: Using survival analysis to predict cross-selling opportunities. *Journal of Financial Services Marketing*, *6*(3), 229–239.
Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008, September). Random survival forests. *The Annals of Applied Statistics*, *2*(3). Retrieved from https://doi.org/10.1214/08-aoas169
Kalbfleisch, J. D., & Prentice, R. L. (2011). *The statistical analysis of failure time data*. John Wiley & Sons.
Kauffman, R. J., & Wang, B. (2001). The success and failure of dotcoms: A multi-method survival analysis. In *Proceedings of the 6th informs conference on information systems and technology (cist).*
Lee, C., Zame, W., Yoon, J., & der Schaar, M. V. (2018, April). DeepHit: A deep learning approach to survival analysis with competing risks. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*(1). Retrieved from https://doi.org/10.1609/aaai.v32i1.11842
Li, Z., Li, A., Bellotti, A., & Yao, X. (2023, April). The profitability of online loans: A competing risks analysis on default and prepayment. *European Journal of Operational Research*, *306*(2), 968–985. Retrieved from https://doi.org/10.1016/j.ejor.2022.08.013
Lunn, M., & McNeil, D. (1995). Applying cox regression to competing risks. *Biometrics*, 524–532.
Mayr, A., Binder, H., Gefeller, O., & Schmid, M. (2014). The evolution of boosting algorithms. *Methods of information in medicine*, *53*(06), 419–427.
Narain, B. (1992). Survival analysis and the credit granting decision. *Credit scoring and credit control*,

109, 121.

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, *135*(3), 370–384.

Pintilie, M. (2006). *Competing risks: a practical perspective*. John Wiley & Sons.

Ridgeway, G. (1999). The state of boosting. *Computing science and statistics*, 172–181.

Schuster, N. A., Hoogendijk, E. O., Kok, A. A., Twisk, J. W., & Heymans, M. W. (2020, June). Ignoring competing events in the analysis of survival data may lead to biased results: a nonmathematical illustration of competing risk analysis. *Journal of Clinical Epidemiology*, *122*, 42–48. Retrieved from https://doi.org/10.1016/j.jclinepi.2020.03.004

Steinbuks, J. (2015). Effects of prepayment regulations on termination of subprime mortgages. *Journal of Banking & Finance*, *59*, 445–456.

Stepanova, M., & Thomas, L. (2002). Survival analysis methods for personal loan data. *Operations Research*, *50*(2), 277–289.

Stepanova, M., & Thomas, L. C. (2001, September). PHAB scores: proportional hazards analysis behavioural scores. *Journal of the Operational Research Society*, *52*(9), 1007–1016. Retrieved from https://doi.org/10.1057/palgrave.jors.2601189

Thackham, M., & Ma, J. (2022). On maximum likelihood estimation of competing risks using the cause-specific semi-parametric cox model with time-varying covariates–an application to credit risk. *Journal of the Operational Research Society*, *73*(1), 5–14.

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

Tracy, J., & Wright, J. (2016). Payment changes and default risk: The impact of refinancing on expected credit losses. *Journal of urban Economics*, *93*, 60–70.

Vaněk, T., & Hampel, D. (2017). The probability of default under ifrs 9: Multi-period estimation and macroeconomic forecast. *Acta universitatis agriculturae et silviculturae mendelianae brunensis*.

Wright, M. N., Dankowski, T., & Ziegler, A. (2017, January). Unbiased split variable selection for random survival forests using maximally selected rank statistics. *Statistics in Medicine*, *36*(8), 1272–1284. Retrieved from https://doi.org/10.1002/sim.7212

Yang, W., Huang, T., Zeng, J., Tang, Y., Chen, L., Mishra, S., & Liu, Y. E. (2019). Purchase prediction in free online games via survival analysis. In *2019 ieee international conference on big data (big data)* (pp. 4444–4449).

Zhang, J., & Thomas, L. C. (2012, January). Comparisons of linear regression and survival analysis using single and mixture distributions approaches in modelling LGD. *International Journal of Forecasting*, *28*(1), 204–215. Retrieved from https://doi.org/10.1016/j.ijforecast.2010.06.002