



Article

Modeling the Probability of Default Term Structure Using Different Methodologies Under IFRS 9

Kgotso Rudolf Moremoholo, Sandile Charles Shongwe *  and Frans Frederick Koning † 

Department of Mathematical Statistics and Actuarial Science, Faculty of Natural and Agricultural Sciences, University of the Free State, Bloemfontein 9301, South Africa; rkmoremoholo@hotmail.com (K.R.M.); koningf@ufs.ac.za (F.F.K.)

* Correspondence: shongwesc@ufs.ac.za

† Fellow of the Actuarial Society of South Africa (FASSA) & Chartered Enterprise Risk Actuary (CERA).

Abstract

To mitigate credit risk, banks are required to set aside a specific amount as a safety net to absorb the expected loss on a banks' loan portfolio called loan loss provisions (LLPs) or provisions for bad debts. All banks worldwide had to adopt International Financial Reporting Standard 9 (IFRS 9) as the financial reporting standard. Unlike its predecessor (i.e., International Accounting Standard 39, IAS 39), IFRS 9 accelerates the recognition of losses by requiring provisions to cover both already-incurred losses and some losses expected in the future by calculating the expected credit loss (ECL). To evaluate if the obligor's credit quality has deteriorated, the IFRS 9 standard requires banks to compare the obligor's probability of default (PD) at the inception phase of the loan and at the reporting date. Thus, three methodologies are explored in this study (i.e., Cox proportional hazard (PH), Extended Cox PH, and Random Boosting Forest (RBF)) for computation of the PD term structures using Kaplan–Meier as the benchmark model under IFRS 9. The purpose of this research is to illustrate the application of three methodologies on the publicly available mortgage loan portfolio from Freddie Mac using different measures of goodness-of-fit and the predictive accuracy measure, i.e., the Concordance index (C-index). The comparison analysis reveals that the extended Cox PH and RBF models provide better predictive accuracy (higher C-index) but at the cost of increased complexity and potential overfitting (higher information criteria). However, Cox PH has shown the most efficient fit, and offers a stable and understandable hazard trajectory. Finally, for reproducibility, the SAS and R codes are included to illustrate how each of the results (in form of a table or figure) were obtained.

Keywords: Akaike Information Criterion (AIC); Bayesian Information Criterion (BIC); Conditional Akaike Information Criterion (CAIC); Concordance index (C-index); Cox proportional hazard; credit risk; expected credit loss; International Financial Reporting Standard 9 (IFRS 9); loan loss provisions; probability of default



Academic Editor: Indranil SenGupta

Received: 12 December 2025

Revised: 28 January 2026

Accepted: 2 February 2026

Published: 3 March 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction and Literature Review

To mitigate credit risk, financial institutions are obligated to hold regulatory capital to address unexpected losses, which are unpredictable and may occur due to extreme or adverse circumstances. For expected losses, financial institutions are required to set aside a specific amount as a safety net to absorb the expected loss on financial institution's loan portfolio; this amount is referred to as loan loss provisions (LLPs) or provisions for bad

debts. The LLP estimate is a credit risk management tool used by financial institutions to mitigate expected losses on the financial institutions' loan portfolios (Ozili & Outa, 2017). The LLP is an important subject in the banking sector because the banks' large number of loans on their balance sheets makes them vulnerable to default. The vulnerability is due to the deteriorating economic conditions which might affect borrowers' ability to repay. According to Laeven and Majnoni (2003), banks are required to keep sufficient LLPs in anticipation of expected loan losses.

In the past, banks used International Accounting Standards 39 (IAS 39) (International Financial Reporting Standards Foundation, 2011) to establish their loan loss provisions. IAS 39 has played an instrumental role in shaping the financial landscape by defining standards for the recognition and measurement of financial assets and liabilities, as well as guidance on impairment. These principles have provided a foundation for consistency and transparency in financial reporting. The aim of IAS 39 was to establish principles for recognizing and measuring financial assets, financial liabilities, and some contracts to buy or sell non-financial items. Under IAS 39, banks would make provisions after the default event has occurred, which might be "too little, too late" provisioning for loan losses. Scholars often refer to IAS 39 as a reactive approach because IAS 39 waits for the default event to occur to start recognizing LLP. The 2007–2009 Global Financial Crisis and the subsequent credit crunch sparked a series of concerns around accounting reporting standards. As a result, the IASB introduced a new reporting standard which is the International Financial Reporting Standard 9 (IFRS 9) for the classification and measurement of financial instruments and the impairment of financial assets (International Accounting Standards Board, 2014).

IFRS 9 expedites the recognition of credit losses by requiring provisions that encompass both incurred losses and those expected in the future (International Accounting Standards Board, 2014). The new IFRS 9 standard adopts a forward-looking approach and requires provisions to be made based on the assumption of credit losses at the time assets are first recognized (International Accounting Standards Board, 2014). According to Novotny-Farkas (2015), IFRS 9 will mitigate the amplifying effect of the incurred loss approach on procyclicality and enhance financial stability. The impairment provision as per IFRS 9, referred to as Expected Credit Loss (ECL), is determined by assessing the anticipated economic loss associated with a loan or asset in question. The ECL from the reporting date until the maturity of a loan is calculated as follows:

$$ECL = \sum_t^T PD_t \times LGD_t \times EAD_t, \quad (1)$$

where PD_t is the probability of default, LGD_t is the loss given default, EAD_t is the exposure at default, T is the maturity time, and the subscript t denotes the time point (International Accounting Standards Board, 2014).

The IFRS 9 standard introduced the staging buckets approach for calculating the ECL based on the stage allocated to the loan/financial instruments; see Figure 1. The stages are the main criteria for calculating the ECL and continuous monitoring of the credit risk. The IFRS 9 standard is divided into three stages which are stage 1, stage 2, and stage 3; see Figure 1. According to the International Accounting Standards Board (2014), a loan/financial instrument will be moved between stages. The transition between stages will depend on the credit quality of the obligor and if a significant increase in credit risk (SICR) is observed.

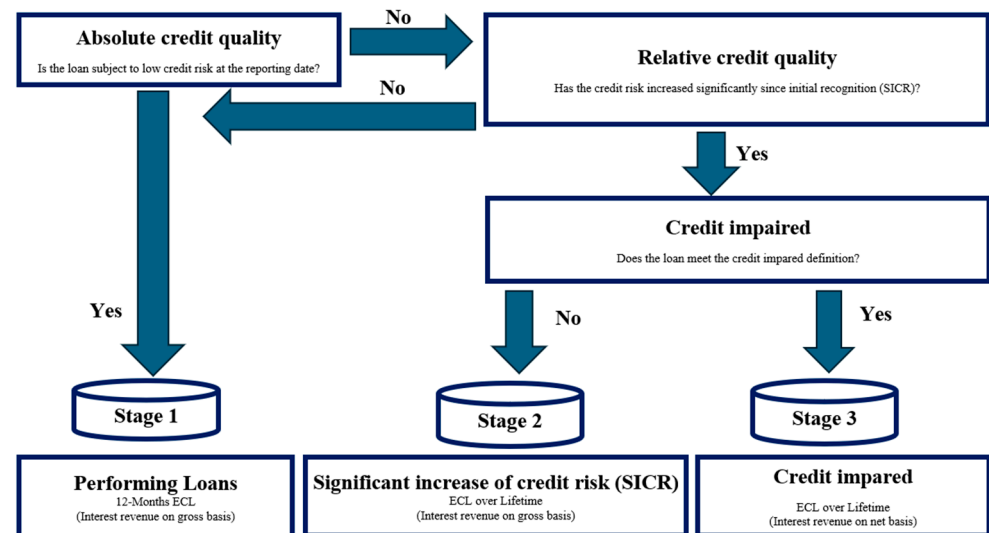


Figure 1. Three stages decision tree.

As shown in Figure 1, the absolute credit quality involves judging the financial instrument's credit quality solely based on a fixed threshold or a specific level of credit risk at the reporting date, regardless of its initial quality-like default definition or regulatory thresholds. The relative assessment requires an entity to compare the credit risk of a financial instrument at the reporting date with the credit risk at the date of initial recognition.

Ertan and Gansmann (2015), in their thesis at the Stockholm School of Economics, used survival analysis methodology but proposed a semi-parametric PD model that also complies with IFRS 9 standards. Their analysis incorporates both firm-specific (idiosyncratic) and macroeconomic (systematic) variables to evaluate alternative PD estimation models. The study identifies the semi-parametric Cox PH model as particularly effective, demonstrating strong predictive accuracy and alignment with IFRS 9 requirements with an area under the curve (AUC) of 0.87.

Brunel (2016) provides a comprehensive account of PD modeling techniques using credit portfolio analytics, with a particular focus on PD curve calibration. The study highlights the growing importance of dynamic and long-term risk assessment, especially considering IFRS 9 implementation, which exposed the lack of standardized methodologies for PD estimation. The paper outlines the distinction between wholesale and retail portfolios, emphasizing how their structural and risk characteristics influence the choice of modeling techniques. Their study identifies Cox's PH model as particularly effective for lifetime PD estimation due to its ability to incorporate multiple covariates and account for cure rates, thereby reducing estimation bias.

Santos (2018) employs survival analysis to estimate Observed Default Rates (ODRs), distinguishing between lifetime and 12-month horizons through tailored censorship adjustments. To derive through-the-cycle PD, the study applies a smoothed ODR curve using an adapted Nelson–Siegel model. For PIT PD estimation, Santos (2018) utilizes the Cox PH model, enabling sensitivity to prevailing economic conditions.

Bank and Eder (2021) present a comprehensive review of various PD modeling approaches relevant to IFRS 9, including survival models. The authors emphasize that the selection of an appropriate modeling technique is contingent upon several factors such as data availability, computational efficiency, forecasting horizon, and the clarity with which results can be communicated. The key focus of the paper is survival analysis models, also referred to as event history models, which are particularly well-suited to credit risk modeling due to their ability to estimate the time until default. Semi-parametric models,

such as the Cox PH model, allow for the inclusion of time-varying covariates, enabling the dynamic assessment of risk factors over time.

Ptak-Chmielewska and Gonzalez (2024) investigate the application of the Cox PH model within the IFRS 9 ECL framework, emphasizing its suitability for lifetime PD estimation. The study highlights the model's dynamic nature and its capacity to incorporate macroeconomic covariates, aligning with IFRS 9's forward-looking requirements. The authors report that the Cox PH model demonstrates strong discriminatory power, with an average AUC of approximately 0.8, particularly high in the early years of the credit lifecycle. The study concludes that survival analysis—especially the Cox PH model—is a robust and flexible tool for PD estimation under IFRS 9, as it effectively captures time-varying internal and external risk factors, offering a comprehensive view of credit risk over the loan's lifetime. Botha and Verster (2025) conducted a comprehensive study that offers a detailed and practical guide to modeling lifetime PD in alignment with IFRS 9, with a strong focus on discrete-time survival analysis. Their goal was to examine and contrast various modeling methods for estimating the term structure of default risk. Moreover, they supply practical tools and diagnostics for the implementation, validation, and performance assessment of these models.

The research conducted by Hardin and Ingre (2021) investigates the application of supervised statistical learning techniques to predict the PD for 'Buy Now Pay Later' credit agreements, while also examining the influence of macroeconomic factors in accordance with IFRS 9 requirements. Their findings suggest that extreme gradient boosting (XGBoost) is a suitable method and highlight the significance of macroeconomic variables in 'Buy Now Pay Later' PD modeling, which ultimately aids the financial sector and society by enhancing credit risk management practices.

Englund and Mostberg (2022) investigate the effectiveness of ML in modeling the PD term structure, comparing deep neural networks and gradient boosted trees with the traditional Markov chain model. They assessed the performance of the models using calibration and discrimination metrics at each time point, as well as aggregating results over the entire time horizon. The objective of their study was to investigate whether ML models can effectively predict PD term structures and surpass the traditional Markov chain method. Their findings suggested that ML can be effectively utilized in PD term structure modeling because the deep neural networks model consistently outperformed the Markov chain model across all performance metrics, whereas Gradient Boosted Trees excelled in all but one metric. Furthermore, ML models demonstrated superiority in long-term predictions, while in short-term predictions, they only marginally outperformed the Markov chain model.

Saavedra et al. (2024) presented a ML model and considered a competing risks survival analysis model aimed at evaluating the PD aspect of credit risk, which is relevant for assessing lifetime ECL in accordance with the IFRS 9 regulation. Kottayil et al. (2025) investigate the use of ML algorithms, particularly XGBoost and RBF, to enhance the accuracy of PD term structure forecasting within the IFRS 9 regulatory framework. Their research underscores the significance of precise PD term structure forecasts for effective business planning and adherence to regulations that necessitate the estimation of PD throughout the entire duration of a credit contract. The paper compares the ML approaches with conventional methods such as logistic regression and Markov chains, emphasizing the benefits of ML in understanding intricate relationships and adjusting to changing risk patterns. The findings of the research demonstrate that ML models are proficient in forecasting PD term structures and surpass the baseline discrete homogeneous Markov chain model in the adjusted average metric.

This study focuses on the PD component to be used as an input in Equation (1) to calculate the ECL. The aim of the study is to derive a PD term structure component, which is the probability of the obligor defaulting at time t given that the obligor survives time $(t - 1)$, hence the following methodologies explored in this study: (i) Survival Analysis—Cox proportional hazard (PH); (ii) Survival analysis—Extended Cox PH (Ext Cox PH); and (iii) Machine Learning—Random Boosting Forest (RBF). These methodologies are widely used for modeling time to default and predicting the obligor's future credit behaviour.

One of the main objectives of exploring survival analysis is that the model can predict whether the obligors will default on their loans and also when they are likely to default (Narain, 2004). According to Smuts and Allison (2020), survival analysis can estimate the probability that a borrower will still be repaying a loan at each time point along the survival curve (e.g., each month), and these probabilities can be estimated with high accuracy. Another major strength of survival analysis is that it allows censored data to be included in the model. For instance, in the consumer credit context, censored customers are customers who have never experienced an event of interest, which is defaults (Stepanova & Thomas, 2002).

Machine learning employs various methods broadly categorized into supervised, unsupervised, semi-supervised, and reinforcement learning. Supervised learning uses labeled data to train models, enabling them to predict outcomes based on explicit input–output examples (Mahesh, 2020). Unsupervised learning analyzes unlabeled data to uncover hidden structures and patterns, making it ideal for tasks like clustering and anomaly detection (Hastie et al., 2009). Semi-supervised learning leverages both labeled and unlabeled data, balancing the strengths of supervised and unsupervised methods which is especially useful when labeled data is scarce or expensive to obtain (James et al., 2023). Reinforcement learning trains algorithms through a process of trial and error, where the model learns to take actions that maximize cumulative rewards, thus excelling in dynamic and sequential decision-making tasks.

The mandate under IFRS 9 requires financial institutions to establish a thorough assessment of obligors to predict their behaviour throughout the credit life and estimate the ECL continually, instead of annually. The modeling technique, like logistic regression, on its own is inadequate for estimating the lifetime PD because it gives a static PD for a fixed outcome period (Banasik et al., 1999, Bellotti & Crook, 2009). Traditionally, many scholars prefer and have explored the Markov chain modeling technique to build the IFRS 9 PD term structure model and very few scholars have explored other alternative modeling techniques (Ludwig & Bank, 2019).

There are few research studies that have compared the accuracy of more than two modeling techniques for term structure PDs, using Kaplan–Meier as the benchmark model under IFRS 9 PD term structures. Englund and Mostberg (2022) compared the machine learning techniques against the traditional modeling techniques (Markov chain) but did not use Kaplan–Meier as a benchmark. Bank and Eder (2021) review a couple of models' strengths and weaknesses but did not benchmark them against the Kaplan–Meier. The objective of considering Kaplan–Meier is that its estimation is based solely on observed survival times and censoring, reflecting the empirical survival experience of the population without imposing a modeling structure. In this study, we did not consider the competing risk which involves the risk of prepayments. This is one of the limitations of the study that can be addressed in future studies.

This is a data-driven study that aims to illustrate how to derive the term-structured PDs that need to be used as inputs into the IFRS 9 ECL calculation using specifically the mortgage loans portfolio data by Freddie Mac. Currently, there is no thorough study in the credit risk literature that illustrates how this valuable and rich Freddie Mac dataset can be implemented under the IFRS 9 regulation that is currently used by almost all (if not all)

banks nowadays. Therefore, the specific objective of this study is to separately use survival analysis (Cox PH and Ext Cox PH) methods and the RBF to derive the term structure PDs by computing both the lifetime PD of the obligor and the time to default. In addition, the Kaplan–Meier model will be used as a benchmark. That is, the three methodologies will be compared against the Kaplan–Meier model and against each other, and the top-performing methodologies will be systematically identified using different evaluation metrics. This research work intends to answer the following questions:

- (RQ1) Which evaluation metric can be used to systematically identify the best-performing modeling technique?
- (RQ2) For the three implemented methods to formulate the PD, which one tends to outperform the rest?
- (RQ3) For the three implemented methods to formulate the PD, which one is the most time-efficient to use?

The remaining sections of the paper are structured as follows: Section 2 outlines the methodology used in this paper. Section 3 provides a detailed analysis and the results of the applied methodologies on the Freddie Mac dataset. Then, a detailed discussion of the comparison of the three methods we focus on in this research is provided in Section 4. Lastly, the conclusions of the study are provided in Section 5.

2. Methodology

This section discusses the theoretical aspects of the methodologies used in this study.

2.1. Kaplan–Meier

Kaplan–Meier is a non-parametric technique widely used in survival analysis to estimate the probability that a subject survives beyond a certain time. The survival curve is built incrementally at each event time using the proportion of subjects who survive:

$$S(t) = P(T > t), \quad (2)$$

where T is the time until the event occurs. The method is fundamentally empirical (Turkson et al., 2021) and does not rely on assumptions about the underlying distribution of survival times (George et al., 2014). Instead, it uses observed data—specifically, the actual times at which events occur—to estimate survival probabilities. These probabilities are calculated directly from the observed event times and censoring patterns. At each time point where an event occurs, the methodology calculates the probability of surviving past that time, then multiplies this probability by the previous survival probability to obtain the cumulative survival. This process continues for all observed event times, producing a Kaplan–Meier survival curve that reflects the actual experience of the population,

$$S(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right) \quad (3)$$

where $S(t)$ is the estimated probability of surviving beyond time t , t_i is time of the ordered i^{th} event (e.g., default, death), d_i is the number of events (e.g., defaults) that occurred at time t_i , and n_i is the number of individuals at risk just before time t_i .

The hazard rate $H(t_i)$ at time t_i is the rate of experiencing default, given that the customer has survived up to that point and it is given by

$$H(t_i) = \frac{d_i}{n_i}, \quad (4)$$

where d_i is the number of events at time t_i and n_i is number of individuals at risk at time t_i . The i -th distinct event time refers to the ordered time points at which at least one event occurs (Herrmann et al., 2021). Predictions from other methodological approaches will be compared against the actual outcomes derived using the Kaplan–Meier method. This is because the Kaplan–Meier methodology accurately reflects the real-world experience of the population.

2.2. Cox Proportional Hazard (PH)

The Cox PH model is a statistical technique in survival analysis, which is widely used in finance, economics, and the health sector, especially for investigating the correlation between the survival duration of subjects and one or more predictor variables (Tibshirani, 2022). The model is commonly employed to model the hazard function, and it is thought of as a semi-parametric model for estimating the hazard function. The hazard function represents the probability of default taking place at time t , given that the customers have survived up to the time t . The Cox PH model is made up of two components: the baseline hazard rate, which varies with time, and the effect of parameters, which represent a collection of explanatory variables that account for variations in the hazard rate:

$$\lambda(t|X) = \lambda_0(t) \cdot \exp(\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p), \quad (5)$$

where $\lambda(t|X)$ is the hazard at time t for a subject with covariates X and $\lambda_0(t)$ is the baseline hazard function. The PH model assumes that the hazard ratio among individuals remains constant through time (Caselli et al., 2021). This model handles censored data, particularly when the event has not transpired for all subjects during the observation period. Equation (6) calculates the hazard ratio and can be obtained by rearranging the terms in Equation (5):

$$\frac{\lambda(t|X)}{\lambda_0(t)} = \exp(\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p). \quad (6)$$

The hazard ratio represents the proportion of the hazard rate compared to the baseline hazard rate. The hazard ratio can be interpreted as the instantaneous risk of an event occurrence with the effect of a variable over the control arm, i.e., without the effect of the variable. The Cox PH model survival function, $S_0(t)$, is borrower-specific survival function that considers the covariates:

$$S(t|X) = S_0(t)^{\exp(\beta^T X)} \quad (7)$$

where $S_0(t)$ is the baseline survival function (non-parametric), β is a vector of regression covariates and X is a vector of covariates. The estimation of the regression coefficient β_j is performed using the partial likelihood method, a robust technique introduced by Cox, which is expressed in the following manner (Cox, 1972):

$$L(\beta_j) = \prod_{i=1}^m \frac{\exp\{\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p\}}{\sum_{k \in R_i(t)} \exp\{\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p\}}, \quad (8)$$

where $R(t_i)$ is the risk set at time t_i (i.e., individuals still at risk just before t_i) and m is number of observed events.

The objective of the Cox PH model in this research is to derive the lifetime PDs, namely $PD(t|X)$, as defined in terms to $\lambda(t|X)$. The lifetime PD refers to the likelihood associated with the remaining duration of the loan; this probability can manifest in various forms.

$$\text{Lifetime PD}(t) = 1 - S(t|X) \quad (9)$$

However, the marginal probability is:

$$\text{Marginal PD} = \text{Lifetime PD}(t+1) - \text{Lifetime PD}(t) \quad (10)$$

2.3. Extended Cox Proportional Hazard

The standard Cox PH assumes that covariates do not change during the follow-up period and that their impact on the hazard remains constant over time. In the standard Cox PH, each subject is represented by a single row. When hazard ratios stay constant throughout time, this model performs well with static variables. Nevertheless, the validity of the model's conclusions is jeopardized when the proportionate hazards assumption is violated. These presumptions frequently fall short in many real-world situations, particularly in industries like finance and health (Lin & Zelterman, 2002). The Ext Cox PH is a versatile framework for survival analysis that can account for non-proportional hazards and time-dependent covariates, making it ideal for modeling dynamic risk over time (Kulinskaya, 2019). Time-dependent variables allow information that varies throughout the course of the investigation to be incorporated. The proportional hazards assumption is relaxed by the Ext Cox PH, which permits covariates to change over time. The model also allows for various baseline hazards across strata (e.g., gender, region), and these covariates might affect the hazard in different ways over time. Ext Cox PH provide a more adaptable and realistic way to analyze survival data by taking dynamic processes into consideration. The Ext Cox PH is used in place of the standard Cox PH when the standard Cox PH assumptions are violated or when the data structure requires greater flexibility (Kleinbaum & Klein, 2012). The hazard function becomes:

$$\lambda(t|X) = \lambda_0(t) \times \exp(\beta(t)^T X(t)), \quad (11)$$

where $\lambda(t|X(t))$ is the hazard at time t given covariates $X(t)$ that varies with time, $\lambda_0(t)$ is the baseline hazard, $\beta(t)^T$ is covariate that may vary with time. The Equation (12) calculates the hazard ratio and can be obtained by rearranging the terms in Equation (11):

$$\frac{\lambda(t|X(t))}{\lambda_0(t)} = \exp(\beta(t)^T X(t)). \quad (12)$$

The survival probability at time t depends on the entire history of covariates up to time t

$$S(t|X(t)) = \exp\left(-\int_0^t S_0(u) \exp(\beta^T X(u) du)\right), \quad (13)$$

where $S_0(t)$ is the baseline survival function. Each subject in the expanded Cox PH has several rows, each of which represents a time period, in a counting process data structure. Based on the martingale-based estimate and counting process theory, it enables robust inference under time-varying conditions, flexible hazard modeling, and compatibility with partial likelihood estimation. The Ext Cox PH is used in credit risk to monitor borrower behaviour and macroeconomic changes, in clinical trials to represent treatment effects that fluctuate over time, and in epidemiology to model disease progression with shifting exposures (Bellotti & Crook, 2009).

The objective is to use the Ext Cox PH model to derive the lifetime PDs, namely $PD(t|X)$, as defined in terms to $\lambda(t|X(t))$. The lifetime PD refers to the likelihood of a borrower defaulting within the remaining duration of the loan; this probability can manifest in various forms:

$$\text{Lifetime PD}(t) = 1 - S(t|X(t)). \quad (14)$$

However, the marginal probability is:

$$\text{Marginal PD} = \text{Lifetime PD}(t+1) - \text{Lifetime PD}(t). \quad (15)$$

2.4. Machine Learning—Random Boosting Forest (RBF)

RBF is a machine learning technique that enhances the conventional random forest algorithm to manage time-to-event data with right-censored outcomes, such as in survival analysis. RBF is a hybrid ensemble approach that integrates concepts from both random forests and boosting to effectively model time-to-event data, particularly in the presence of censoring. It merges the randomness characteristic of random forests with the sequential learning process of boosting.

This ensemble consists of independent sets of gradient boosted trees, rather than the standard decision trees. Trees are constructed sequentially, similarly to boosting, but each tree is trained on a randomly selected subset of features and data, akin to random forest (Englund & Mostberg, 2022). In essence, each decision tree in the traditional random forest is substituted with a relatively small ensemble of XGBoost models. Unlike the bootstrap sampling method employed in random forests, the input data for the RBF is randomly sub-sampled for each XGBoost ensemble, denoted as first, second until Bth on Figure 2. This design aims to balance the bias–variance trade-off, enhance robustness, and effectively manage complex survival data (Kottayil et al., 2025).

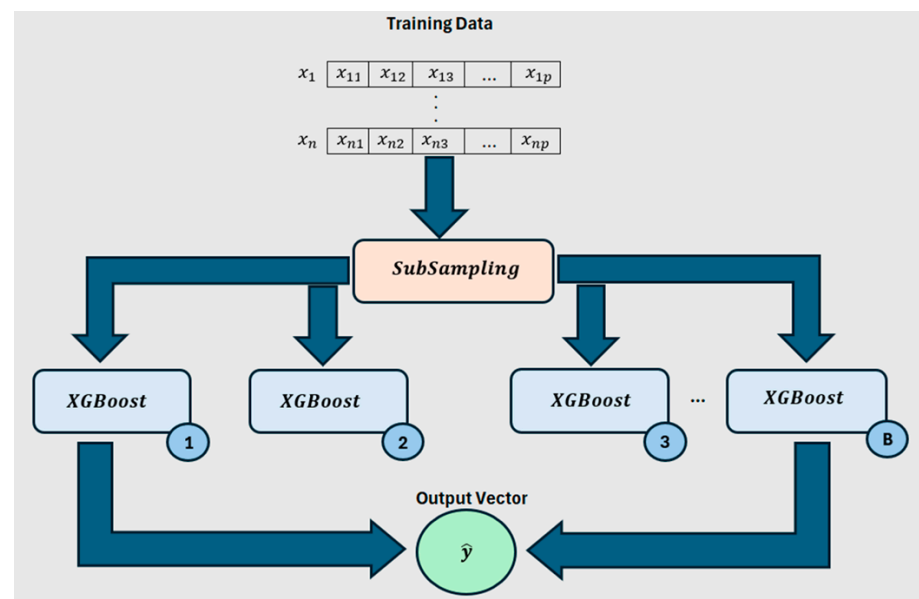


Figure 2. Random Boosting Forest model architecture.

The RBF model is developed within a Cox regression framework, while its predictions rely on the Kaplan–Meier estimator. Before discussing the training process of the RBF, it is necessary to revisit the Cox PH:

$$\lambda(t|X) = \lambda_0(t) \cdot \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p). \quad (16)$$

In RBF, the linear predictor is replaced by an XGBoost ensemble; hence, one ensemble is characterized as

$$\lambda^{(b)}(t|X_i) = \lambda_0^{(b)}(t) \delta_b(x_i), \quad (17)$$

where δ_b is the b^{th} XGBoost ensemble, see Figure 2. As a result, the complete model transforms into:

$$\lambda^{(*)}(t|X_i) = \frac{1}{B} \sum_{b=1}^B \lambda_0^{(b)}(t) \delta_b(x_i). \quad (18)$$

The optimization of the RBF model pertains to the negative Cox partial log-likelihood and because the Cox regression model is unable to directly represent the survival function, the Kaplan–Meier estimator is utilized at the terminal nodes of the XGBoost ensembles. The XGBoost ensembles are permitted to forecast the training data through their foundational Cox PH. Subsequently, at each terminal node, a survival curve is estimated using the Kaplan–Meier estimator, relying on the observations allocated to the terminal node (Kottayil et al., 2025). To predict an unseen observation, each ensemble assigns it to a terminal node according to its foundational Cox PH and provides the corresponding survival curve. Ultimately, the predictions from each ensemble are averaged to yield a final prediction.

The rationale for utilizing multiple parallel XGBoost ensembles is to enhance stability of the predicted term structures. Given that the Kaplan–Meier estimator is non-parametric, predictions derived from a single ensemble would rely on a limited number of observations at each leaf node. Consequently, these predictions would inherently exhibit instability and high variance. Therefore, by employing several parallel ensembles, we can address the variance problem and enhance the model’s predictive performance.

The RBF model provides robustness and reduces variance while achieving high accuracy through the sequential correction of errors (Ghosal & Hooker, 2020). The randomization inherent in RBF mitigates overfitting, while the boosting process contributes to bias reduction, resulting in improved generalization on previously unseen data. The RBF effectively manages complex datasets and can appropriately handle censored data when implemented correctly, adapting to loss functions specific to survival analysis. Additionally, the independent boosted ensembles of RBF can be trained concurrently, enhancing scalability and offering interpretable insights into feature importance.

2.5. Software

This study uses SAS (Statistical Analysis Software) 9.4 (TS1M9) for Windows Desktop to perform data cleaning and data preparation. The dataset prepared for modeling was then exported to a Microsoft Excel file. For modeling purposes, this study uses the R programming software version 4.5.2 with the following libraries and their main purpose provided in parentheses: the readxl (to read and import the data in the excel), survival (to model the survival data), survminer (to create survival analysis results), ggplot2 (to create a customizable graphic for data visualization), ggfortify (to provide a unified, automatic way to visualize the results of many statistical models and analyses), survAUC (to evaluate the prediction accuracy of survival models), timeROC (to compute time-dependent ROC curves and time-dependent AUC), XGBoost (for optimized, high-performance implementation of gradient boosting), dplyr (for fast, consistent, and intuitive data manipulation).

3. Analysis and Results

This section outlines the analysis performed on the model development data and the results derived from the models utilized in this study. All the SAS and R codes to reproduce the results herein are provided as a Supplementary Material.

3.1. Data Description

This study makes use of publicly available mortgage data provided by Freddie Mac. The dataset, accessible through the Freddie Mac website, includes both origination and performance information. The origination data captures key loan characteristics at the time the mortgage was issued, while the performance data tracks how each loan behaves over

time, from the first payment through to maturity, which offers valuable insights into default patterns and credit risk dynamics. Freddie Mac was chartered by Congress in 1970 to support the U.S. Housing Finance System and help ensure a reliable and affordable supply of mortgage funds across the country. Freddie Mac operates in the U.S. secondary mortgage market, purchasing from lenders' single-family and multifamily residential mortgage loans. For more information about Freddie Mac, please visit <https://www.freddiemac.com/>.

The file downloaded from Freddie Mac includes the origination dataset which captures the accounts behaviour at acquisition and performance data to capture the customers' behaviour over the lifetime of the loan. The origination data captured at loan origination is important in credit risk modeling because it identifies the type of borrower, the loan acquisition strategy, and the approval pattern. These features are critical for modeling credit risk and default probability. The performance data consists of credit behavioural variables, which offer a dynamic view of how customers manage their loan obligations from the time the loan is originated until it reaches maturity. The behavioural variables help identify customer patterns, such as repayment consistency, credit utilization, and delinquent behaviour. They also provide valuable insights into the customers' risk matrix. The origination dataset consists of 50,000 unique accounts, all of which were originated in 2004. The performance dataset comprises 3,974,583 rows of performance data and 50,000 unique accounts. See Table A1 in Appendix A for the original 32 variables and the Supplementary Material for the detailed account of removed variables and the corresponding reasons thereof.

The study assesses and estimates default times within the mortgage portfolio under IFRS 9. The task was performed using survival analysis and machine learning methodologies due to the versatility of these models and their ability to model time to default on a forward-looking basis. The study uses the Kaplan–Meier method as the benchmark since it is derived from actual data.

In this study, the derived models were examined (to assess the models' discrimination performance) using the following methods: Receiver Operating Characteristic (ROC) curve, area under the curve (AUC), Concordance index (C-index), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Consistent Akaike Information Criterion (CAIC). Note that C-index in credit risk is a statistical metric used to evaluate the discriminatory power of a predictive model (e.g., a credit scoring model) by measuring its ability to correctly rank borrowers by their relative risk of default (Bone-Winkel & Reichenbach, 2024). The C-index measures the model discrimination power, meaning it measures how well the model ranks individuals by risk. The ROC is a statistical tool used to visually assess and compare the discriminatory power of credit scoring models (Kochański, 2022). The AUC is derived from the ROC curve, and it is used to evaluate the discriminatory power of a model to distinguish between defaulting (bad) and non-defaulting (good) lenders (Kraus & Küchenhoff, 2014). The AUC is a single numerical value (ranging from 0.5 to 1.0) that summarizes the performance of a credit scoring model across all possible classifications. The AIC evaluates model fit while penalizing complexity. The BIC is like AIC but imposes a stronger penalty for model complexity. The CAIC is a variant of AIC that further penalizes complexity to ensure consistency in model selection. Model diagnostics, such as Schoenfeld residuals, were also applied to evaluate model fit. Finally, nomograms were used to display covariate information and survival probabilities.

The data points beyond the loan age of 120 months are excluded in the development dataset. This exclusion is based on Figure 3, which demonstrates that the default rate post 120 months starts becoming volatile, with an outlier observed at month 187. The number of defaults post 120 months are low, with some months recording zero defaults, and there is no clear default trend after 120 months.

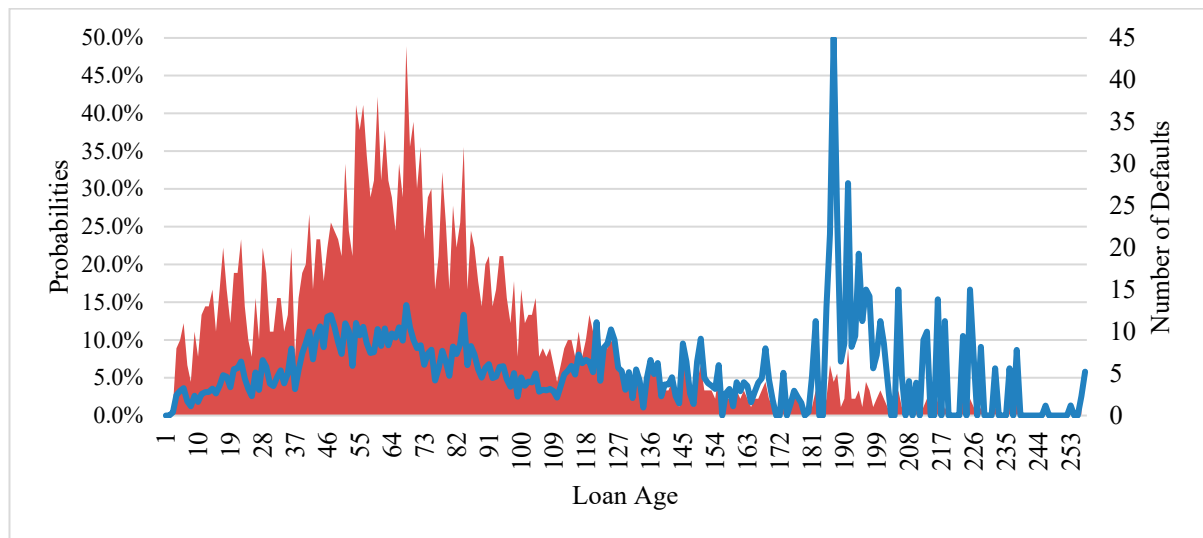


Figure 3. Default rate and number of defaults distribution.

Figure 4 displays the default rate and the number of defaults after excluding data points post 120 months. The graph shows that the default rate follows a clear trend over time, as indicated by the yellow trend line. This trend suggests that the probability of an account defaulting is low during the initial stage of the loan but gradually increases as the loan progresses, until it reaches a point where it begins to decline. This pattern can be explained by the fact that customers have become more accustomed to managing their debt, leading to a decrease in the default rate beyond a certain point.



Figure 4. Default rate and number of defaults distribution (excl. points after 120 months).

3.2. Kaplan–Meier

To estimate the survival function from time-to-event data and especially when censoring is present, the Kaplan–Meier estimator is used. It provides a stepwise survival curve that shows the probability of surviving beyond each observed event time. The Kaplan–Meier model in this study is used as a benchmark. To evaluate the performance of more complex models, such as Cox PH or parametric survival models it is common to employ Kaplan–Meier as the benchmark model. This is because Kaplan–Meier makes no assumptions about the hazard function and has no covariate effects. Its estimator is based solely on observed survival times and censoring, reflecting the empirical survival experience of the population without imposing a modeling structure. It offers a clear

picture of survival probabilities over time, which helps contextualize the performance of more complex models.

3.3. Cox Proportional Hazard

This section outlines the model selection criterion performed in selecting the best performing model. The model diagnostics such as Schoenfeld residuals test is also performed to the best-fitting model to evaluate if the Cox PH model adheres to the PH assumption. The covariates that violated the PH assumption are removed and reconsidered in subsequent models. Table 1 presents the details of the initial model iteration and the variables that were included and excluded in the model. The C-index for this model is 0.757.

Table 1. Initial Cox PH model.

Variables	Coefficient	Se (coef)	Exp (coeff)	p-Value	C-Index
<i>Original Loan Term</i>	0.00602	0.000360	1.006000	<0.001 ***	
<i>Loan Purpose</i>	−0.21220	0.027920	0.808800	<0.001 ***	
<i>Number of Borrowers</i>	−0.35370	0.046480	0.702100	<0.001 ***	
<i>Channel</i>	0.14910	0.044500	1.161000	0.000806 ***	
<i>Credit Score</i>	−0.00685	0.000376	0.993200	<0.001 ***	0.757
<i>Mortgage Insurance%</i>	0.00710	0.000501	1.007000	<0.001 ***	(0.005)
<i>Original UPB</i>	0.00000	0.000000	1.000000	0.094	
<i>DTI</i>	0.00006	0.000182	1.000000	0.762	
<i>Number of Units</i>	−0.10990	0.101400	0.895900	0.279	
<i>Current Non-Interest Bearing UPB</i>	−0.00418	0.158000	0.995800	0.979	

Note: *** significant at 1% level of significance.

The *p*-value of a variable is used to outline the variable explanatory power for the dependent variable and variables with high *p*-values are excluded in the model. Due to high *p*-values, the variables *Original UPB*, *DTI*, *Number of Units*, and *Current Non-Interest Bearing UPB* were excluded from the model as these variables have demonstrated low explanatory power for the dependent variable.

The Schoenfeld residuals test was conducted on the significant variables. Figure 5 outlines the results of the test where the Schoenfeld test plots the residuals of each variable against the survival time. This plot serves as the basis for testing the correlation between Schoenfeld residuals and survival time, which demonstrate the goodness-of-fit test. The null hypothesis of the test states that there is zero correlation between residuals and survival time. If the *p*-value is greater than the 0.05 significance level, the null hypothesis is not rejected and it is concluded that the PH assumption is satisfied. For four variables, the null hypothesis was rejected meaning that the variables did not meet the PH assumption, while for two variables the null hypothesis was accepted. The two variables adhered to the PH assumption, and the two variables are *Original Loan Term*, with a *p*-value of 0.7767, and *Loan Purpose*, with a *p*-value of 0.9886.

Table 2 presents all statistically significant variables which form part of the final Cox model. The variables were tested for PH assumptions using Schoenfeld residuals and the variables passed the test. Then, the final Cox model is made up of two variables which are *Original Loan Term* and the *Loan Purpose* variables.

The variables that did not meet the PH assessment criteria were removed; after that, we ran the Schoenfeld residuals on the two variables that satisfied the PH assumption. The *p*-values for the two variables remained above 0.05, and the global Schoenfeld residual value increased to above 0.05. Figure 6 shows the Cox PH model Schoenfeld residual global test and the *p*-values of the two variables.

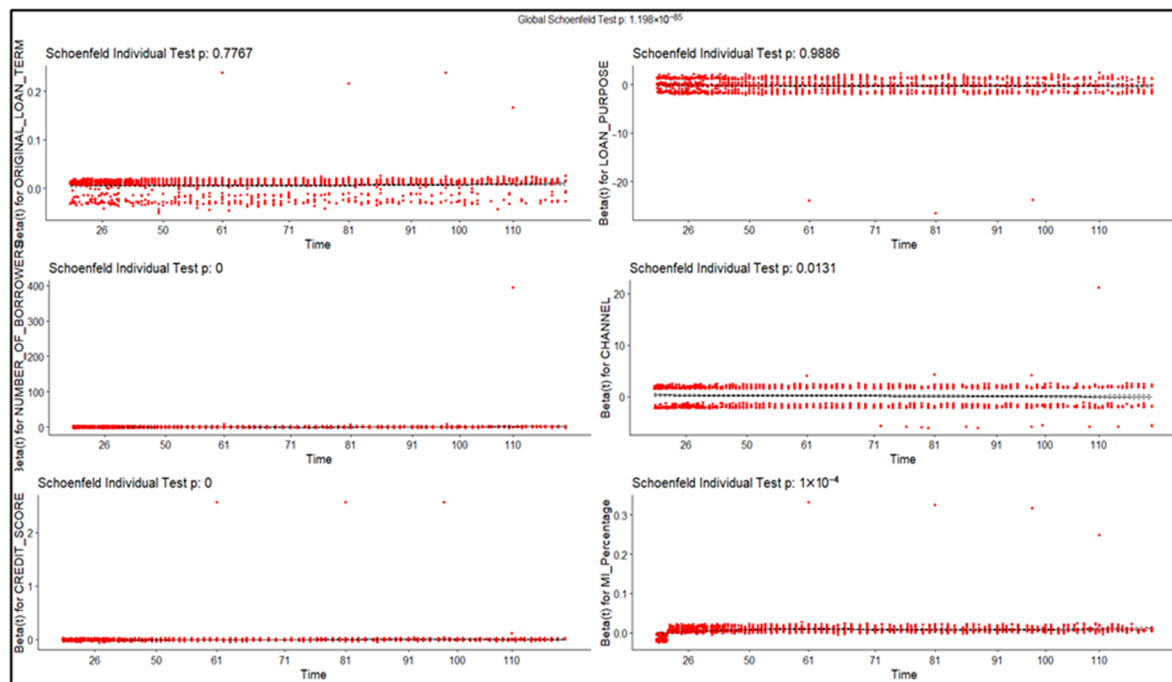


Figure 5. Initial PH test using Schoenfeld residuals.

Table 2. Final Cox model.

Variable	Coefficient	Se (coef)	Exp (coeff)	p-Value	C-Index
Original Loan Term	0.0070202	0.0003492	1.0070	<0.001	0.639
Loan Purpose	−0.2576336	0.0275811	0.7729	<0.001	(0.006)

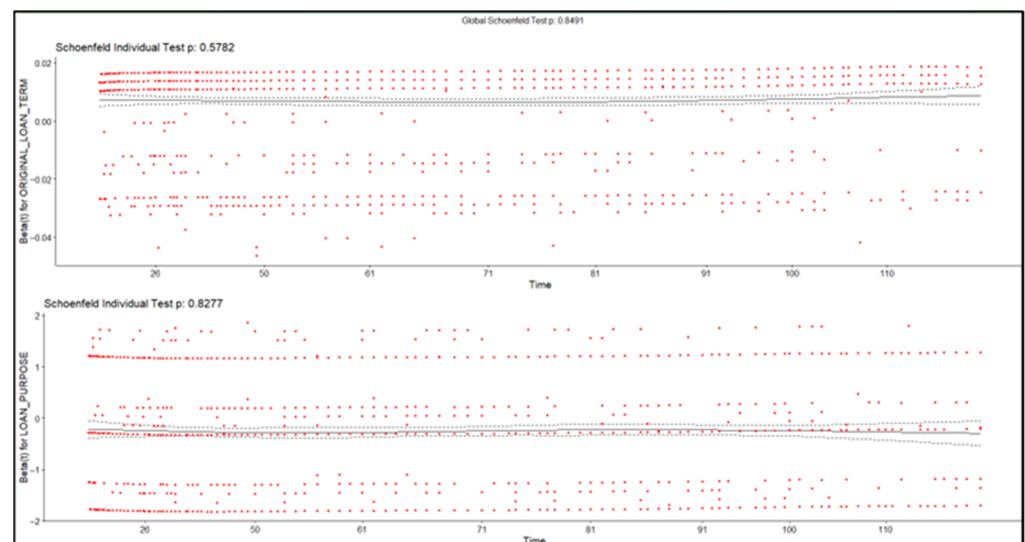


Figure 6. Final PH test using Schoenfeld residuals.

3.4. Extended Cox Proportional Hazard

The Ext Cox PH model allows the violation of the proportional hazards' assumption. Violating the PH assumption allows the inclusion of time-varying variables in the model, contrary to excluding them in the Cox PH model (Kleinbaum & Klein, 2012). The time-varying variables are recorded in the performance dataset and are captured over time.

The same set of variables identified under the Cox PH model after the correlation matrices step are used in the Ext Cox PH model. These variables are used as the starting point for selecting the best fit explanatory variables under Ext Cox PH. The stepwise selection method is used to select variables, and the variables marked with an asterisk are retained. Table 3 presents the details of the initial model iteration and the variables that formed part of the model.

Table 3. Initial Extended Cox PH model.

Variables	Coefficient	Se (coeff)	Exp (coeff)	p-Value	C-Index
<i>Original Loan Term</i>	0.00538	0.00032	1.00500	<0.001 ***	0.761 (0.005)
<i>Loan Purpose</i>	−0.22210	0.02561	0.80080	<0.001 ***	
<i>Number of Borrowers</i>	−0.37220	0.04235	0.68920	<0.001 ***	
<i>Channel</i>	0.07893	0.04111	1.08200	0.0548	
<i>Credit Score</i>	−0.00725	0.00033	0.99280	<0.001 ***	
<i>Mortgage Insurance%</i>	0.00657	0.00045	1.00700	<0.001 ***	
<i>Original UPB</i>	0.00000	0.00000	1.00000	0.6086	
<i>DTI</i>	0.00031	0.00015	1.00000	0.0443 *	
<i>Number of Units</i>	−0.03847	0.08296	0.96230	0.6428	
<i>Current Non-Interest Bearing UPB</i>	0.00000	0.00004	1.00000	0.9866	

Note: * significant at 5% level of significance; *** significant at 1% level of significance.

Due to high *p*-values, the variables *Channel*, *Original UPB*, *Number of Units* and *Current Non-Interest Bearing UPB* are excluded from the model; so, as a result, these variables cannot form part of the model. Then, the following results in Table 4 were observed after removing the insignificant variables. The final model is made up of six variables which are *Original Loan Term*, *Loan Purpose*, *Number of Borrowers*, *Credit Score*, *Mortgage Insurance Percentage* and *Debt to Income Ratio*.

Table 4. Final Extended Cox PH model.

Variables	Coefficient	Se (coef)	Exp (coeff)	p-Value	C-Index
<i>Original Loan Term</i>	0.005474	0.000311	1.005489	<0.001 ***	0.761 (0.005)
<i>Loan Purpose</i>	−0.224787	0.025584	0.798686	<0.001 ***	
<i>Number of Borrowers</i>	0.000310	0.000152	1.000311	<0.001 ***	
<i>Credit Score</i>	−0.371441	0.041322	0.689740	<0.001 ***	
<i>Mortgage Insurance%</i>	−0.007251	0.000328	0.992776	<0.001 ***	
<i>DTI</i>	0.006602	0.000447	1.006624	<0.001 ***	

Note: *** significant at 1% level of significance.

3.5. Random Boosting Forest

The single decision tree from an ensemble used in a boosted forest, such as those produced by models like XGBoost, is presented in Algorithm 1. Each tree in the ensemble contributes a small portion to the final prediction. Unlike random forests, boosting builds trees sequentially, with each tree correcting the errors of its predecessor. This tree is not standalone; it is one of many. The model aggregates predictions from all trees, typically using a weighted sum. These trees are shallow and specialized, focusing on correcting specific errors rather than modeling the entire decision space.

Algorithm 1. Steps implemented to conduct the RBF model

Input: dataset with variables:

CREDIT_SCORE, ORIGINAL_LOAN_TERM, NUMBER_OF_BORROWERS, Mortgage_Insurance%,
Current_Actual_UPB%, LOAN_AGE, Default_Flag

Step 1: Build feature matrix X

- Select predictors: CREDIT_SCORE, ORIGINAL_LOAN_TERM, NUMBER_OF_BORROWERS, Mortgage_Insurance%, Current_Actual_UPB%
- Encode predictors into numeric design matrix
- Remove intercept column

Step 2: Define target variables

- Loan_Age ← Dataset.LOAN_AGE // survival time
- default_Flag ← Dataset.Default_Flag // event indicator (defined but not used here)

Step 3: Create training dataset object

- Training Dataset ← Matrix(data = X, label = Loan_Age)

Step 4: Set model parameters

- objective = "survival:cox" // Cox proportional hazards model
- eval_metric = "cox-nloglik" // Negative log-likelihood for Cox
- max_depth = 2 // Maximum depth of trees
- eta = 0.1 // Learning rate
- subsample = 0.8 // Fraction of rows sampled per tree

Step 5: Train XGBoost model

For round = 1 to 100:

- Build a new decision tree using current parameters
- Add tree to the ensemble
- Update predictions from ensemble
- Compute evaluation metric (cox-nloglik if survival:cox)
- Adjust gradients and weights for next round

End loop

Output: trained survival model (ensemble of 100 boosted trees)

Each tree in the ensemble contributes a partial prediction, typically a log-risk score. These predictions are summed across trees to produce a final risk score for everyone. The risk scores are then used to estimate survival curves through the Cox PH stratification. Hazard functions are derived by modeling the instantaneous risk of event occurrence over time. The forest for the RBF model is provided in Appendix A as Figure A1.

3.6. Survival Probability Graphs

This section presents the survival probability graphs for different methodologies which shows the probability of a customer surviving over time. The survival probability graph in Figure 7 is derived using the Kaplan–Meier methodology. The graph visually represents the estimated probability that a customer will survive beyond a given time point. It is based on observed time-to-event data, and we observed a decreasing trend of the survival probabilities.

The survival probability graph in Figure 8 is derived using the Cox PH model and displays the customers' probabilities of surviving. The survival probabilities start high and gradually decrease as time progresses and a similar pattern is observed in the Kaplan–Meier survival probability graph.

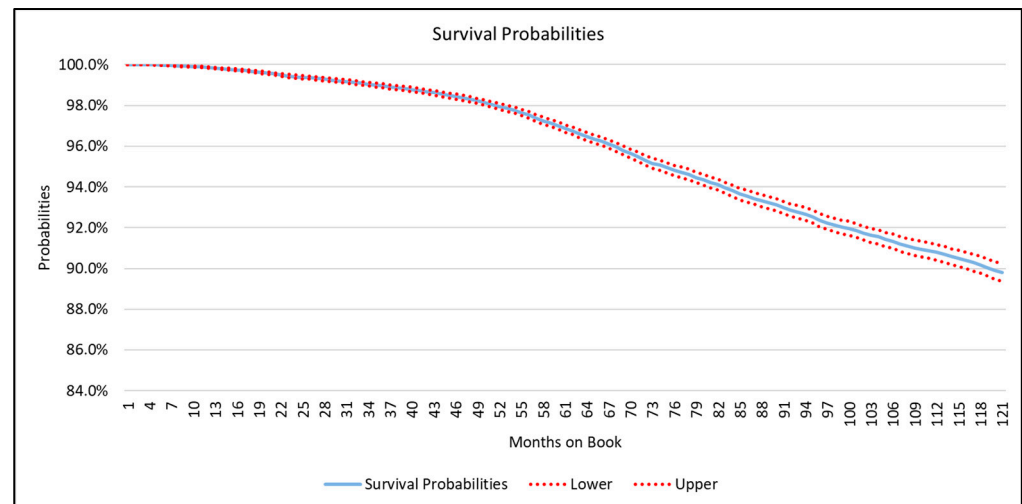


Figure 7. Kaplan–Meier survival probability curve.

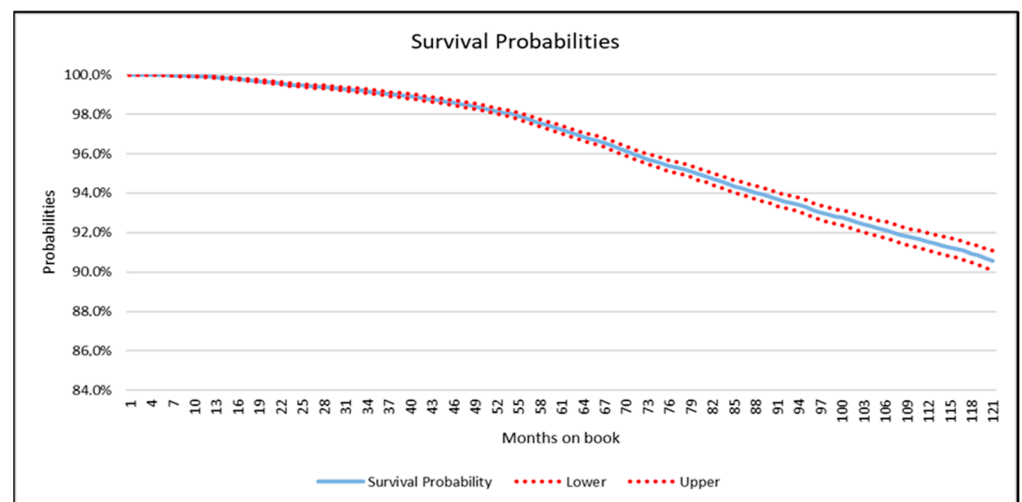


Figure 8. Cox PH model survival probabilities.

Due to the fact that the Ext Cox PH model allows time-dependent covariates, the hazard ratios change as time progresses. This flexibility results in distinct survival curves for different covariate patterns.

The purpose of this study is to compare the Ext Cox PH model to other models, as well as to the benchmark model, which is the Kaplan–Meier model. The average survival probability curve was produced by combining all the distinct survival curves from different covariate patterns at each time point depicted in Figure 9 into a single survival probability curve. This combination serves as a reference for comparing the Ext Cox PH survival probability curve with other models. The average survival probability curve that resulted from this process is shown in Figure 9.

The survival probability graph in Figure 10 is derived using the RBF model and displays the customers' probabilities of surviving. The RBF survival probability graph shows that the probability of a customer surviving is high from month zero to month 120, remains high for a long time, but gradually decreases as time progresses; a similar pattern is observed from the Kaplan–Meier survival probability graph.

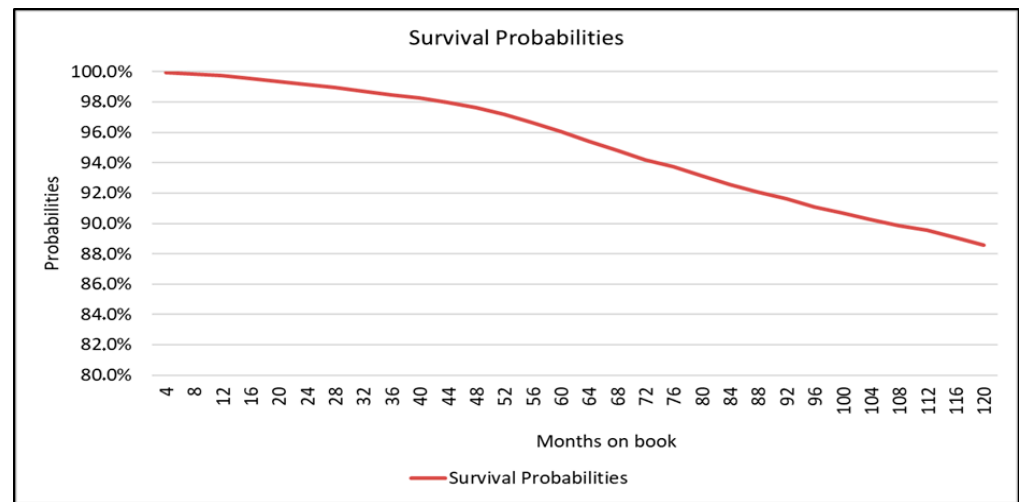


Figure 9. Average survival probability graph.

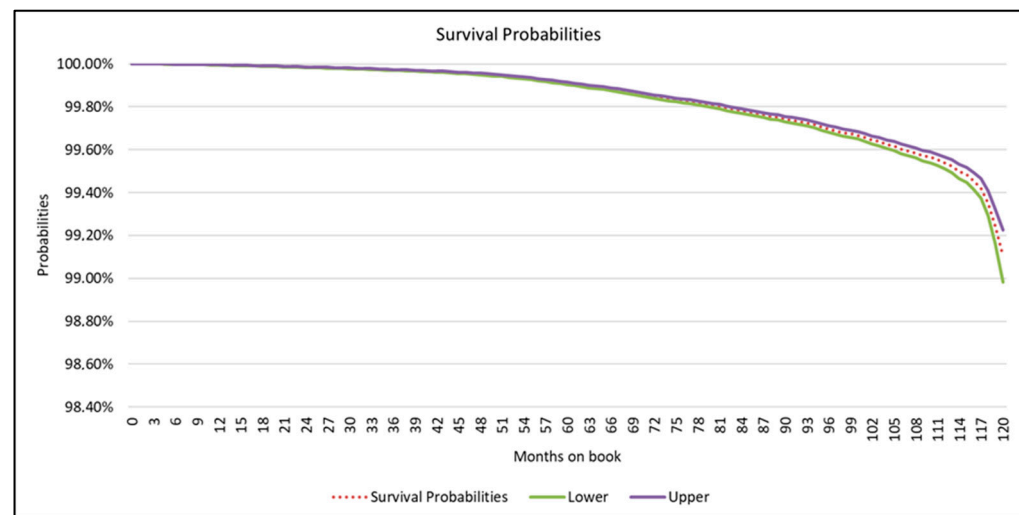


Figure 10. RBF model survival curve.

3.7. Cumulative Hazard Rate Graphs

The cumulative hazard rate graph outlines the customers cumulative default rate from month zero to month 120. The months on book on the horizontal axis indicate the time spent in follow-up, measured from 0 to 120 months, while the vertical axis indicates the probability that a customer will default, which is reflected in the cumulative hazard. The cumulative hazard rate graph in Figure 11 shows how the risk of experiencing the default, in this study, accumulates over time for the population under observation. The cumulative hazard rate is derived using the Kaplan–Meier methodology.

The cumulative hazard rate outlined in Figure 12 is derived using the Cox PH model. We observe a similar hazard rate trajectory from the Kaplan–Meier methodology, which has a cumulative hazard rate that gradually increases as the months on book increases. The difference is where the Ext Cox PH model total cumulative default rate increases below 10% while the Kaplan–Meier total increase is just above 10% default rate.

Figure 13 shows the cumulative hazard rate curve derived using the Ext Cox PH model, which reaches a maximum point of 12.7%. The cumulative hazard graph outlines how the hazard or risk accumulates over time. Each step corresponds to an event, and the size of the step reflects the impact of that event (default) on the overall risk.

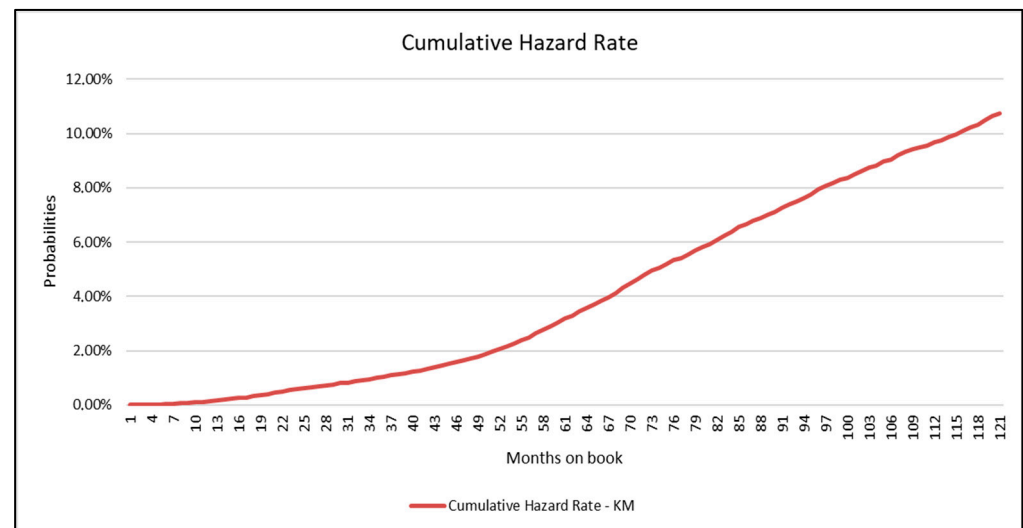


Figure 11. Kaplan–Meier cumulative hazard rate.

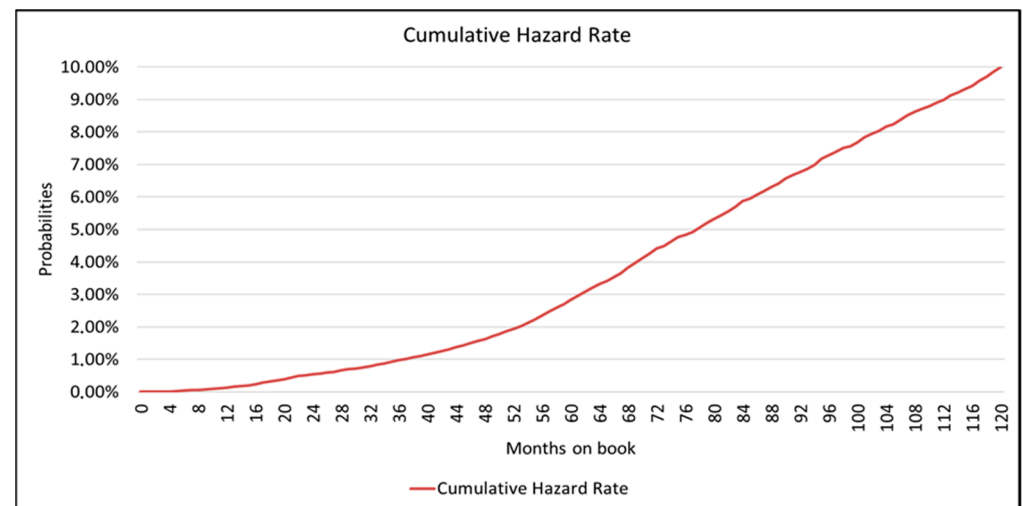


Figure 12. Standard Cox PH model cumulative hazard rate.

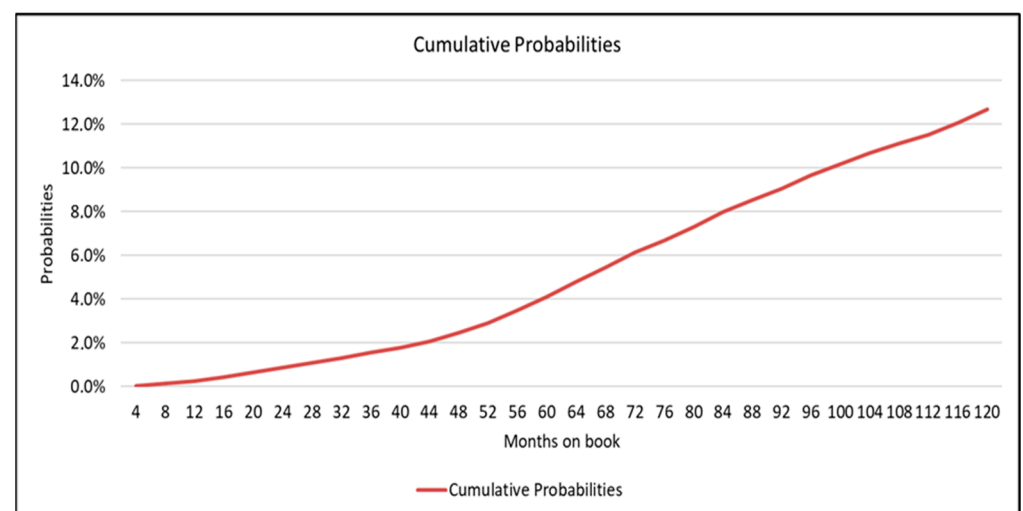


Figure 13. Ext Cox PH model cumulative hazard curve.

Figure 14 shows the cumulative hazard rate curve derived using the RBF model. The cumulative hazard graph gradually increases over time and indicates that the probabilities

of the customers defaulting increases as time passes. In more recent months, the slope of the graph becomes steeper, indicating that there was a higher concentration of defaults or a greater risk of events near the end of the follow-up period. The timing and intensity of risk throughout the follow-up period is demonstrated in Figure 14.

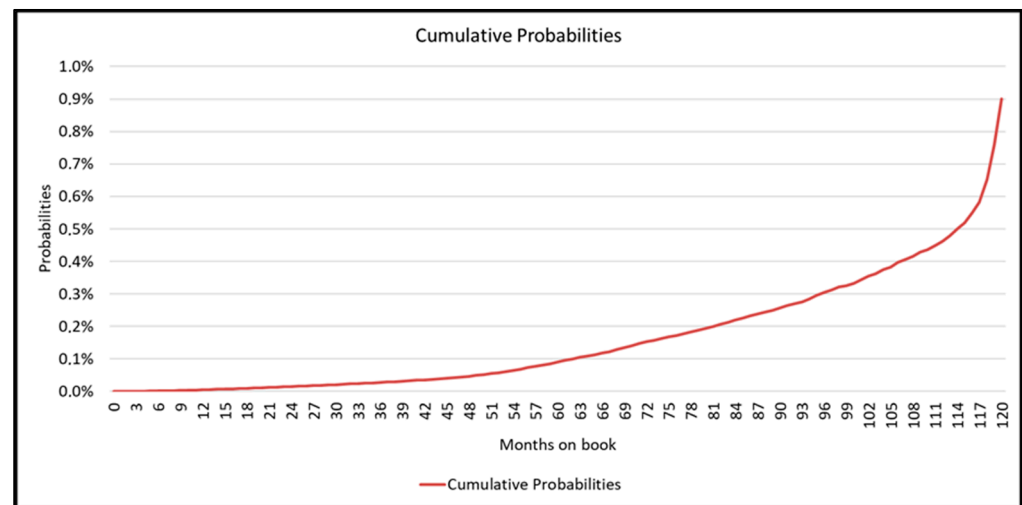


Figure 14. RBF model cumulative hazard curve.

3.8. Model Evaluation Metrics

The following tests: C-index, AIC, BIC, and CAIC, were done on the training dataset, and the Cox PH, Ext Cox PH and RBF model results are presented in Table 5. The results are for the final three models explored in this study. The confidence interval (CI) for each model's C-index is added to Table 5. The CI illustrates the range within each C-index operates, with the Cox PH model exhibiting a wider CI followed by RBF.

Table 5. Comparison results of the C-index, AIC, BIC, CAIC and AUC for the three methods.

Model	C-Index	CI (Lower)	CI (Upper)	AIC	BIC	CAIC	AUC
Cox PH	0.639	0.626	0.653	38,286.27	38,297.44	38,305.46	0.613
Ext Cox PH	0.761	0.749	0.774	45,834.02	45,868.63	45,886.97	0.746
RBF	0.780	0.765	0.795	65,123.81	65,129.61	65,137.61	0.508

Table 5 presents a comparative evaluation of the three models, Cox PH, Ext. Cox PH, and RBF, based on key performance metrics. The RBF and Ext. Cox PH all had a superior ability to discriminate between high and low risk, as evidenced by their C-indexes of 0.780 and 0.761, respectively; Cox PH had a C-index of 0.639. Although the Cox PH is more efficient due to having the lowest AIC = 38,286.27; BIC = 38,297.44; and CAIC = 38,305.46 values, the RBF and Ext. Cox PH demonstrated improved predictive capabilities; maybe this is due to the more complex nature associated with these two models. Therefore, even though the Cox PH had a lower ability to discriminate, it produced a greater level of statistical efficiency and provided a better generalization of unknown future data compared to the other two models. This trade-off between predictive accuracy and model complexity is particularly important to consider when attempting to produce hazard function estimates and interpretable survival curves for purposes of credit risk assessment.

The ROC curve demonstrates the model's ability to differentiate between the two classes of outcomes. The diagonal line represents the model's discrimination power meaning that if a model has no discriminating power the model is basically random guessing the outcome. Specificity indicates the percentage of true negatives or corrects non-predictions, while higher

specificity will have fewer false positives. Sensitivity measures how many true positives or correct predictions there are, while higher sensitivity would produce fewer false negatives.

The ROC curve in Figure 15 is derived using the Cox PH model. Predictive power is represented by the curve above the diagonal line in Figure 15 and the overall AUC for the Cox PH model is 0.613, which is good but not great.

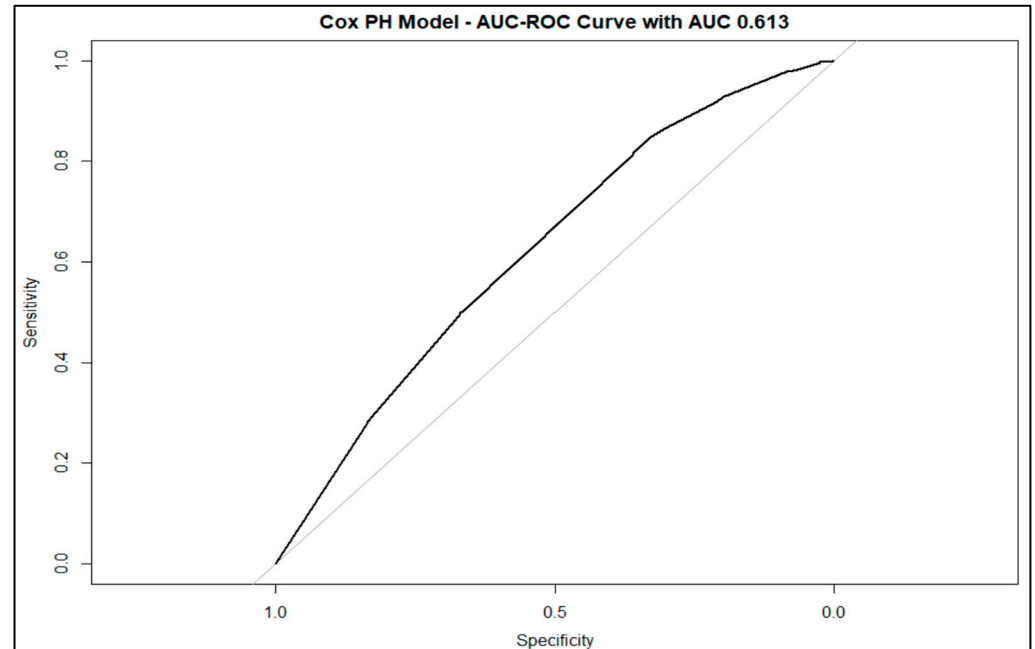


Figure 15. Standard Cox PH AUC-ROC curve.

The ROC curve in Figure 16 is derived using the Ext Cox PH model. The predictive accuracy is illustrated by any curve above the grey diagonal line as it moves to the upper left corner. The more pronounced this move, the more successful the Ext Cox PH model is in discriminating the two class types. The Ext Cox PH model is 0.746, meaning the Ext Cox PH model discriminates between classes well.

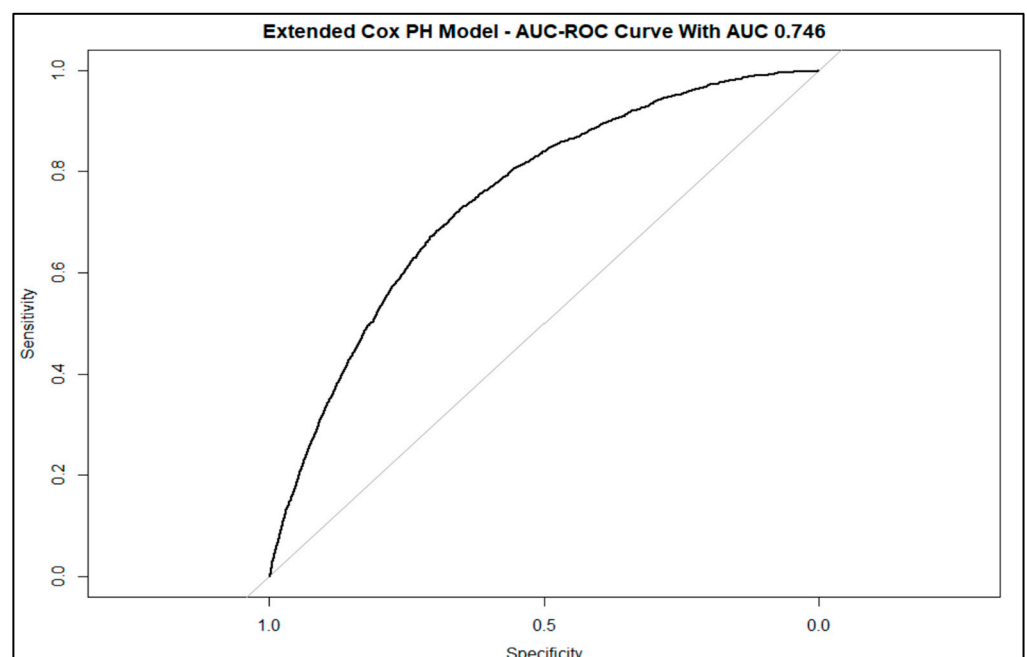


Figure 16. Ext Cox PH model AUC-ROC curve.

The grey diagonal line in Figure 17 represents that the model has no ability to discriminate between the two classes or is the same as guessing. A model with better predictive power will produce curves that lie above this grey diagonal line. A curve that is furthest from the grey diagonal line is said to be the ideal plot of the ROC curve, which implies that this is the best performance that the model achieved. When the line is below the horizontal line it means that the model is performing poorly in certain situations, resulting in model misclassification. Therefore, the RBF model with an AUC of 0.508 and an ROC curve below the grey line performed poorly.

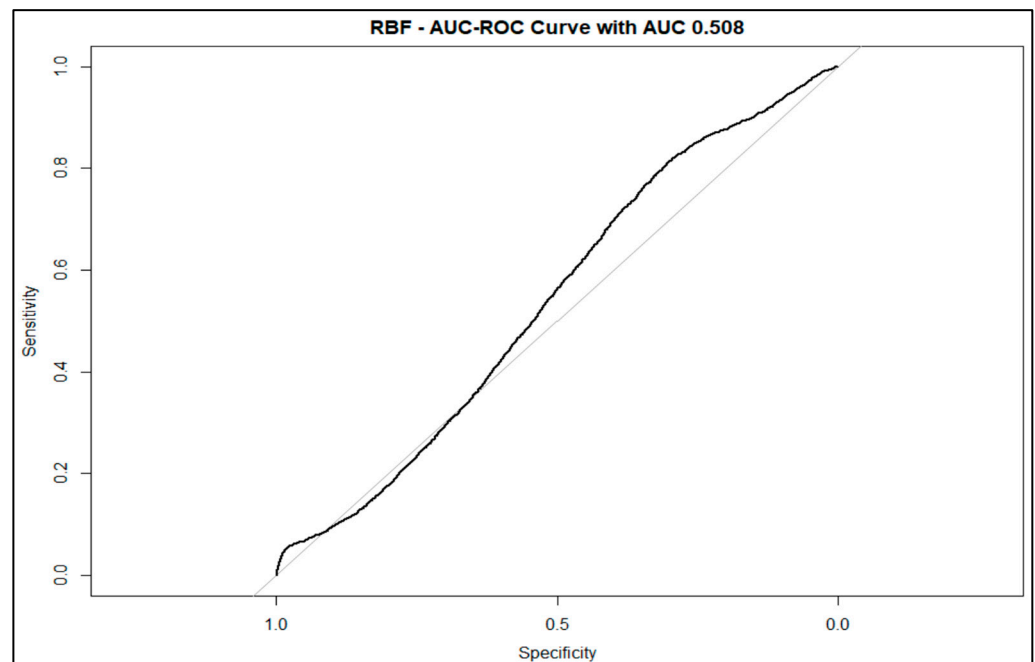


Figure 17. RBF AUC-ROC curve.

4. Discussion

A comparative analysis of the three models discussed above is presented in this section. All three models were evaluated based on key performance metrics, and the corresponding performance metric results were then used to determine which model performed well and which did not. Furthermore, the results provide insight into which aspects of the superior model made it perform well, thus assisting in understanding why that model is able to make accurate and robust predictions.

The number of months of book indicates the period that customers have been “on book.” These months can be anywhere from 4 to 120. The survival probabilities show the likelihood of a customer not defaulting at each month, which is between 80 and 100%. Of the models depicted in Figure 18, the RBF model consistently predicts that a customer will survive the longest. This may suggest that this model is more optimistic or less sensitive to risk than some of the other models. On the other hand, the Ext Cox model predicts that customers will default the earliest and therefore suggests that Ext Cox model provides a more cautious or risk-averse estimate. In between the two models is the Cox Model, which initially appears to be aligned with the Kaplan–Meier model but slowly diverges from that trajectory. The Kaplan–Meier model is the empirical benchmark because it was created strictly from observed data and not based upon model assumptions.

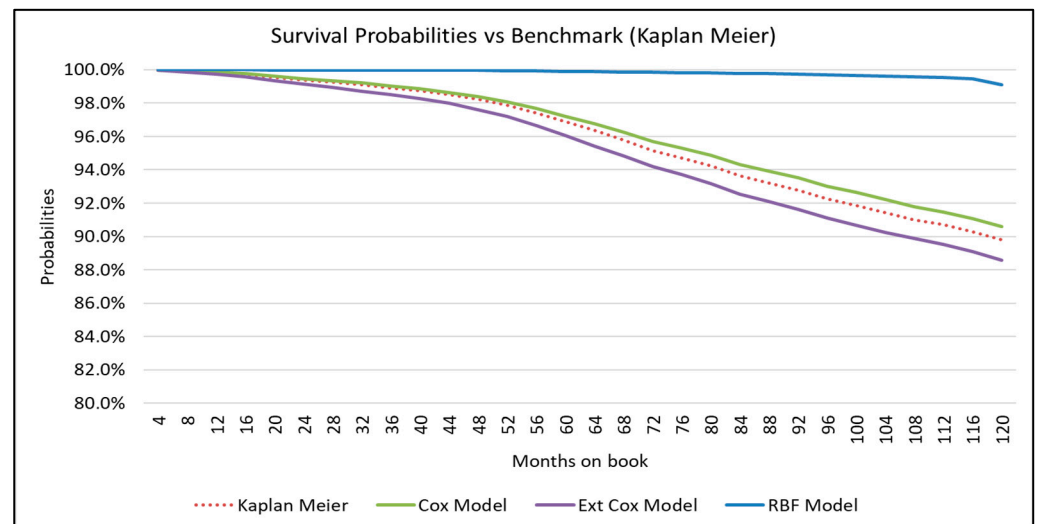


Figure 18. Comparison of survival probabilities.

Figure 19 shows that the Cox PH provides the most reliable and comprehensible survival trajectory. The Ext Cox PH maintains structural integrity while offering increased flexibility. Although the RBF model delivers the most expressive risk modeling, it does so at the expense of interpretability and potential overfitting.

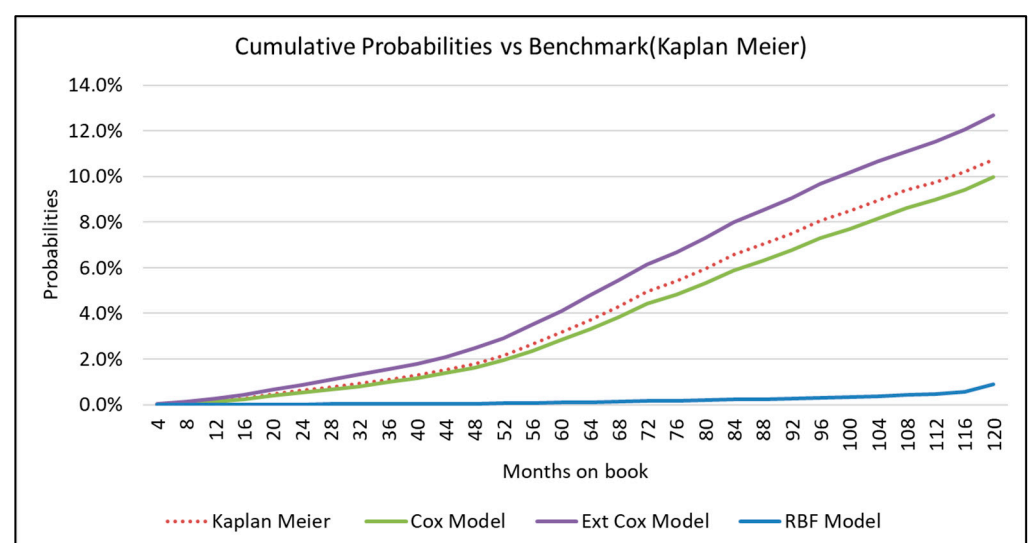


Figure 19. Comparison of cumulative probabilities.

5. Conclusions

The comparative analysis demonstrates a clear trade-off between a simple and a more complicated model that provides a greater degree of discriminatory predictive power. The following metrics (AIC, CAIC, BIC, C-index and AUC) are used to systematically identify the best performing modeling methodology. The RBF achieved a higher C-index of 0.780, indicating it was better able than either Cox's or Ext Cox's to accurately rank the survival time. Then, closely behind the RBF model, Ext Cox's model had a C-index of 0.761, which outperformed Cox's model, which had a C-index of 0.639. Thus, both the RBF and Ext Cox's models exhibited a high degree of power to discriminate between various risk groups. The model's discriminative power describes how well the predictive model can rank or classify individuals who are at risk.

The Cox PH model had the lowest AIC (38,286.27), BIC (38,297.44), and CAIC (38,305.46) of all the models, indicating that it provided the best fit of the three models. The RBF model had a much higher AIC (65,123.81) despite a higher discrimination ability, most likely due to overfitting and it being a more complex model. Model fit indicates how closely the predictions from a statistical or ML model align with the observed data and therefore how well the model describes the underlying data structure and relationships.

The Ext Cox PH model produced the best AUC results of 0.746. AUC indicates the optimal balance between sensitivity and specificity as compared to other models. The RBF model has a high C-index but a low AUC of only 0.508, which indicates that the model performs poorly at classifying events. Thus, the predictive accuracy can be viewed as a direct indication of how well the model is able to differentiate between individuals with and without the event occurring.

Therefore, the Ext Cox PH model produces the best combination of discrimination and predictive accuracy, so this is the most robust model for survival prediction in the context of IFRS 9 using the data presented. Due to it exhibiting the best combination of discrimination and predictive accuracy, the Ext Cox PH model is considered the best performing model from the three models explored in this study. It is important in the IFRS 9 to select the modeling methodology that can most accurately predict the customer's probability of default and the timing of that default. Underprediction of the model will result in the financial institution's under providing of their impairment amount while over prediction will result in the financial institution's over providing of the impairment amount. In the case of under providing, the financial institution will not have enough reserve to observe any significant impairment whereas over providing will reduce the financial institution's profit.

Whilst comparing the three models to the benchmark model which is the Kaplan–Meier, we observe that the Cox PH model and Ext Cox PH model closely track the Kaplan–Meier survival probability curve and the cumulative hazard rate curve. The RBF underperforms in tracking the benchmark model because it is not close to the benchmark model.

The RBF model has a strong ability to rank (as indicated by the C-index); however, it does not provide accurate calibration and lacks proper model fit. The RBF was applied under default settings which might be the reason for underperformance; maybe with better tuning we would have observed different results. The Cox PH model produces an efficient model with a good overall fit; however, it does not offer the same predictive strength as does the Ext Cox PH model. Simplicity is sometimes considered good, but in the IFRS 9 it is important to emphasize the prediction accuracy because prediction inaccuracy affects the financial institution's profit. The Cox is identified as the most time-efficient model because of its simplicity to develop and implement.

This is a data-driven study which means that the general results of the performance of the three methods are only applicable for this dataset on the condition that we follow the elimination of variables as stated in Appendix A as well as the Supplementary Material. Stated differently, it does not imply that the Ext Cox PH always has a better predictive performance than the Cox PH or RBF. Should a different method of subset selection of variables be used (see, for example, [Hastie et al., 2020](#)), then, the chances are, the three methods may vary in terms of their performance.

For future research, it would be valuable to include a parametric method and compare its performance with the methodologies applied under IFRS 9 using Freddie Mac data. Furthermore, applying the same three methodologies explored in this study to different datasets would help assess whether the overall conclusion remain valid in other contexts. It would also be a worthwhile endeavour to apply the principles and methodologies used in [Englund and Mostberg \(2022\)](#) to the Freddie Mac dataset and investigate whether their findings still hold. For future research, we will consider Lasso and forward stepwise

procedures for selecting variables (Hastie et al., 2020). Given that this research work mainly focused on RBF under machine learning discussion, for the future, readers may also evaluate the difference between the results obtained through the boosting versus the non-boosting methods under the IFRS 9 regulation (see, for instance, Cha et al., 2021, in the context of demolition waste). Similarly, assuming the IFRS 9 regulation and since the data from lenders can be formatted in structured tabular form, readers can follow the analytical approach illustrated in Stumpf and Shongwe (2026) for analyzing structured tabular data, where linear/logistic regression, support vector machines, XGBoost, and multi-layered perceptrons are applied for regression and classification purposes. We did not consider hyperparameter tuning for the RBF in this study; we used the default settings, which is considered one of the limitations of this study. We refer the reader to the recent paper by Stumpf and Shongwe (2026) where different machine learning techniques were compared under ‘default’ and ‘tuned’ settings however, for structured tabular data. Thus, such a study can be extended, in the context of research done here, as the following topic: “Machine learning techniques for ‘default’ vs. ‘tuned’ settings under the IFRS 9 regulation”. Since there is no literature that has attempted the modeling of real-life data using stacked regression, readers are referred to Ersoy et al. (2024) and Ahrens et al. (2022) for the general algorithm of this interesting approach. In addition, the DeLong’s test or bootstrapping can be considered as comparison metrics in future research.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijfs14030062/s1>. It contains information on how some of the variables were reduced from the models and the SAS and R codes used in this research work.

Author Contributions: Conceptualization, K.R.M., S.C.S. and F.F.K.; Methodology, K.R.M.; Software, K.R.M.; Validation, S.C.S. and F.F.K.; Formal analysis, K.R.M.; Investigation, K.R.M.; Resources, S.C.S. and F.F.K.; Data curation, K.R.M.; Writing—original draft, K.R.M.; Writing—review and editing, S.C.S. and F.F.K.; Supervision, S.C.S. and F.F.K.; Project administration, S.C.S.; Funding acquisition, S.C.S. and F.F.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in [Freddie Mac] at [<https://www.freddiemac.com/>], accessed on 30 March 2025.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

The origination dataset consists of 50,000 unique accounts, all of which were originated in 2004. The dataset contains 32 variables captured at the time of loan origination; the names and descriptions of the variables are provided in Table A1. The performance dataset comprises 3,974,583 rows of performance data and 50,000 unique accounts. It contains 32 variables captured over the duration of the loan and their names along with their descriptions are provided in Table A2. After the explanation in Supplementary Material S1, the final variables’ statistics are given in Table A3. The forest for the RBF model discussed in Section 3.5 is provided in Figure A1.

Table A1. Initial list of variables from the origination dataset.

Position	Attribute Name	Type
1	1_Credit Score	Numeric
2	2_FIRST PAYMENT DATE	Date
3	3_FIRST TIME HOMEBUYER FLAG	Alpha
4	4_MATURITY DATE	Date
5	5_METROPOLITAN STATISTICAL AREA (MSA)	Numeric
6	6_MORTGAGE INSURANCE PERCENTAGE (MI %)	Numeric
7	7_Number of Units	Numeric
8	8_OCCUPANCY STATUS	Alpha
9	9_ORIGINAL COMBINED LOAN-TO-VALUE (CLTV)	Numeric
10	10_ORIGINAL DEBT-TO-INCOME (DTI) RATIO	Numeric
11	11_Original UPB	Numeric
12	12_ORIGINAL LOAN-TO-VALUE (LTV)	Numeric
13	13_ORIGINAL INTEREST RATE	Numeric Literal decimal
14	14_Channel	Alpha
15	15_PREPAYMENT PENALTY MORTGAGE (PPM) FLAG	Alpha
16	16_AMORTIZATION TYPE	Alpha
17	17_PROPERTY STATE	Alpha
18	18_PROPERTY TYPE	Alpha
19	19_POSTAL CODE	Alpha
20	20_LOAN SEQUENCE NUMBER	Alpha-numeric
21	21_Loan Purpose	Alpha
22	22_Original Loan Term	Numeric
23	23_Number of Borrowers	Numeric
24	24_SELLER NAME	Alpha-numeric
25	25_SERVICER NAME	Alpha-numeric
26	26_SUPER CONFORMING FLAG	Alpha
27	27_PRE-RELIEF REFINANCE LOAN SEQUENCE NUMBER	Alpha-numeric
28	28_SPECIAL ELIGIBILITY PROGRAM	Alpha-numeric
29	29_RELIEF REFINANCE INDICATOR	Alpha
30	30_PROPERTY VALUATION METHOD	Numeric
31	31_INTEREST ONLY INDICATOR (I/O INDICATOR)	Alpha
32	32_MI CANCELLATION INDICATOR	Alpha Numeric

Table A2. Initial list of variables from the monthly performance dataset.

Position	Attribute Name	Data Type and Format	Max Length
1	1_Loan Sequence Number	Alpha Numeric	12
2	2_Monthly Reporting Period	Date	6
3	3_Current Actual UPB	Numeric—12,2	12
4	4_Current Loan Delinquency Status	Alpha Numeric	3
5	5_Loan Age	Numeric	3
6	6_Remaining Months to Legal Maturity	Numeric	3
7	7_Defect Settlement Date	Date	6
8	8_Modification Flag	Alpha	1
9	9_Zero Balance Code	Numeric	2
10	10_Zero Balance Effective Date	Date	6
11	11_Current Interest Rate	Numeric—8,3	8
12	12_Current Deferred UPB	Numeric	12
13	13_Due Date of Last Paid Installment (DDLPI)	Date	6
14	14_MI Recoveries	Numeric—12,2	12
15	15_Net Sales Proceeds	Alpha-Numeric	14
16	16_Non MI Recoveries	Numeric—12,2	12
17	17_Expenses	Numeric—12,2	12
18	18_Legal Costs	Numeric—12,2	12
19	19_Maintenance and Preservation Costs	Numeric—12,2	12
20	20_Taxes and Insurance	Numeric—12,2	12
21	21_Miscellaneous Expenses	Numeric—12,2	12
22	22_Actual Loss Calculation	Numeric—12,2	12
23	23_Modification Cost	Numeric—12,2	12
24	24_Step Modification Flag	Alpha	1
25	25_Deferred Payment Plan	Alpha	1
26	26_Estimated Loan-to-Value (ELTV)	Numeric	4
27	27_Zero Balance Removal UPB	Numeric—12,2	12
28	28_Delinquent Accrued Interest	Numeric—12,2	12
29	29_Delinquency Due to Disaster	Alpha	1
30	30_Borrower Assistance Status Code	Alpha	1
31	31_Current Month Modification Cost	Numeric—12,2	12
32	32_Interest Bearing UPB	Numeric—12,2	12

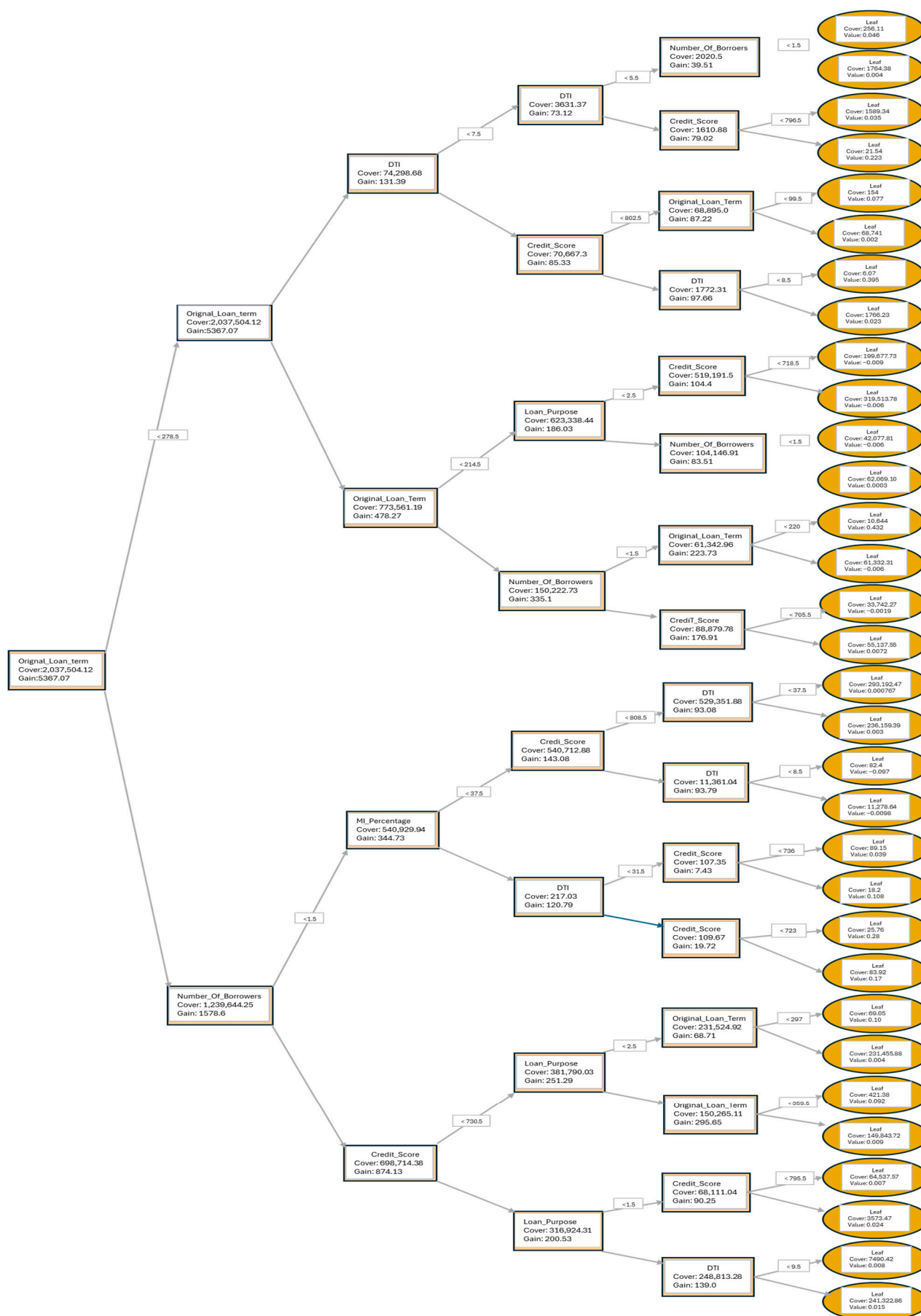
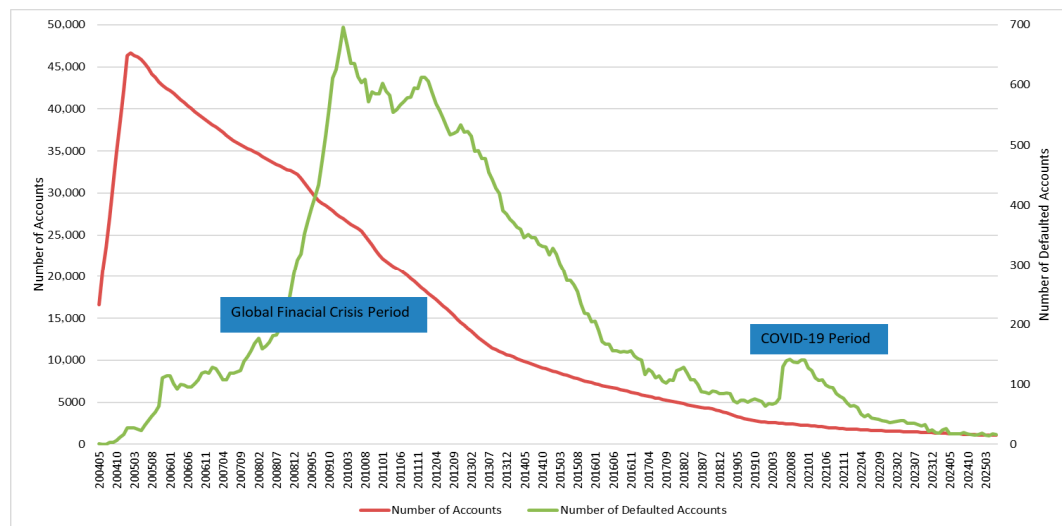


Figure A1. Forest for the RBF model.

Table A3. Variables considered in the final models (Cox PH, Ext Cox PH and RBF).

Variable	Frequency	Mean	Std Dev	Min	Max
Original Loan Term	3,357,837	292.77	86.84	60	361
Loan Purpose	3,357,837	2.06	0.80	1	3
Number of Borrowers	3,357,837	1.63	1.59	1	99
Credit Score	3,357,837	733.02	268.92	300	9999
Mortgage Insurance%	3,357,837	3.14	8.65	0	999
DTI	3,357,837	49.88	127.01	1	999

The Freddie Mac data we use includes the global financial crisis period (i.e., around 2007 to 2009) as well as the coronavirus disease (COVID-19) period (i.e., late 2019 to 2021). Figure A2 shows that there is a significant increase in the number of defaults during the US financial crisis and a minor increase in the number of defaults during the COVID-19 period.

**Figure A2.** Number of accounts and defaulted accounts over time.

References

- Ahrens, A., Ersoy, E., Iakovlev, V., Li, H., & Schaffer, M. E. (2022). An introduction to stacking regression for economists. In S. Sriboonchitta, V. Kreinovich, & W. Yamaka (Eds.), *Credible asset allocation, optimal transport methods, and related topics*. TES 2022. *Studies in systems, decision and control* (Vol. 429). Springer. [\[CrossRef\]](#)
- Banasik, J., Crook, J. N., & Thomas, L. C. (1999). Not if but when will borrowers default. *The Journal of the Operational Research Society*, 50(12), 1185–1190. [\[CrossRef\]](#)
- Bank, M., & Eder, B. (2021). A review on the probability of default for IFRS 9. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3981339 (accessed on 25 July 2025). [\[CrossRef\]](#)
- Bellotti, T., & Crook, J. (2009). Credit scoring with macroeconomic variables using survival analysis. *Journal of the Operational Research Society*, 60(12), 1699–1707. [\[CrossRef\]](#)
- Bone-Winkel, G. F., & Reichenbach, F. (2024). Improving credit risk assessment in P2P lending with explainable machine learning survival analysis. *Digit Finance*, 6, 501–542. [\[CrossRef\]](#)
- Botha, A., & Verster, T. (2025). Approaches for modelling the term-structure of default risk under IFRS 9: A tutorial using discrete-time survival analysis. *arXiv*, arXiv:2507.15441. [\[CrossRef\]](#)
- Brunel, V. (2016). *Lifetime PD analytics for credit portfolios: A survey*. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2857183 (accessed on 30 July 2025). [\[CrossRef\]](#)
- Caselli, S., Corbetta, G., Cucinelli, D., & Rossolini, M. (2021). A survival analysis of public guaranteed loans: Does financial intermediary matter? *Journal of Financial Stability*, 54, 100880. [\[CrossRef\]](#)

- Cha, G.-W., Moon, H.-J., & Kim, Y.-C. (2021). Comparison of random forest and gradient boosting machine models for predicting demolition waste based on small datasets and categorical variables. *International Journal of Environmental Research and Public Health*, 18(16), 8530. [CrossRef] [PubMed]
- Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187–220. [CrossRef]
- Englund, H., & Mostberg, V. (2022). *Probability of default term structure modelling: A comparison between machine learning and Markov chains* [Master's thesis, Umeå University (Sweden)]. Available online: <https://www.diva-portal.org/smash/get/diva2:1667201/FULLTEXT03> (accessed on 30 March 2025).
- Ersoy, E., Li, H., Schaffer, M. E., & Szendrei, T. (2024). Stacking regression for time-series, with an application to forecasting quarterly US GDP growth. In T. N. Ngoc, V. Kreinovich, D. T. Ha, & N. D. Trung (Eds.), *Optimal transport statistics for economics and related topics. studies in systems, decision and control* (Vol. 483). Springer. [CrossRef]
- Ertan, C., & Gansmann, A. (2015). *A semi-parametric probability of default model* [Master's dissertation, Stockholm School of Economics]. Available online: <https://arc.hhs.se/download.aspx?MediumId=2818> (accessed on 31 July 2025).
- George, B., Seals, S., & Aban, I. (2014). Survival analysis and regression models. *Journal of Nuclear Cardiology*, 21(4), 686–694. [CrossRef]
- Ghosal, I., & Hooker, G. (2020). Boosting random forests to reduce bias; one-step boosted forest and its variance estimate. *Journal of Computational and Graphical Statistics*, 30(2), 493–502. [CrossRef]
- Hardin, P., & Ingre, R. (2021). *BNPL probability of default modelling including macroeconomic factors: A supervised learning approach*. KTH Royal Institute of Technology, School of Engineering Sciences. Available online: <https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-307334> (accessed on 30 June 2025).
- Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer. Available online: <https://link.springer.com/book/10.1007/978-0-387-84858-7> (accessed on 30 June 2025).
- Hastie, T., Tibshirani, R., & Tibshirani, R. (2020). Best subset, forward stepwise or lasso? Analysis and recommendations based on extensive comparisons. *Statistical Science*, 35(4), 579–592. [CrossRef]
- Herrmann, M., Probst, P., Hornung, R., Jurinovic, V., & Boulesteix, A. L. (2021). Large-scale benchmark study of survival prediction methods using multi-omics data. *Briefings in Bioinformatics*, 22(3), bbaa167. [CrossRef]
- International Accounting Standards Board. (2014). *International financial reporting standard 9 financial instruments*. International Accounting Standards Board. Available online: <https://www.ifrs.org/issued-standards/list-of-standards/ifrs-9-financial-instruments/> (accessed on 30 June 2025).
- International Financial Reporting Standards Foundation. (2011). *International accounting standard 39 financial instruments: Recognition and measurement*. Available online: <https://www.ifrs.org/issued-standards/list-of-standards/ias-39-financial-instruments-recognition-and-measurement/#:~:text=IAS%2039%20establishes%20principles%20for,instruments%20and%20for%20hedge%20accounting> (accessed on 30 June 2025).
- James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). Statistical learning. In *An introduction to statistical learning: With applications in Python*. Springer. [CrossRef]
- Kleinbaum, D. G., & Klein, M. (2012). *Survival analysis: A self-learning text* (3rd ed.). Springer. [CrossRef]
- Kochański, B. (2022). Which curve fits best: Fitting ROC curve models to empirical credit-scoring data. *Risks*, 10(10), 184. [CrossRef]
- Kottayil, N. M., Jailia, M., & Verma, S. (2025). Machine learning approach for PD term structure modeling under IFRS 9 regulatory framework. *International Journal of Science and Research*, 14(4), 1115–1119. [CrossRef]
- Kraus, A., & Küchenhoff, H. (2014). Credit scoring optimization using the area under the curve. *Journal of Risk Model Validation*, 8(1), 31–67. [CrossRef]
- Kulinskaya, E. (2019). *Modelling non-proportional hazards: Time-dependent coefficients, parametric “double Cox” regression and landmark analysis*. University of East Anglia. Available online: https://vle.actuaries.org.uk/pluginfile.php/148705/mod_resource/content/1/Plenary%20Elena%20Kulinskaya.pdf (accessed on 30 March 2025).
- Laeven, L., & Majnoni, G. (2003). Loan loss provisioning and economic slowdowns: Too much, too late? *Journal of Financial Intermediation*, 12(2), 178–197. [CrossRef]
- Lin, H., & Zelterman, D. (2002). Modeling survival data: Extending the Cox model. *Technometrics*, 44(1), 85–86. [CrossRef]
- Ludwig, T., & Bank, M. (2019). *Markov Chain based PD term structure modelling in an IFRS 9 framework* [Master's dissertation, Department of Banking and Finance, University of Innsbruck]. Available online: <https://diglib.uibk.ac.at/ulbtirolhs/download/pdf/4347915?originalFilename=true> (accessed on 30 June 2025).
- Mahesh, B. (2020). Machine learning algorithms—A review. *International Journal of Science and Research*, 9(1), 381–386. [CrossRef]
- Narain, B. (2004). Chapter 16: Survival analysis and the credit granting decision. In L. C. Thomas, D. B. Edelman, & J. N. Crook (Eds.), *Readings in Credit Scoring: Foundations, developments, and aims* (online edition). Oxford Academic. [CrossRef]
- Novotny-Farkas, Z. (2015). The significance of IFRS 9 for financial stability and supervisory rules. *EPRS: European Parliamentary Research Service*. Available online: <https://coilink.org/20.500.12592/0kn91j> (accessed on 30 March 2025).

- Ozili, P. K., & Outa, E. R. (2017). Bank loan loss provisions research: A review. *Borsa Istanbul Review*, 17(3), 144–163. [CrossRef]
- Ptak-Chmielewska, A., & Gonzalez, J. P. E. (2024). Default prediction using the cox regression model and macroeconomic conditions—A lifetime perspective. *Econometrics. Ekonometria. Advances in Applied Data Analytics*, 28(2), 50–61. [CrossRef]
- Saavedra, C. A. P. B., Fachini-Gomes, J. B., de Castro Gomes, E. M., & Kimura, H. (2024). Probability of default for lifetime credit loss for IFRS 9 using machine learning competing risks survival analysis models. *Expert Systems with Applications*, 249, 123607. [CrossRef]
- Santos, B. L. (2018). *Practical approach for probability of default estimation under IFRS 9* [Doctoral dissertation, Instituto Superior de Economia e Gestão]. Available online: <https://repositorio.ulisboa.pt/bitstream/10400.5/17350/1/DM-BLS-2018.pdf> (accessed on 3 September 2025).
- Smuts, M., & Allison, J. (2020). An overview of survival analysis with an application in the credit risk environment. *ORioN*, 36(2), 89–110. [CrossRef]
- Stepanova, M., & Thomas, L. (2002). Survival analysis methods for personal loan data. *Operations Research*, 50(2), 277–289. [CrossRef]
- Stumpfe, S. F., & Shongwe, S. C. (2026). Comparative analysis and optimisation of machine learning models for regression and classification on structured tabular datasets. *Mathematics*, 14(3), 473. [CrossRef]
- Tibshirani, R. (2022). What is Cox’s proportional hazards model? *Significance*, 19(2), 38–39. [CrossRef]
- Turkson, A. J., Ayiah-Mensah, F., & Nimoh, V. (2021). Handling censoring and censored data in survival analysis: A standalone systematic literature review. *International Journal of Mathematics and Mathematical Sciences*, 2021(1), 9307475. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.