Mathematical Modelling and Applications
2026, Vol. 11, No. 4, pp. 90–97
https://doi.org/10.11648/j.mma.20261104.11

SciencePG
Science Publishing Group

Research Article

# A Modelling Consumer Credit Risk Using Survival Analysis: A District-level Study from Badulla, Sri Lanka

**Rajapaksha Pathira Gamage Amithani[1, *], Chathuri Lakshika Jayasinghe[2], Miuran Dencil Jayaweera[1]**

[1]Department of Mathematics, University of Moratuwa, Moratuwa, Sri Lanka

[2]Department of Statistics, University of Sri Jayewardenepura, Nugegoda, Sri Lanka

## Abstract

Credit risk models in many rural lending settings are still largely static, focusing on whether default occurs rather than when it occurs. This study develops a time-to-event (survival) framework to analyze the timing and determinants of consumer-loan default in the Badulla District of Sri Lanka. Using microdata on consumer instalment loans originated between January 2015 and December 2020 (followed up to end-2020), loans were modelled in months from loan origination to first default (≥90 days past due), early full settlement, or right-censoring. Kaplan–Meier estimates and log-rank tests were used to describe overall loan survival and compare survival patterns across borrower and loan groups. Kaplan–Meier results show high early survival at 12 and 36 months (S (12) =0.943, S (36) =0.938) followed by a pronounced deterioration after four years, with survival falling from S (48) =0.923 to S (60) =0.343. A multiple Cox proportional hazards model was estimated using 1,168 loans. Education was a consistent determinant: compared with higher education, primary, secondary, and tertiary (non-degree) groups had significantly higher hazards of default. Higher interest rates and lower income categories were associated with elevated default hazards. Model diagnostics suggested that the proportional hazards assumption was broadly reasonable for most covariates, but collateral exhibited time dependence. Furthermore, the findings highlight a critical risk window beyond 48 months and support strengthening monitoring, borrower support, and risk-based pricing for vulnerable groups in rural Sri Lankan credit portfolios.

## Keywords

Credit Risk, Survival Analysis, Kaplan–Meier, Cox Proportional Hazards, Loan Default

## 1. Introduction

Consumer credit is a major role in supporting household welfare by enabling consumption smoothing, meeting emergency needs, financing education and health expenditure, and sustaining small scale enterprise activity. These are especially important in rural and semi-rural economies where income is often seasonal and uncertain. At the same time, the expansion of consumer lending exposes financial institutions to heightened credit risk, particularly when borrowers face unemployment shocks, illness-related costs, price instability or sudden declines in household cash flows [1]. According to this situation accurate credit risk modeling is essential not only for profitability and portfolio quality. Furthermore, safeguarding

*Correspondence: Rajapaksha Pathira Gamage Amithani (amithani94@gmail.com)

**Received:** 5 August 2026; **Accepted:** 17 August 2026; **Published:** 6 November 2026

Copyright: © The Author(s), 2026. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

institutional resilience and supporting broader financial stability. Within Sri Lanka, district level household credit dynamics are influenced by local labor conditions, informal employment, agricultural livelihoods, and macroeconomic pressures [2]. Badulla district provides a particularly relevant setting for district level credit risk analysis due to its socio-economic structure and reliance on income sources that may fluctuate over time. Therefore, understanding consumer credit risk in such environments requires models that represent repayment beavior realistically and support early intervention strategies.

In many lending institutions, credit risk assessment has traditionally been approached through "static" classification methods that predict whether default will occur within a fixed horizon. Thomas et.al 2002 show that logistic regression scorecards remain one of the most widely used tools because they are interpretable and operationally convenient, offering estimated probabilities of default based on borrower demographics and loan characteristics [3]. However, static models typically treat default s a binary endpoint and ignore the timing of default events, even though the period during which default occurs is highly relevant for lenders. Moreover, consumer loan portfolio commonly includes a substantial proportion of accounts that remain active at the end of observation, resulting in right-censoring that cannot be properly handled within standard binary frameworks [4]. When time is ignored, loans that are still performing but have not yet completed their term are implicitly treated the same as fully repaid loans, which can distort the practical usefulness of static credit scoring for monitoring portfolios over time and for identifying critical risk windows during a loan's lifecycle.

Survival analysis provides a more realistic framework for consumer credit risk because it explicitly models the time until an event occurs while properly accommodating censored observations [15]. The Kaplan-Meier estimator is a broadly used non-parametric technique for describing survival patterns and estimating the probability that an account remains non-default beyond a given time point [5]. In credit risk setting, Kaplan-Meier survival curves allow lenders to compare default timing across borrower groups and loan segments. Furthermore, log-rank tests offer a statistical basis for evaluating whether survival experience differ between groups [6]. This descriptive evidence is valuable because it highlights whether default risk accumulates steadily or clusters in specific periods, supporting targeted monitoring. However, descriptive survival models alone do not quantify the simultaneous contribution of multiple predictors. For that reason, regression-based survival models have become increasingly important for applied credit risk research.

Among survival regression methods, the cox proportional hazards model is especially influential because it estimates how covariates shift the hazard of default without requiring a strict parametric form for the baseline hazard function [7]. The model produces interpretable hazard ratios that quantify how borrower characteristics and loan design variables increase or decrease the instantaneous risk of default. Cox based approaches have been broadly adopted in credit risk research because they support multivariable inference and allow institutions to translate model outputs into actionable rules for monitoring and risk segmentation [4]. Nevertheless, a key assumption of the cox proportional hazards model is that covariate effects remain constant over time. In practice, repayment behavior can evolve across the loan lifecycle due to changes in borrower, motivation, exposure to economic shocks, financial stress, meaning that proportional hazards may not hold. Diagnostic methods based on Schoenfeld residuals provide formal tools for testing this assumption and identifying covariates whose effects vary overtime [8, 13]. Where non-proportionality exists, time varying covariate effects can be introduced to maintain model validity and improve interpretation.

A further complexity in consumer credit portfolios is not the only possible loan outcome. Many accounts may exit through early full settlement or prepayment, which removes a borrower from the risk set and competes with default as a terminal event. If early settlement is treated as simple censoring, standard survival models may overestimate the probability of default. Because the analysis implicitly assumes that competing events do not alter the default process. Competing risks methods explicitly address this issue by distinguishing the probability of default from the probability of alternative outcomes over time [9]. In particular, the cumulative incidence function provides a direct estimate of the chance that default occurs by time $t$ in the presence of competing events, and regression approaches such as sub distribution hazard modeling offer a coherent method for covariate-based inference under competing risks [10]. Incorporating competing risks is particularly important for credit portfolios. Because early settlement may reflect either improved repayment capacity, borrower strategy refinancing behavior, or institutional repayment incentives, each of which can meaningfully change observed default patterns.

Empirical evidence consistently demonstrates that credit risk is shaped by borrower socio-economic characteristics and loan design attributes [3, 14]. Variables such as employment stability, education, income capacity are frequently associated with repayment performance, as they capture both ability and willingness to repay and reflect the borrower's resilience to financial shocks. Loan factors such as term length, role of collateral requirements, and the repayment burden are also important because they influence affordability and strategic default behavior. Collateral requirements in particular are widely used as a risk mitigation mechanism because they reduce loss severity and can strengthen borrower discipline. In some portfolios, collateral may be highly protective during early periods but less influential later when financial stress becomes persistent. This possibility highlights the need for models that test time stability rather than assume it.

Despite the relevance of these methodological developments, evidence from low- and middle-income settings re-

mains comparatively limited, and district level studies are especially scare. In Sri Lanka, many lending institutions continue to rely on arrears based monitoring and traditional scorecard approaches, while loan level time to time event modeling is less commonly applied. This creates an important gap in both practice and research. Because static methods may fail to capture when risk accelerates, and thus may delay interventions that could prevent arrears from progressing into default. District level analysis is particularly valuable because local economic structures and income cycles can strongly influence repayment timing, meaning that risk patterns observed nationally my not adequately represent rural districts [11].

This study addresses this gap by adopting a time to event framework to model consumer credit risk in Badulla district using loan level microdata. Default is defined in line with institutional standards based on severe delinquency. While early full settlement is treated as a competing outcome, and loan that remain active by the end of follow up are treated as right censored. The analysis begins with Kaplan-Meier survival estimation and log rank comparisons identify unadjusted differences in loan survival across borrower and loan segments. Furthermore, applies cox proportional hazard modeling to estimate multivariable effects on default timing, supported by diagnostics for proportional hazards and appropriate adjustments through time-varying covariate effects when necessary [8]. Competing risks techniques are further used to provide a realistic separation between default and early settlement processes [10]. Moreover, a logistic regression model is estimated as a benchmark, enabling direct comparison between conventional static credit scoring and survival-based approaches [6, 15]. By integrating these methods, the study provides timing aware evidence on default risk and generates practical insights that can strengthen early warning systems, improve borrower segmentation, and support risk-based portfolio decision in rural Sri Lankan consumer credit.

## 2. Methodology

This study used a quantitative research approach with a retrospective cohort design to model consumer credit risk as a time to event process. The analysis was conducted for Badulla district, Sri Lanka. Also using secondary loan level data obtained from a regulated financial institution. The dataset included consumer instalment loans issued from January 2015 to December 2020 and observed until the end of 2020. The sample unit was an individual loan account. After cleaning and validation, the final analytical sample consisted of 1 168 loans, including 281 defaults, 20 early full settlements, and 867 right censored loans. The event of interest was loan default, defined as $\geq 90$ days past due according to institutional standards. Early full settlement was treated as a competing event, and loans that did not experience either outcome within the study period were treated as right censored. Time to event was measured in months, calculated from loan disbursement due to the earliest of default, early settlement, or the study end date.

Descriptive analysis was performed to summarize borrower and loan characteristics. Default timing patterns were initially examined using the Kaplan-Meier estimator and survival curves were compared across key borrower using the log rank test. Multivariable default risk was modeled using the cox proportional hazards model [7] expressed as:

$$h_i(t) = h_0(t)\exp(\beta^\top X_i) \quad (1)$$

Where $h_i(t)$ is the hazard of default for loan $i = 1,2,\ldots,1168$ at time $t, h_0(t)$ is the baseline hazard, and $X_i$ represents borrower and loan covariates such as gender, employment status, education level etc. The proportional hazards assumption was assessed using Schoenfeld residual based diagnostics [8]. When non proportionality was detected, time varying effects were incorporated using interactions with log time. The cox model assumes linearity of the log hazard with respect to continuous covariates. However, in this study, the covariates were categorical into ordered categories. Therefore, the linearity assumption for continuous predictors was not applicable and did not require formal assessment. Because early settlement removes loans from default exposure, competing risk patterns were evaluated using cumulative incidence functions to estimate default probability in the presence of early settlement [10]. For comparison with conventional credit scoring, a binary logistic regression model was estimated as a benchmark. All statistical analyses were performed using the R statistical environment.

## 3. Results

The Kaplan-Meier estimator was applied to assess the probability that loans remain active over time while accounting for right censored observations. Survival remained high during the first three years, with estimates of 0.943 at 12 months, 0.941 at 24 months, and 0.938 at 36 months, indicating minimal early-stage defaults. A gradual decline was observed at 48 months (0.923). However, the most pronounced deterioration occurred between 48 and 60 months, where survival dropped sharply to 0.343, highlighting a critical default risk window during the fourth and fifth years. Beyond this period, survival continued to decline rapidly, reaching 0.068 at 84 months, and approaching zero by 120 months. The Kaplan-Meier curve in Figure 1 visually confirms these patterns, showing relatively stable early phase followed by a steep decline after approximately 48 months, suggesting that portfolio monitoring and intervention should be strengthened before loan enter the mind to late term risk phase.

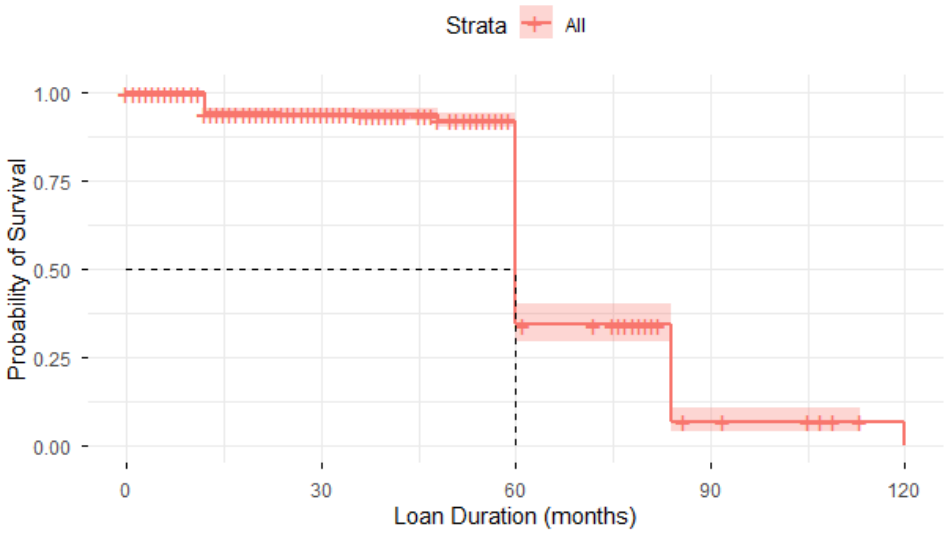



***Figure 1.*** *Kaplan–Meier survival curve showing loan survival probability over time.*

Kaplan-Meier survival curves were estimated to examine whether loan survival differs across key borrower and loan characteristics. Visual inspection of the curves showed that group differences were minimal during the first 36 months, but separation became clear after approximately 48 months, indicating that borrower heterogeneity influences default risk mainly in the mid-to-late stages of the loan life.

Survival differed significantly across income groups (p=0.020), here the low-income group showed the steepest decline beyond 48 months, while the high-income group maintained the highest survival throughout follow up. Survival also differed by employment status with borderline statistical evidence (p=0.050), suggesting that borrowers without stable income sources experience weaker repayment durability in later periods. Education level demonstrated significant group differences (p=0.034), with borrowers in the higher education category showing the strongest long-term survival and those with primary education showing the weakest.

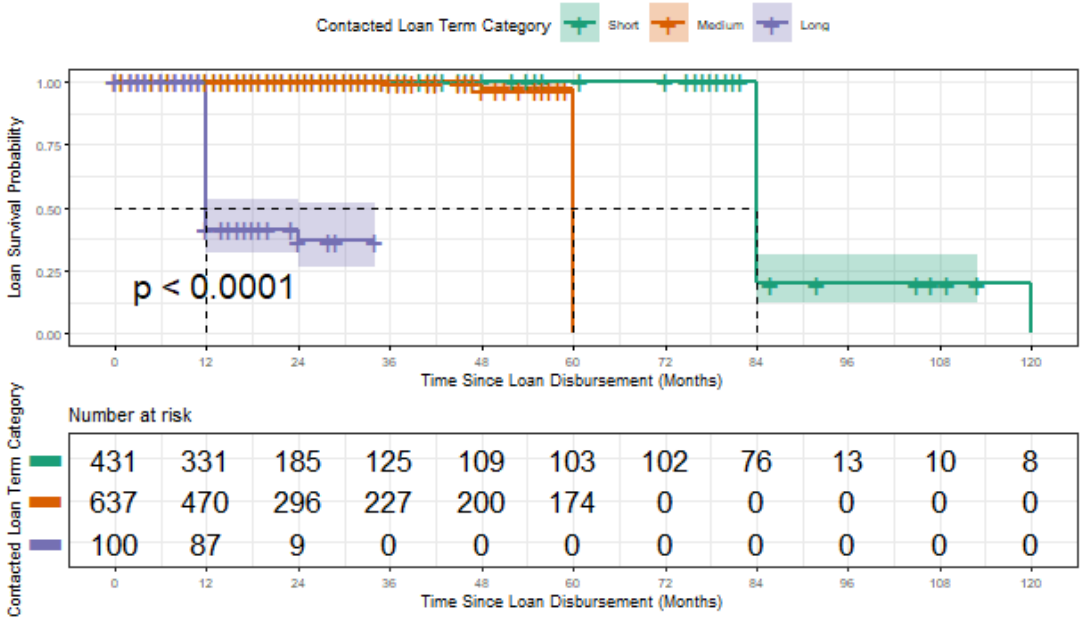



***Figure 2.*** *Kaplan–Meier survival curves comparing loan survival across contracted loan terms.*

Contracted loan term showed the strongest survival separation among factors (p<0.0001), confirming that repayment outcomes vary substantially across loan maturity structures. Therefore, collateral status was one of the most influential risk discriminators (p<0.0001). secured loans maintained substantially higher survival across the loan horizon, while unsecured loans displayed early and mid-term default concentration. These results indicate that income capacity, education, loan

term, and collateral are major factors for long-run repayment survival, and their effects become most visible after the fourth year of loan duration.

***Table 1.*** *Log-rank tests for Kaplan–Meier survival differences.*

| Factor | Groups | p-value |
|---|---|---|
| Income group | Low / Medium / High | 0.020 |
| Employment status | Formal / Informal / Self-employed / Unemployed | 0.050 |
| Education level | Higher / Primary / Secondary / Tertiary | 0.034 |
| Contracted loan term | Short / Medium / Long | <0.0001 |
| Collateral status | No / Yes | <0.0001 |

Default and early settlement were treated as competing terminal outcomes, with ongoing loans treated as right censored. Most loans were censored (867; 74.2%), also 281 (24.1%) defaulted and 20 (1.7%) were settled early. Cumulative incidence functions showed that early settlement was rare, reaching only about 4.1% by 60 months. In contrast, default increased sharply after 48 months, incising to approximately 61.6% by 60 months and continuing to increase at longer durations. These results confirm that default is the dominant event, and the 48-60 months period represent the key default risk window.

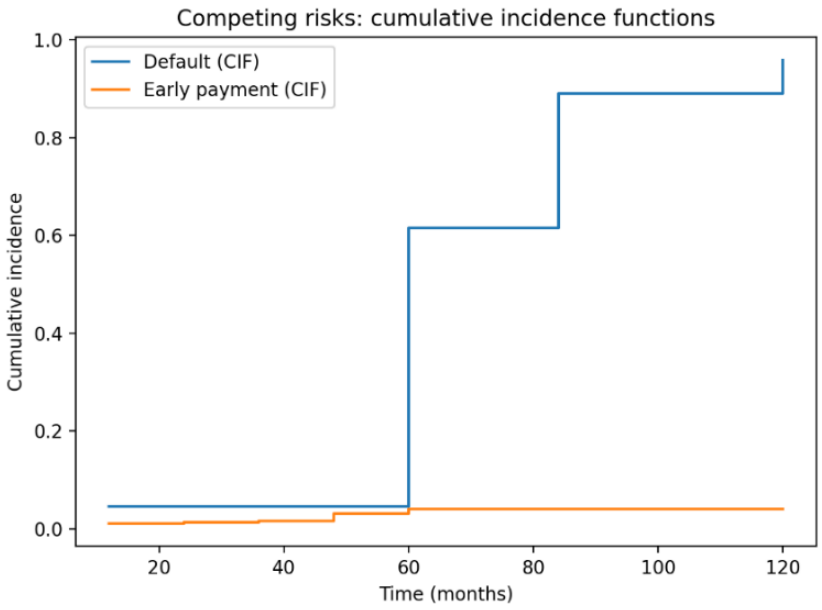



***Figure 3.*** *Cumulative incidence functions for default and early settlement.*

A multivariable cox proportional hazard regression model was fitted to evaluate determinants of time to default among 1 168 consumer loans. The model was statistically significant overall (likelihood ratio test p< 2e-16) and showed strong discrimination (concordance = 0.983), indicating clear separation in default timing across borrower and loan profiles.

After adjustment, income group and education level were significant socio demographic predictors. Compared with low-income borrowers, default hazards were higher for the medium income group and high-income group. Education was also associated with default timing: borrowers with primary, secondary, tertiary non-degree education had significantly higher default hazards compared with the higher education reference category. Furthermore, loan design characteristics including contracted loan term, interest rate category, and loan amount showed significant associations with time to default.

*Table 2. Cox proportional hazards model: selected significant predictors.*

| Predictor | Reference | Hazard Ratio (HR) | 95% CI | p-value |
|---|---|---|---|---|
| Income: Medium | Low | 1.56 | 1.24–1.96 | <0.001 |
| Income: High | Low | 1.43 | 1.14–1.79 | 0.002 |
| Education: Primary | Higher | 1.58 | 1.24–2.01 | <0.001 |
| Education: Secondary | Higher | 1.37 | 1.04–1.79 | 0.024 |
| Education: Tertiary (non-degree) | Higher | 1.59 | 1.10–2.31 | 0.015 |

The proportional hazards assumption was assessed using Schoenfeld residual tests. Most covariates satisfied the PH assumption. However, collateral violated the proportional hazard assumption and the global test was also significant indicating overall non proportionality driven mainly by collateral.

*Table 3. Schoenfeld residual test for proportional hazards.*

| Test | $\chi^2$ | df | p-value | Conclusion |
|---|---|---|---|---|
| Collateral | 19.45 | 1 | 0.00001 | PH violated |
| Global test | 24.92 | 9 | 0.0031 | Overall violation |

A binary logistic regression model was estimated as a benchmark to identify factors associated with default when default is treated as a static binary outcome. The model showed acceptable discrimination with an AUC of 0.744 (Gini=0.488), indicating moderate ability to separate default and non-default loans. In the adjusted model, unemployment was associated with higher odds of default. Education categories were also strongly associated with default relative to the reference group. Medium income borrowers showed increased odds of default compared with low-income borrowers. Although this benchmark model provides useful correlates of default risk. It does not incorporate time-to-default, censoring or competing outcomes. Therefore, survival-based models are more suitable for understanding how default risk evolves over the loan lifecycle.

*Table 4. Logistic regression benchmark model.*

| Predictor | Odd ratio (OR) | p-value |
|---|---|---|
| Unemployed (vs formal) | 1.565 | 0.028 |
| Income: Medium (vs low) | 2.559 | 0.045 |
| Education: Primary (vs higher) | 5.119 | 0.001 |
| Education: Secondary (vs higher) | 4.161 | 0.004 |
| Education: Tertiary (vs higher) | 7.639 | <0.001 |

Model performance: AUC=0744 (Gini=0488)

## 4. Discussion

The findings of this study confirm that consumer credit risk is inherently time-dependent, with default risk increasing significantly in later loan durations. The Kaplan-Meier estimates show relatively stable survival during the early loan period, followed by a sharp decline after 48 months, identifying the 48-60 months' interval as a critical window. This pattern is consistent with prior studies, which emphasize that default risk is not constant over time and tends to increase as loans mature. Such evidence reinforces the importance of adopting time-to-event approaches rather than static credit scoring methods.

The significant differences observed across borrower and loan characteristics further support existing literature. In particular, the strong separation by contracted loan term and collateral align with findings by Bellotti and crook, who highlight that loan structure and underwriting conditions are key determinants of default timing [4]. These results suggest that longer term loans and collateral arrangements may influence borrower behavior differently across the loan lifecycle, especially in the later stages where financial stress accumulates.

The competing risks analysis reveals that default is the dominant outcome, while early settlement remains relatively rare. This is in line with Fare et.al, who argue that ignoring competing events can lead to biased risk estimates [12]. The use of cumulative incidence functions in this study provides a more accurate representation of event-specific probabilities, demonstrating that the sharp increase in default incidence after 48 months marks the primary escalation phase of credit risk.

The cox proportional hazard model identifies income and education as significant predictors of time-to-default, consistent with the broader credit risk literature, which links socio-economic characteristics to repayment capacity. However,

the violation of the proportional hazards assumption for collateral indicates that its effect varies over time. This finding supports the argument by Bellotti and crook that credit factors may exhibit time-varying effects, and that extended cox models or time-dependent covariates are necessary to capture such dynamics accurately [4].

The logistic regression benchmark demonstrates moderate predictive performance; it fails to incorporate the timing of default and censoring. This limitation has been widely noted in the literature, where survival-based models are shown to provide superior risk stratification and more actionable insights for credit risk management. The results highlight the value of survival and competing risks approaches in developing stage-specific monitoring and intervention strategies across the loan lifecycle.

## 5. Conclusions

This study highlights that consumer credit risk in Badulla District is strongly time-dependent, with default risk increasing substantially in later loan durations and peaking within the 48-60 months' period. The results demonstrate that both borrower characteristics and loan design features significantly influence repayment behavior, with socio-economic factors such as income and education playing a key role in determining time-to-default. The analysis also reveals that default is the predominant outcome compared to early settlement, underscoring the importance of accounting for competing risks in credit risk modeling.

Furthermore, the findings indicate that certain risk factors, particularly collateral, exhibit time-varying effects, suggesting that traditional modelling assumptions may be insufficient for capturing the full dynamics of credit risk. The study confirms that survival and competing risks approaches provide more meaningful, lifecycle-oriented insights than conventional static models. These insights are particularly valuable for improving risk monitoring, supporting early intervention strategies, and enhancing decision making in consumer lending contexts at the district level.

## Abbreviations

| | |
|---|---|
| KM | Kaplan-Meier |
| HR | Hazard Ratio |
| dF | Degree of Freedom |
| AUC | Area Under the Curve |

## Acknowledgments

The authors also extend sincere thanks to the relevant officers for their assistance in providing the required data and clarifications. Finally, special appreciation is given to the borrowers in Badulla district whose loan information made this research possible.

## Author Contributions

**Rajapaksha Pathira Gamage Amithani:** Conceptualization, Data curation, Formal Analysis, Methodology, Software, Writing – original draft

**Chathuri Lakshika Jayasinghe:** Supervision

**Miuran Dencil Jayaweera:** Supervision

## Data Availability Statement

The data is available from the corresponding author upon reasonable request.

## Conflicts of Interest

The authors declare no conflicts of interest.

## References

[1] Central Bank of Sri Lanka, "Financial Stability Review 2024," 2024.

[2] Department of Census and Statistics, "NATIONAL ACCOUNTS OF SRI LANKA 2019," 2019.

[3] M. Stepanova and L. Thomas, "Survival analysis methods for personal loan data," *Oper. Res.*, vol. 50, no. 2, pp. 277–289, 2002, https://doi.org/10.1287/opre.50.2.277.426

[4] T. Bellotti and J. Crook, "Credit scoring with macroeconomic variables using survival analysis," *J. Oper. Res. Soc.*, vol. 60, no. 12, pp. 1699–1707, 2009, https://doi.org/10.1057/jors.2008.130

[5] J. P. Sy and J. M. G. Taylor, "Estimation in a Cox Proportional Hazards Cure Model," 2000., https://doi.org/10.1111/j.0006-341X.2000.00227.x

[6] M. Stepanova and L. C. Thomas, "PHAB scores: proportional hazards analysis behavioural scores," 2001. [Online]. Available: www.palgrave-journals.com/jors

[7] D.. Cox, "Regression models and life-tables. 34(2), 187-220.," *J. R. Stat. Soc. Ser. B*, vol. 32, pp. 187–220, 1972., https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

[8] D. Schoeneeld, "Partial residuals for the proportional hazards regression model," 1982. [Online]. Available: http://biomet.oxfordjournals.org/

[9] N. Capon, "credit scoring systems critical analysis," *JournalofMarketing*, p. 46: 2, 1982., https://doi.org/10.1177/002224298204600 20

[10] J. P. Fine and R. J. Gray, "A Proportional Hazards Model for the Subdistribution of a Competing Risk," *J. Am. Stat. Assoc.*, vol. 94, no. 446, pp. 496–509, Jun. 1999, https://doi.org/10.1080/01621459.1999.10474144

[11] Department of Census and Statistics. (2019). *Household Income and Expenditure Survey 2019*. Retrieved August 24, 2026, from https://www.statistics.gov.lk/

[12] R. F ăre, G. Karagiannis, M. Hasannasab, and D. Margaritis, "A benefit-of-the-doubt model with reverse indicators," *Eur. J. Oper. Res.*, vol. 278, no. 2, pp. 394–400, Oct. 2019, https://doi.org/10.1016/j.ejor.2019.02.009

[13] Dietrich, M. T., Dobler, D., & de Gunst, M. C. M. (2025). Wild bootstrap for counting process-based statistics: a martingale theory-based approach. *Lifetime Data Analysis*. https://doi.org/10.1007/s10985-025-09659-w

[14] Laborda, J., & Ryoo, S. (2021). Feature selection in a credit scoring model. *Mathematics*, *9*(7). https://doi.org/10.3390/math9070746

[15] Shang, L., Zhao, J., Li, G., & Zhang, X. (2024). Survival analysis in credit risk management: a review study. *Journal of Credit Risk*, *20*(4), 59–83. https://doi.org/10.21314/JCR.2025.002