risks

MDPI

Article

Survival Analysis for Credit Risk: A Dynamic Approach for Basel IRB Compliance

Fernando L. Dala [1], Manuel L. Esquível [2] and Raquel M. Gaspar [3,*]

[1] Banco Nacional de Angola, Av. 4 de Fevereiro n. 151, Luanda, Angola; fdala@bna.ao
[2] School of Science and Technology and Nova Math, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal; mle@fct.unl.pt
[3] ISEG Research, Lisbon School of Economics and Management, Universidade de Lisboa, Rua do Quelhas, n. 6, 1200-781 Lisboa, Portugal
* Correspondence: rmgaspar@iseg.ulisboa.pt

Abstract

This paper uses survival analysis as a tool to assess credit risk in loan portfolios within the framework of the Basel Internal Ratings-Based (IRB) approach. By modeling the time to default using survival functions, the methodology allows for the estimation of default probabilities and the dynamic evaluation of portfolio performance. The model explicitly accounts for right censoring and demonstrates strong predictive accuracy. Furthermore, by incorporating additional information about the portfolio's loss process, we show how to empirically estimate key risk measures—such as Value at Risk (VaR) and Expected Shortfall (ES)—that are sensitive to the age of the loans. Through simulations, we illustrate how loss distributions and the corresponding risk measures evolve over the loans' life cycles. Our approach emphasizes the significant dependence of risk metrics on loan age, illustrating that risk profiles are inherently dynamic rather than static. Using a real-world dataset of 10,479 loans issued by Angolan commercial banks, combined with assumptions regarding loss processes, we demonstrate the practical applicability of the proposed methodology. This approach is particularly relevant for emerging markets with limited access to advanced credit risk modeling infrastructure.

Keywords: survival analysis; loan portfolios; default probability; Gompertz–Makeham model; hazard function; value at risk (VaR); expected shortfall (ES); Monte Carlo simulation; Basel IRB approach; loss distribution

JEL Classification: C11; G21; G22; H81

check for updates

Academic Editor: Alejandro Balbás

Received: 23 June 2025
Revised: 21 July 2025
Accepted: 5 August 2025
Published: 15 August 2025

Citation: Dala, Fernando L., Manuel L. Esquível, and Raquel M. Gaspar. 2025. Survival Analysis for Credit Risk: A Dynamic Approach for Basel IRB Compliance. *Risks* 13: 155. https://doi.org//10.3390/risks13080155

Copyright: © 2025 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) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Survival analysis—a statistical methodology traditionally used in actuarial science, healthcare, engineering, and reliability studies—has seen growing relevance in the financial sector (Kleinbaum and Klein 2012b). In particular, it offers a robust framework for modeling time-to-event data, such as the time until a loan defaults. This temporal dimension makes survival analysis especially suitable for assessing credit risk in loan portfolios, aligning with the Internal Ratings-Based (IRB) approach under the Basel regulatory framework (BIS 2006).

Unlike traditional credit scoring models that offer static default probabilities, survival analysis enables a dynamic, time-dependent understanding of credit risk. By modeling the time to default, financial institutions can obtain more detailed insights into when defaults are likely to occur, thereby enhancing their ability to manage risk exposure, allocate capital efficiently, and comply with regulatory requirements.

This paper presents a survival analysis-based methodology for estimating default probabilities and assessing credit risk in loan portfolios. We employ the Gompertz–Makeham parametric model to characterize the time to default, explicitly incorporating right-censored observations, which are common in loan data. These occur when the default event is not observed within the observation window—for example, due to early repayment, transfer to another institution, or loan maturity. Right censoring requires special attention during estimation, as it limits the available information on the timing of potential defaults. Properly accounting for censored data is essential to avoid biased estimates of default intensities and survival probabilities.

Survival analysis offers several practical advantages for credit risk management:

- Granular Risk Profiling: It enables the evaluation of credit risk at the level of individual loans or borrower segments.
- Dynamic Monitoring: Time-based risk assessment supports early detection of deteriorating credit quality.
- Strategic Insights: Outputs from survival models inform pricing, capital provisioning, and portfolio allocation decisions.
- Regulatory Alignment: The methodology supports key requirements of the IRB approach within the Basel framework (BCBS 2011).

In implementing this approach, we use a dataset of 10,479 loans from Angolan commercial banks. The dataset includes detailed information on loan characteristics, payment histories, and the timing of default or prepayment. Based on this information, we estimate survival functions and compute hazard ratios that reflect the relative risk of default over time. Unfortunately, the database does not contain direct information on losses. However, we show how the model can be easily extended to incorporate a calibration of loss given default (LGD), allowing for the construction of a complete loss process. This extension enables the estimation of key risk metrics such as Value at Risk (VaR) and Expected Shortfall (ES). In addition, we propose a novel approach to estimate VaR and ES as a function of loan age, providing a dynamic view of credit risk throughout the life cycle of a loan.

The remainder of the paper is organized as follows. Section 2 presents a concise review of the literature on credit risk assessment in loan portfolios. Section 3 introduces the survival analysis framework, outlining the key variables and methodological foundations. Section 4 details our empirical application using loan-level data from Angolan banks. This section is structured to highlight each stage of the methodological approach, demonstrating how to model loss given default (LGD) in conjunction with default probabilities to construct realistic loss processes. It concludes by illustrating how the framework can be used to analyze the evolving risk profile of loan portfolios and derive risk measures by loan age. Finally, Section 6 summarizes the main findings and discusses their implications for credit risk management.

2. Literature Review

Credit risk modeling has evolved significantly, with methodologies ranging from traditional statistical techniques to advanced machine learning (ML) and survival analysis. Early approaches, such as those by Micocci (2000), relied on Monte Carlo simulations to estimate default rates, while Crosbie and Bohn (2003) combined accounting- and market-based indicators to calibrate default probabilities. These foundational works, however, often made simplifying assumptions (e.g., static transition matrices, constant interest rates) that limited their adaptability to dynamic economic environments (Derbali 2018).

Macroeconomic drivers of credit risk have been extensively studied, particularly in emerging markets. For instance, Poudel (2013) demonstrated the impact of economic variables on credit risk in Nepal, highlighting macro-financial linkages. Structural and reduced-form models, such as those by Suisse (1997) and Zokirjonov (2018), extended credit risk assessment to include rating migrations and portfolio-level Value at Risk (VaR). Yet, these models often struggle with high-dimensional, non-linear data—a gap that ML methods have sought to address. Recent work by Bo (2024) underscores how macroeconomic shocks amplify model risk in regulatory frameworks in developing countries' baks.

The advancement of machine learning has introduced powerful alternatives for credit risk modeling (Leo et al. 2019). ML techniques, such as ensemble learning (Charpignon et al. 2014) and deep neural networks (Mahbobi et al. 2023), excel at capturing complex patterns in large datasets. Recent studies, like Sharma et al. (2022) or Hernes et al. (2023), show that ML models (e.g., XGBoost, random forests) often outperform traditional logistic regression in default prediction due to their ability to handle non-linear interactions. Cutting-edge ML approaches continue to push boundaries. Huang et al. (2025) apply language learning to small- and medium-sized enterprises (SMEs) credit risk, achieving superior accuracy but underscoring computational costs. Meanwhile, Kakadiya et al. (2024) leverage transformer architectures for default prediction, demonstrating state-of-the-art performance in high-data environments. However, their effectiveness depends heavily on data quality and volume, posing challenges for banks in emerging markets or smaller institutions with limited data infrastructure (Fejza et al. 2022). Additionally, ML models are often criticized as "black boxes," lacking interpretability—a key requirement for regulatory compliance under frameworks like Basel IRB (Erkkilä 2025; Hurlin and Pérignon 2023).

Amid these developments, survival analysis has emerged as a particularly suitable approach for modeling credit risk in a time-sensitive and probabilistic framework. Originally applied in biomedical and reliability studies, survival models are increasingly being adopted in credit scoring and default prediction due to their ability to handle censored observations and time-varying covariates. Banasik et al. (1999) and Stepanova and Thomas (2002) were among the first to apply survival techniques to consumer credit, enabling a more granular analysis of default timing. McDonald et al. (2010) combine survival analysis with Monte Carlo simulation to price mortgages and forecast cash flow distributions. Likewise, Hassan et al. (2018) use accelerated failure time models to estimate both the probability and timing of default, incorporating borrower characteristics and macroeconomic indicators. Recent advances in survival analysis have focused on addressing real-world constraints in credit risk modeling. Zhang et al. (2021) developed DeepSurv for corporate default prediction, demonstrating how neural networks can enhance traditional Cox models while maintaining interpretability. For federated learning applications—particularly relevant for multi-institutional risk modeling—Kragh Jørgensen et al. (2025) established foundational methods for discrete-time Cox models which can be applied in privacy-preserving environments, though implementation challenges remain for real-time banking applications. These studies focus on modeling the default event, but lack its connection to losses and risk measure, which is key in addressing the existent regulatory framework.

The choice between survival analysis and ML depends on the use case. While ML excels in high-data environments (e.g., Borisov et al. (2022)), survival analysis is more adaptable to sparse datasets and provides explicit time-to-event estimates—critical for regulatory capital calculations under IRB. On the data scarcity and the limitations of ML, see for instance Abdalla et al. (2025). Hybrid approaches are emerging as promising solutions (Baesens et al. 2005). For instance, Siphuma and van Zyl (2025) introduced Transformer survival models for credit risk, achieving state-of-the-art performance in temporal pattern recognition while Djeundje and Crook (2019) developed dynamic survival analysis frameworks specifically for loan portfolios. These approaches address the critical need for models that adapt to evolving economic conditions, a key requirement under Basel III, but also depend on large datasets, which are not available in emerging markets.

The choice between survival analysis and ML thus depends on the use case. While ML excels in high-data environments (e.g., Borisov et al. (2022)), survival analysis is more adaptable to sparse datasets and provides explicit time-to-event estimates—critical for regulatory capital calculations under IRB. On the data scarcity and the limitations of ML, see for instance Abdalla et al. (2025).

We propose a survival-based framework for IRB compliance, emphasizing not only time-to-default estimation, but also dynamic risk metrics (e.g., age-dependent VaR and ES). Unlike ML models, our approach is data-efficient and interpretable, making it suitable for banks with limited data. We extend survival analysis to loss given default (LGD) modeling, enabling end-to-end loss process simulation. In opposition to the existent literature we illustrate our approach on real data from a set of Angolan banks, thus focusing on the difficulties faced by emerging markets' banks.

3. Survival Analysis

Survival analysis is a statistical method used to analyze time-to-event data, where the event of interest—such as failure, recovery, or death—can occur at any point in time. Its primary objective is to estimate the likelihood of the event occurring over time. A key challenge in survival analysis is dealing with right-censored data, where the exact timing of the event is unknown for some observations, but it is known that the event has not occurred up to a certain point.

This modeling framework is particularly relevant for credit risk analysis, where we may observe defaults, but we may also observe loan repayment (at maturity), early settlement, or transfer to another institution —situations that lead to right-censored data.

In what follows, we introduce the survival analysis method, establish the key notation, and demonstrate its application to a credit portfolio. Most of the theoretical foundations are drawn from actuarial references such as Rocha and Papoila (2009), Kleinbaum and Klein (2012a), Dickson et al. (2013), and Smith (2017).

3.1. Setup and Notation

Let us define the "lifetime of a loan" as a non-negative, real-valued random variable T, representing the duration from loan origination to either default or censoring.

Based on the random variable T, we define the following:

- Age of a loan t: The duration (in years) from the origination date until default or the end of the observation period, where $t \geq 0$.
- Loan at age t: Denoted as (t).
- Maturity of a loan M: The time at which the loan is contractually scheduled to end.
- Remaining lifetime of a loan $T - t$: The time left until default or maturity, satisfying $0 \leq t + T \leq M$.
- At the loan's origination ($t = 0$), we have $T = 0$.

The *survival distribution function* $S(t)$ is the probability that a loan survives longer than time t, mathematically expressed as:

$$S(t) = P(T > t) = 1 - F(t), \tag{1}$$

where $F(t)$ is the cumulative distribution function (CDF) of T, representing the probability of default by time t. Denoted by $f(t)$ the probability density function (pdf) of $F(t)$, it follows that $f(t) = -\frac{dS(t)}{dt}$.

The *hazard rate function*, $h(t)$, focuses on the instantaneous probability of default. For a loan that has survived until time t (i.e., $T > t$), the conditional probability of default in the next small interval $(t, t + dt)$ is given by the following:

$$h(t) = \lim_{dt \to 0} \frac{Pr(t \leq T \leq t + dt | T > t)}{dt} = \frac{f(t)}{S(t)} = -\frac{\partial}{\partial t} \log[S(t)] \; . \tag{2}$$

Since $h(t)$ is derived from a probability density function, it satisfies $h(t) \geq 0$. However, because $h(t)$ is not itself a probability, it is not necessarily less than 1. For very small dt, we can interpret $h(t)dt$ as the approximate probability that a loan, having reached age t, defaults before reaching age $t + dt$:

$$h(t)dt \approx P(T \leq t + dt | T > t) \; . \tag{3}$$

Rearranging Equation (2), we get

$$S(t) = \exp\left(-\int_0^t h(u)\, du \right) = \exp[-H(t)], \tag{4}$$

where the *cumulative hazard function* $H(t)$ is defined as

$$H(t) = \int_0^t h(u)\, du. \tag{5}$$

Also from Equation (2), the density function $f(t)$ can then be written in terms of $S(t)$ and $h(t)$,

$$f(t) = S(t)h(t). \tag{6}$$

Considering $N(t)$ the number of defaults at risk at time t, the expected number of defaults at time t, $d(t)$, is given by

$$d(t) = N(t)f(t) \; . \tag{7}$$

3.2. Censoring

As previously mentioned, in the context of loan portfolios, censoring occurs when a loan is repaid in full before its scheduled term, transferred to another financial institution with due compensation to the original lender or reaches its contractual maturity M without defaulting. For those observations, the exact time of default is unknown and is treated as right-censored. Figure 1 illustrates instances of censored data. Right censorship leads to special care when building the key estimates for the survival table, in the next section.

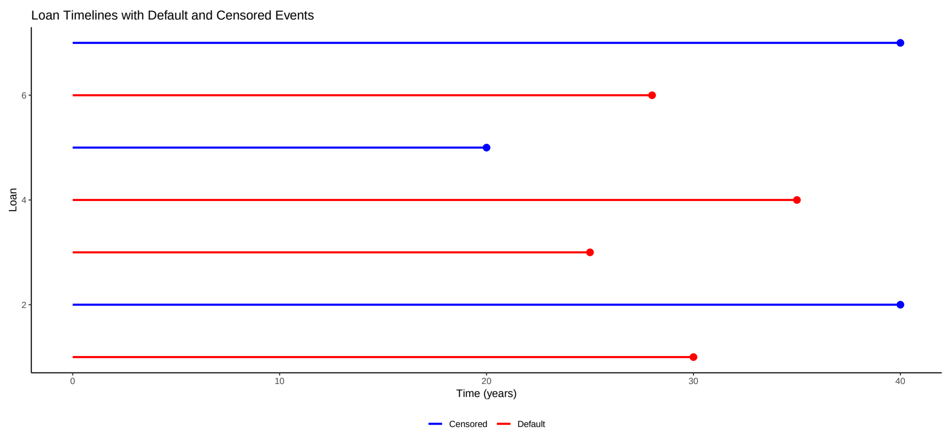


Figure 1. Loan timelines, indicating default and censored events.

Using data from our empirical study, Figure 2 illustrates the relationship between the survival function $S(t)$ and the hazard rate function $h(t)$ over time, with right censoring already considered. As time progresses, the survival probability $S(t)$ (red line) decreases, while the hazard rate $h(t)$ (red dashed line) increases, representing the growing risk of default as loans age.

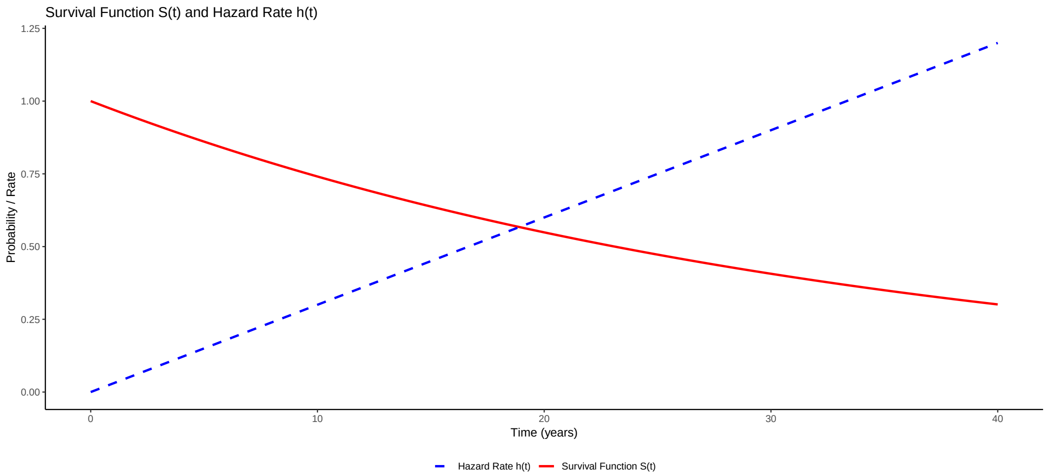


Figure 2. Visual representation of the Survival Function ($S(t)$) and Hazard Rate ($h(t)$).

When determining the number of defaults in a censored dataset, the survival function $S(t)$ is derived from the number of loans at risk (still "alive"). Therefore, defaults should be calculated relative to the remaining loan pool, $N(t)$ at each t.

3.3. Survival Table

Consider a cohort of n loans from a given portfolio. The interval $[0, \infty)$ is divided into $k + 1$ adjacent intervals of fixed width,

$$I_j = [t_{j-1}, t_j), \quad j = 1, \ldots, k+1,$$

with $t_0 = 0$, $t_k = M$, and $t_{k+1} = \infty$, where M is the upper limit of observation.

The data consists of the number of loans active at the beginning of each interval and the number of loans that defaulted or are censored in each interval.

Let us denote

- n_j: number of loans at risk at the beginning of interval I_j;
- d_j: number of defaults observed in I_j;
- m_j: number of censored observations in I_j.

With these we determine the number of loans at risk at the beginning of interval I_j as

$$n_j = n_{j-1} - d_{j-1} - m_{j-1} \qquad \text{for} \qquad j = 2, 3, \ldots, k+1\,.$$

The Kaplan and Meier (1958) estimator is a non-parametric method for estimating the survival function $S(t)$, accounting for right-censored data. The estimator computes the probability of surviving beyond each observed time point. If we had complete data, we could simply calculate the fraction of survivors at each time step. However, right-censored observations introduce uncertainty, so we need to adjust accordingly.

Instead of treating censored cases as failures, we only consider observed default events when updating survival probabilities. At each time t where a default occurs, we update the survival function based on the fraction of loans still "at risk" (i.e., not defaulted or censored before).

To estimate the survival function $S(t)$, we start by determining the estimated probability of default in the interval I_j. This can be calculated using the actuarial estimator,

$$\hat{q}_j = \frac{d_j}{n_j - \frac{m_j}{2}} \quad \text{for } n_j > 0, \tag{8}$$

where $\hat{q}_j$ is the estimated probability of default in interval I_j, adjusted for censoring (Rocha and Papoila 2009). Here, d_j represents the number of defaults, n_j is the number of loans at the beginning of the interval, and m_j denotes the number of censored observations in interval I_j.

Note that $n_j - \frac{m_j}{2}$ is the adjusted number of loans at risk in interval I_j. The adjustment is crucial for accurately estimating the number of loans at risk within interval I_j in the presence of censoring. The adjustment $n_j - \frac{m_j}{2}$ assumes that the instances of censoring are uniformly distributed throughout the interval I_j.

The estimated probability of survival through interval I_j is then as follows:

$$\hat{s}_j = 1 - \hat{q}_j, \tag{9}$$

and the cumulative survival probability $\hat{S}_j$ up to the end of interval I_j is given by the following:

$$\hat{S}_j = \prod_{i=1}^{j} \hat{s}_i. \tag{10}$$

Alternatively, it can be expressed recursively as $\hat{S}_j = \hat{s}_j \cdot \hat{S}_{j-1}$, with $\hat{S}_0 = 1$.

4. Empirical Application

In this section, we present an application of survival analysis to a portfolio of loans, demonstrating its practical use in evaluating credit risk.

The analysis is based on a dataset of 10,479 loans from the portfolios of 10 Angolan commercial banks, all classified within the same rating category and all with the same maturity of 40 years. The observation period spans from the origination date of each loan to the occurrence of a default or the end of the observation period. The latter includes scenarios where loans reach maturity, are prepaid, or are transferred to another financial institution with due compensation to the original lender.

Table 1 gives a detailed statistical overview of loan defaults and survival rates over their lifetime. Its columns are as follows: age (in years), for each interval, d_j shows the number of defaults, m_j shows the number of censored loans, and n_j shows the number of credits at risk in the beginning of the interval. The remaining columns result from the above formulas; $\hat{q}_j$ is the estimated default probability in the interval (Equation (8)), $\hat{s}_j$ the estimated survival probability in the interval (Equation (9)), and $\hat{S}_j$ represents the estimated cumulative survival probability (Equation (10)).

From Table 1, we observe that the estimated cumulative survival probability declines progressively over time, reflecting an increasing cumulative risk of default as loans mature. Default probabilities are very low in the early intervals but rise modestly in the later stages of the loan term. Additionally, the number of censored loans increases with loan age, with roughly half of the original loans reaching maturity without default or censoring. Overall, this table offers valuable insights into the long-term performance and evolving risk profile of the loan portfolio.

Table 1. Performance data of the portfolio.

Age	Interval	d_j	m_j	n_j	$\hat{q}_j$	$\hat{s}_j$	$\hat{S}_j$
1	[0, 1)	0	0	10,479	0.000000	1.000000	1.000000
2	[1, 2)	2	12	10,479	0.000191	0.999809	0.999809
3	[2, 3)	3	0	10,465	0.000287	0.999713	0.999522
4	[3, 4)	3	25	10,462	0.000287	0.999713	0.999235
5	[4, 5)	4	23	10,434	0.000384	0.999616	0.998852
6	[5, 6)	8	3	10,407	0.000769	0.999231	0.998084
7	[6, 7)	9	23	10,396	0.000867	0.999133	0.997219
8	[7, 8)	8	4	10,364	0.000772	0.999228	0.996449
9	[8, 9)	9	7	10,352	0.000870	0.999130	0.995583
10	[9, 10)	9	43	10,336	0.000873	0.999127	0.994714
11	[10, 11)	11	10	10,284	0.001070	0.998930	0.993649
12	[11, 12)	9	24	10,263	0.000878	0.999122	0.992777
13	[12, 13)	13	77	10,230	0.001276	0.998724	0.991511
14	[13, 14)	12	67	10,140	0.001187	0.998813	0.990333
15	[14, 15)	12	13	10,061	0.001193	0.998807	0.989151
16	[15, 16)	13	56	10,036	0.001299	0.998701	0.987866
17	[16, 17)	15	5	9967	0.001505	0.998495	0.986379
18	[17, 18)	20	7	9947	0.002011	0.997989	0.984395
19	[18, 19)	23	8	9920	0.002319	0.997681	0.982112
20	[19, 20)	22	9	9889	0.002226	0.997774	0.979926
21	[20, 21)	24	23	9858	0.002437	0.997563	0.977538
22	[21, 22)	25	0	9811	0.002548	0.997452	0.975047
23	[22, 23)	24	0	9786	0.002452	0.997548	0.972656
24	[23, 24)	27	0	9762	0.002766	0.997234	0.969965
25	[24, 25)	30	15	9735	0.003084	0.996916	0.966974
26	[25, 26)	33	56	9690	0.003415	0.996585	0.963671
27	[26, 27)	33	275	9601	0.003487	0.996513	0.960311
28	[27, 28)	37	106	9293	0.004004	0.995996	0.956465
29	[28, 29)	39	200	9150	0.004309	0.995691	0.952344
30	[29, 30)	40	85	8911	0.004510	0.995490	0.948048
31	[30, 31)	41	98	8786	0.004693	0.995307	0.943599
32	[31, 32)	39	104	8647	0.004538	0.995462	0.939318
33	[32, 33)	39	198	8504	0.004640	0.995360	0.934959
34	[33, 34)	43	155	8267	0.005251	0.994749	0.930050
35	[34, 35)	50	200	8069	0.006274	0.993726	0.924215
36	[35, 36)	49	250	7819	0.006369	0.993631	0.918329
37	[36, 37)	55	900	7520	0.007779	0.992221	0.911185
38	[37, 38)	54	450	6565	0.008517	0.991483	0.903424
39	[38, 39)	55	1000	6061	0.009890	0.990110	0.894489
40	[39, 40)	0	5006	5006	0.000000	1.000000	0.894489

Based upon 10,479 real credit portfolios of 10 commercial banks observed over 40 years. All credits are of the same rating class and maturity equal to 40.

Given the nature of the data, we chose to calibrate a Gompertz (1825) and Makeham (1860) function to the cumulative estimated survival values,

$$S_{GM}(t) = e^{at^2+bt+c(1-e^{dt})} \text{ for } t \geq 0 \tag{11}$$

where a, b, c, and d are constants.

The term e^{at^2+bt} represents the Gompertz Law component and captures the exponential increase in the hazard rate with time. Parameters a and b shape the hazard rate curve over time.

The term $c(1 - e^{dt})$ represents the Makeham term and is the baseline hazard rate, accounting for constant risks not related to time. Parameters c and d control the baseline hazard rate and its influence over time.

In our case, the calibrated values, optimized using gradient-based methods, gave values

$$a = -0.000043044$$
$$b = -6.820428518109184 \times 10^{-23}$$
$$c = 0.00021662$$
$$d = 0.138564\ .$$

The coefficient a reflects a quadratic term, suggesting a decreasing survival probability as time progresses. b is effectively negligible, while c and d end up governing the baseline and exponential growth. Figure 3 represents the calibration versus the estimated values, showing a very good fit.

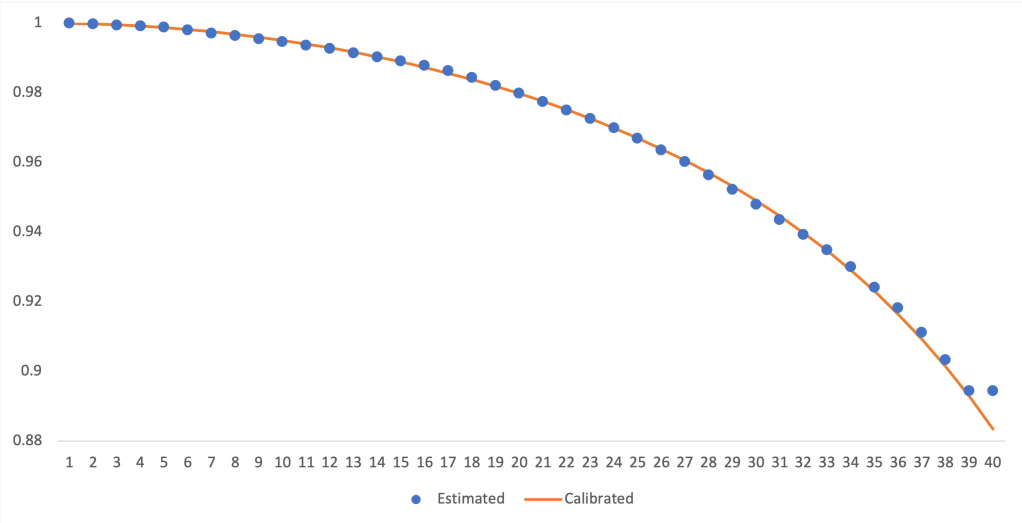


Figure 3. Estimated versus Calibrated Survival Probabilities.

4.1. Hazard Functions

We can use the calibrated survival function above to derive other key variables.

From Equation (2), we can obtain the hazard values from our Gompertz–Makeham survival function:

$$h_{GM}(t) = -\frac{\partial}{\partial t}\log[S_{GM}(t)], \tag{12}$$
$$h_{GM}(t) = 0.000086088t + 6.820428518109184 \times 10^{-23} + 2.999564228 \times 10^{-5}e^{0.138564t},$$

which represent the calibrated instantaneous rate of occurrence of default on a loan at each age or time point. In the context of a loan portfolio, the hazard value indicates the probability that a borrower defaults on their loan within a specific age or time interval $([t, t + dt[)$, given that they have survived up to that point (t).

Figure 4 represents the hazard function corresponding to a Gompertz–Makeham fitting to the survival function. As the age or time interval increases, the hazard value also rises. The increasing hazard values imply that the likelihood of experiencing default tends to grow as time progresses.

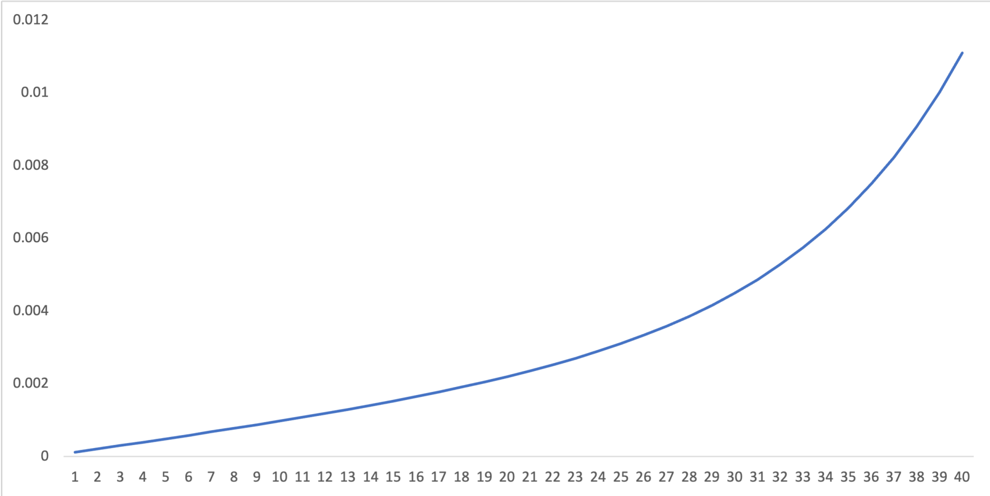


Figure 4. Hazard function.

4.2. Default Probabilities

The probability density of default defined in (6). $f(t)$ is related to the hazard function and the survival function by the following:

$$f_{GM}(t) = h_{GM}(t)S_{GM}(t).$$

Simplifying from (2), we obtain $f_{GM}(t) = -S'_{GM}(t)$, where $'$ denotes the derivative.

Finally, given the previous calibration, we get

$$f_{GM}(t) = 1.00021664346381\left(0.000086088\,t + 3.001573368 \times 10^{-5} \cdot e^{0.138564\,t} + 6.820428518109184 \times 10^{-23}\right) \times \\ \times \exp\left(-0.000043044\,t^2 - 6.820428518109184 \times 10^{-23}\,t - 0.00021662 \cdot e^{0.138564\,t}\right).$$

Figure 5 shows the calibrated density of defaults. Figure 6 compares probabilities of default from the calibrated model against the estimated probabilities of default from Table 1. It shows that although the fit is not perfect, its closeness to the 45-degree line attest the quality of our calibration, along with the Brier Score for the calibration curve of 3.64241333364887 $\times 10^{-6}$, indicating very low mean squared differences (Brier 1950; Steyerberg et al. 2010).

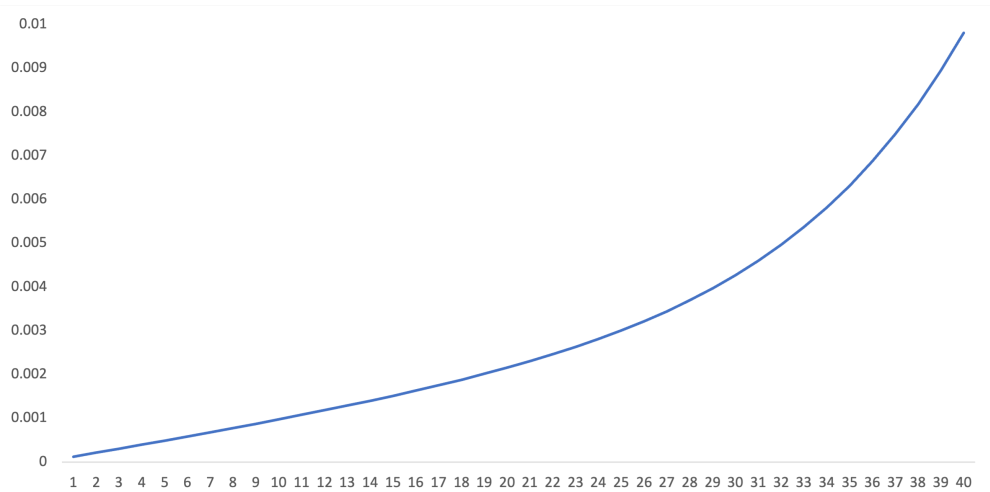


Figure 5. Probability density of default.

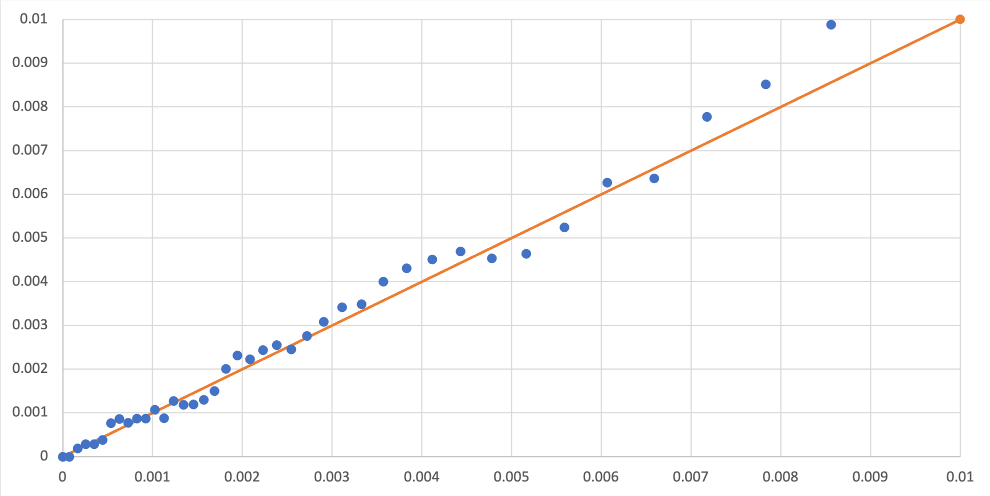


Figure 6. QQ-plot between estimated versus calibrated probabilities of default.

Given the density of defaults, the actual number of defaults can be obtained, provided that we also establish a framework for the number of loans at risk at each moment in time, $N(t)$. Accurately defining defaults at risk requires accounting for right censoring and properly integrating attrition effects. That is, calibrating the estimated number of loans at risk, n_j from Table 1, is crucial to model the number of default occurrences at each time interval.

From Figure 7 we see the number of loans at risk in our portfolio does not need more than a polynomial fit. So we use

$$N_{pol}(t) = -0.0124 \times t^4 + 0.07716^3 - 16.7t^2 + 102t + 10600. \tag{13}$$

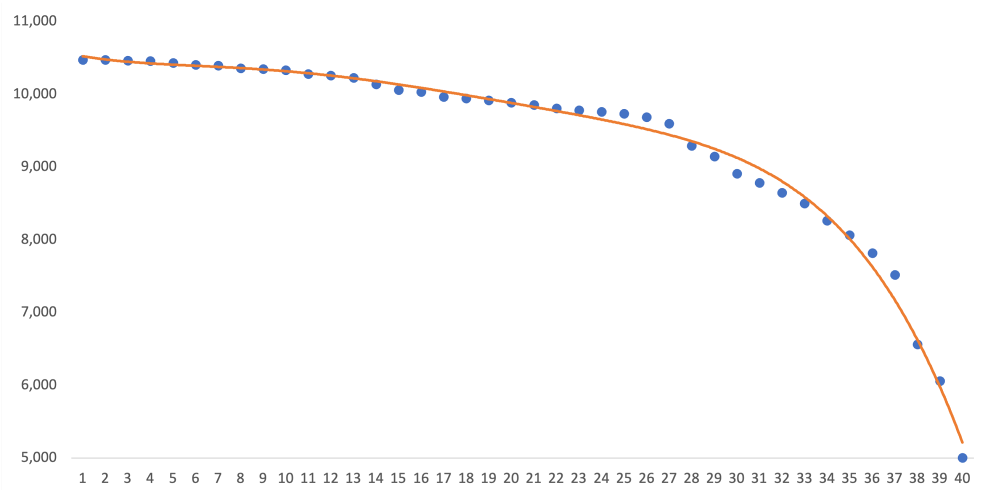


Figure 7. Number of loans at risk—estimated versus calibrated.

To obtain the number of defaults we use Equation (7) to get

$$d_{GM}(t) = N_{pol}(t) f_{GM}(t). \tag{14}$$

Figure 8 displays the estimated number of defaults, q_j, from Table 1, alongside the calibrated function $d_{GM}(t)$.

As we proceed with the analysis, it is important to exercise caution when interpreting results near maturity, given this potential underestimation.

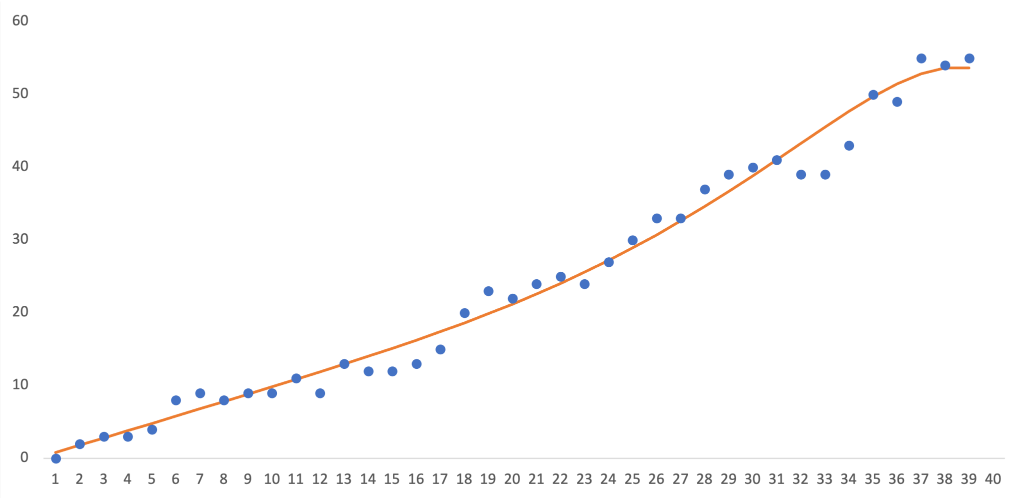


Figure 8. Number of defaults—estimated versus calibrated.

In practical applications, on the one at hand, we may need to be able to simulate the number of defaults per time interval $I_j = [t_{j-1}, t_j]$. By now we already have all ingredients and we are only missing a distribution for the number of defaults, which we can take as a non-homogeneous Poisson process, where the default intensity varies over time:

$$P(d_j = k) = \frac{\Lambda(t_{j-1}, t_j)^k e^{-\Lambda(t_{j-1}, t_j)}}{k!}\,, \qquad k = 0, \ldots, +\infty \tag{15}$$

where

$$\Lambda(t_{j-1}, t_j) = \int_{t_{j-1}}^{t_j} h(t) S(t) N(t) dt\,.$$

The Poisson distribution is a natural choice for modeling defaults within a survival analysis framework, particularly when defaults are viewed as rare, discrete events occurring over time. In this context, the number of defaults in a given interval can be modeled as a Poisson process, with the mean linked directly to the default intensity (hazard rate) and the number of loans at risk.

This approach is especially suitable when default intensities are known or can be estimated from historical data, allowing the Poisson mean to reflect time-varying risk.

Assuming the survival function, hazard rates and the number of loans at risk are approximately constant in each interval $I_j = [t_{j-1}, t_j]$, we get

$$\Lambda(t_{j-1}, j) \approx h(t_{j-1}) S(t_{j-1}) N(t_{j-1})(t_j - t_{j-1})\,. \tag{16}$$

In our case we use S_{GM}, h_{GM}, and N_{pol}, from Equations (11), (12), and (13), respectively.

Figure 9 displays 10 simulated paths of the number of defaults, generated using our calibrated functions. For each time interval $I_j = [t_{j-1}, t_j]$, the cumulative intensity Λ_j is computed as in Equation (16), and default counts are drawn using the inverse of the Poisson probability function defined in Equation (15).

It is interesting to note that the simulations not only capture the tendency for the number of defaults to increase as maturity approaches, but also reveal a rise in the volatility of potential defaults. This increase in variability stems from the uncertainty of results—larger default numbers naturally lead to greater fluctuations. These findings further confirm that the use of the Poisson distribution is appropriate.

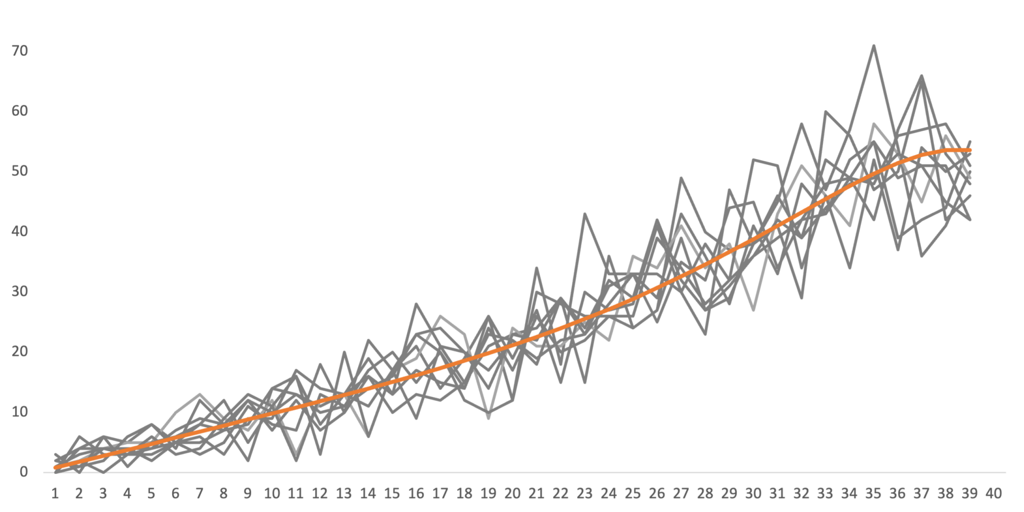


Figure 9. Simulated number of defaults.

4.3. Loss Given Default

Although our empirical data lacks direct information on losses given a default, a comprehensive credit risk assessment requires this crucial insight.

For simplicity, we assume an homogeneous portfolio, where all 10,479 loans share the same notional value of 25,000 monetary units (m.u.). The portfolios is, thus, worth 261,975,000 m.u. initially. Furthermore, we estimate the average loss given default to be 40% of the notional, amounting to 10,000 m.u. per loan.

We assume a 40% loss on *average* , but different scenarios in terms of the functional form of our loss function, $l(t)$, over time:

1. Decreasing: $l(t) = (M - t) \times 500$.
2. Increasing then Decreasing: $l(t) = t \times 975$ for $t \leq 20$, $l(t) = (M - t) \times 975$ for $t > 20$.
3. Constant: $l(t) = 10,000$.
4. Increasing: $l(t) = t \times 500$.

Scenario 1 recognizes that as maturity approaches, previous payments remain intact, reducing potential losses over time. Naturally, losses diminish as the loan nears completion—in the extreme case, when only one final payment is due, the loss is minimal. The multiplication by 500 ensures that the average loss remains at 10,000 m.u., maintaining comparability across different loss functions.

Scenario 2 considers a more complex pattern, where losses initially increase before declining as the credit matures.

Scenario 3 assumes a naïve approach, where a default at any point in time consistently results in a fixed loss of 10,000 m.u., regardless of when it occurs.

Finally, Scenario 4 explores an opposite trend, modeling steadily increasing losses over time.

While scenarios 2 and 4 may lack direct economic interpretability, their inclusion allows us to examine diverse possibilities in the absence of a confirmed empirical loss given default process. Figure 10 shows the four scenarios under consideration.

In reality, loss given default (LGD) is likely to follow a non-linear pattern, and we only observe discrete data points that should be calibrated using a non-linear function. The calibrated function can then serve as the foundation for computing the loss process and key risk measures. In this analysis, however, we apply the previously defined scenarios' functions.

It is important to note that the above approach provides LGD as a function of time but does not yield the aggregated loss process. The latter depends on the actual number of defaults at each point in time from the previous section.

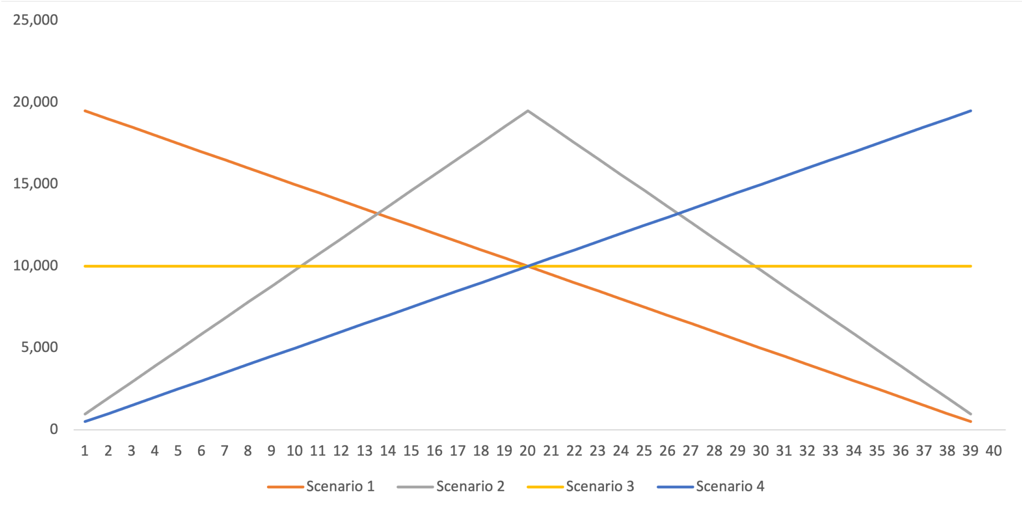


Figure 10. Loss Given Default.

In the next section, we bring all these elements together to construct a comprehensive view of the loss dynamics.

4.4. Loss Process

To determine the evolution of the loss process, $L(t)$, over time, we need two ingredients: (i) the number of defaults over time, $d_{GM}(t)$ (from Equation (6)), and (ii) the loss given default over time, $l(t)$ (as defined in the previous section),

$$L(t) = d_{GM}(t) \times l(t) \ .$$

Figure 11 shows the loss process, $L(t)$, associated with the four scenarios for loss given default, and assuming the calibrated function for d_{GM} as in Figure 8.

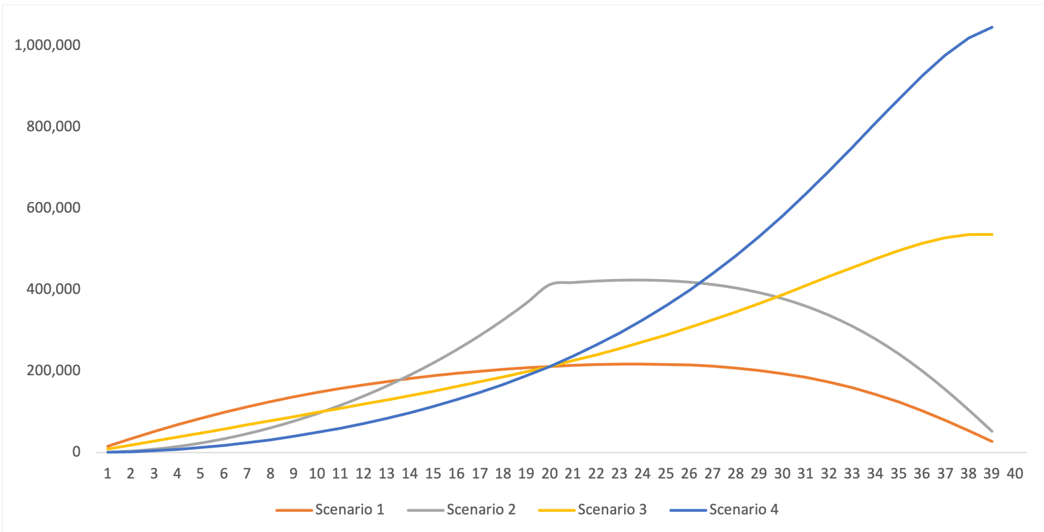


Figure 11. Loss processes.

Notice that the change in the pattern of LGD leads to very different loss processes. When LGD is considered constant (Scenario 3), of course the loss function inherits the shape of the number of defaults $d_{GM}(t)$. When LGD decreases with the age of the loan (Scenario 1), we see the loss process starts by increasing, but from approximately age 20 onwards, it actually decreases, despite the fact that the number of defaults increase over time. Scenarios 2 and 4 are less realistic, but the point here is to show the evolution of losses over time and to show they are determined by both default probability and LGD, and one should not neglect modeling LGD for proper modeling of credit risk.

4.5. Computing Risk Measures

The classical approach to risk measurement involves estimating standard credit risk metrics based on the distribution of the four loss processes under consideration. Figure 12 displays the histograms of the loss processes associated with these four scenarios (from Figure 11). Table 2 presents the corresponding values for Value at Risk (VaR) and Expected Shortfall (ES) for each scenario.

Both VaR and ES vary considerably depending on the assumed LGD scenario. Notably, the most realistic case—Scenario 1—yields values less than half of those obtained under a constant LGD assumption (Scenario 3), providing a strong incentive to model LGD accurately for reliable credit risk assessment.

Both the histograms and the risk measures presented thus far, however, do not account for the *time structure* of the loans' age—that is, they reflect losses aggregated on an annual basis without considering the specific timing within the loans' lifetimes. As a result, these measures are associated with annual losses, independently of ages of the loans.

From the discussions throughout the paper, it is evident that the risk profile of a loan aged 1-year is not the same as that of a loan with, for example, 20-years. The primary purpose of employing survival analysis is in fact to explicitly incorporate the timing and duration aspects into the risk assessment. Fortunately, we have all the necessary tools to compute risk measures that are adapted to the varying ages of the different loans.

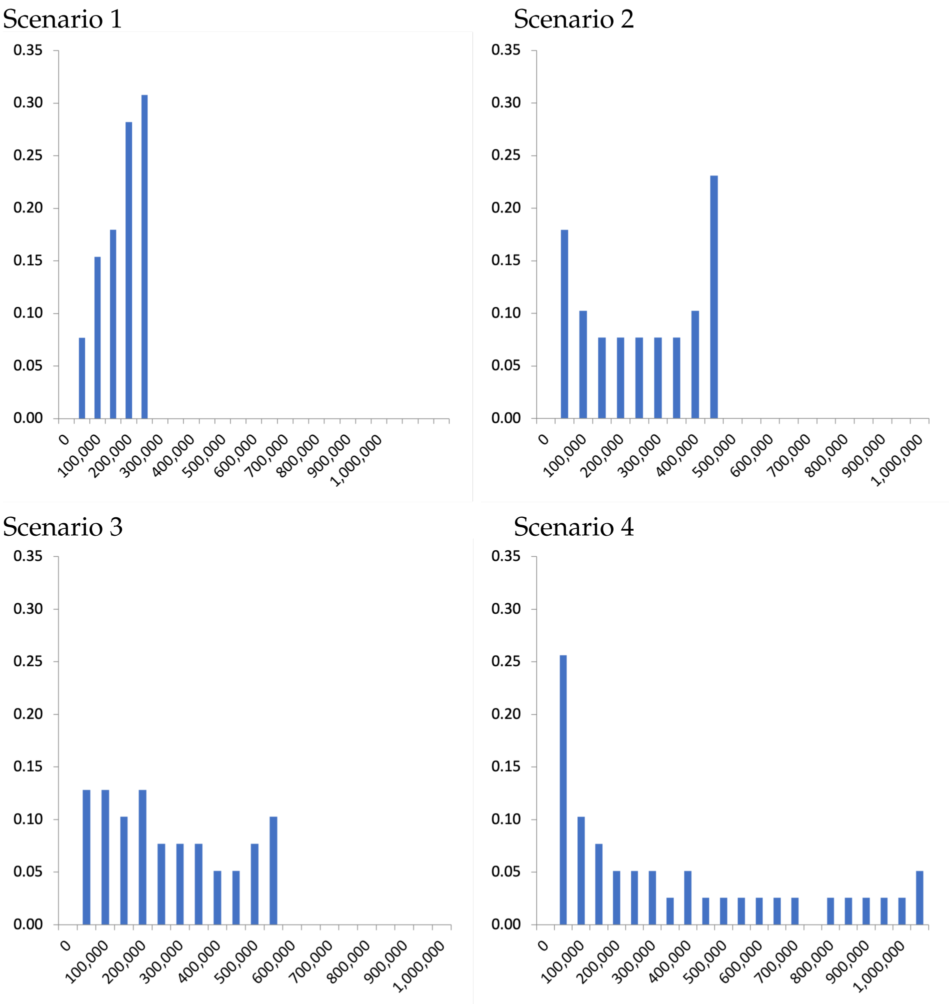


Figure 12. Loss histograms.

Table 2. Risk measures: VaR and ES.

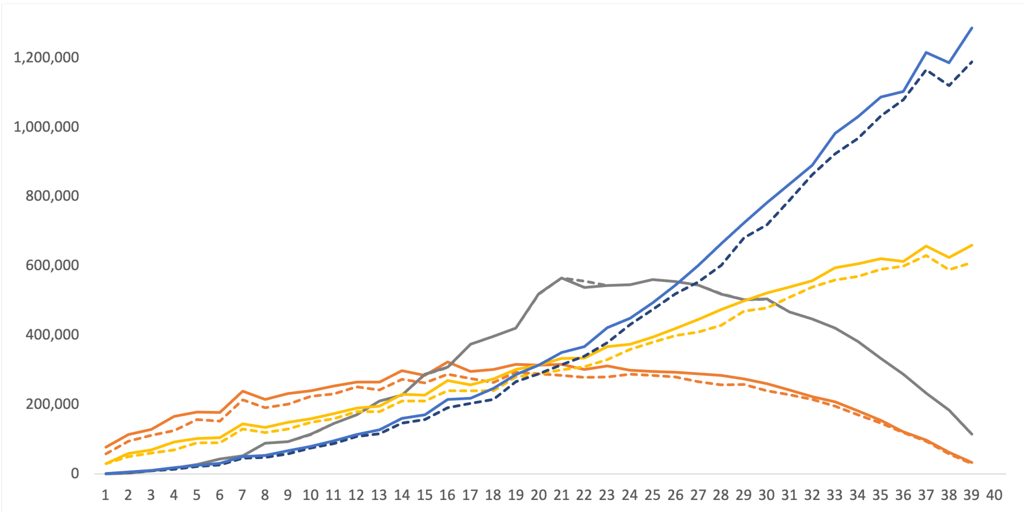
	Scenario 1	Scenario 2	Scenario 3	Scenario 4
		VaR		
95%	216,647.65	422,462.92	528,868.31	981,093.91
99%	217,283.22	423,702.27	536,135.02	1,035,259.28
		ES		
95%	217,260.50	423,657.97	536,130.62	1,032,050.98
99%	217,355.16	423,842.57	536,148.96	1,045,418.92

To generate Figure 12, we used the calibrated function $d_{GM}(t)$ shown in Figure 8. Now, we perform simulations similar to those illustrated in Figure 9, but instead of 10 paths, we consider 10,000 paths. Each simulation path multiplies the four possible LGD profiles depicted in Figure 11.

Thus, for each scenario in terms of LGD, we obtain 10,000 possible loss outcomes at each loan age. This approach enables us to obtain the distribution of losses per age of loan and compute the associated VaR and ES, providing a much deeper understanding of the risk profile of our loan portfolio.

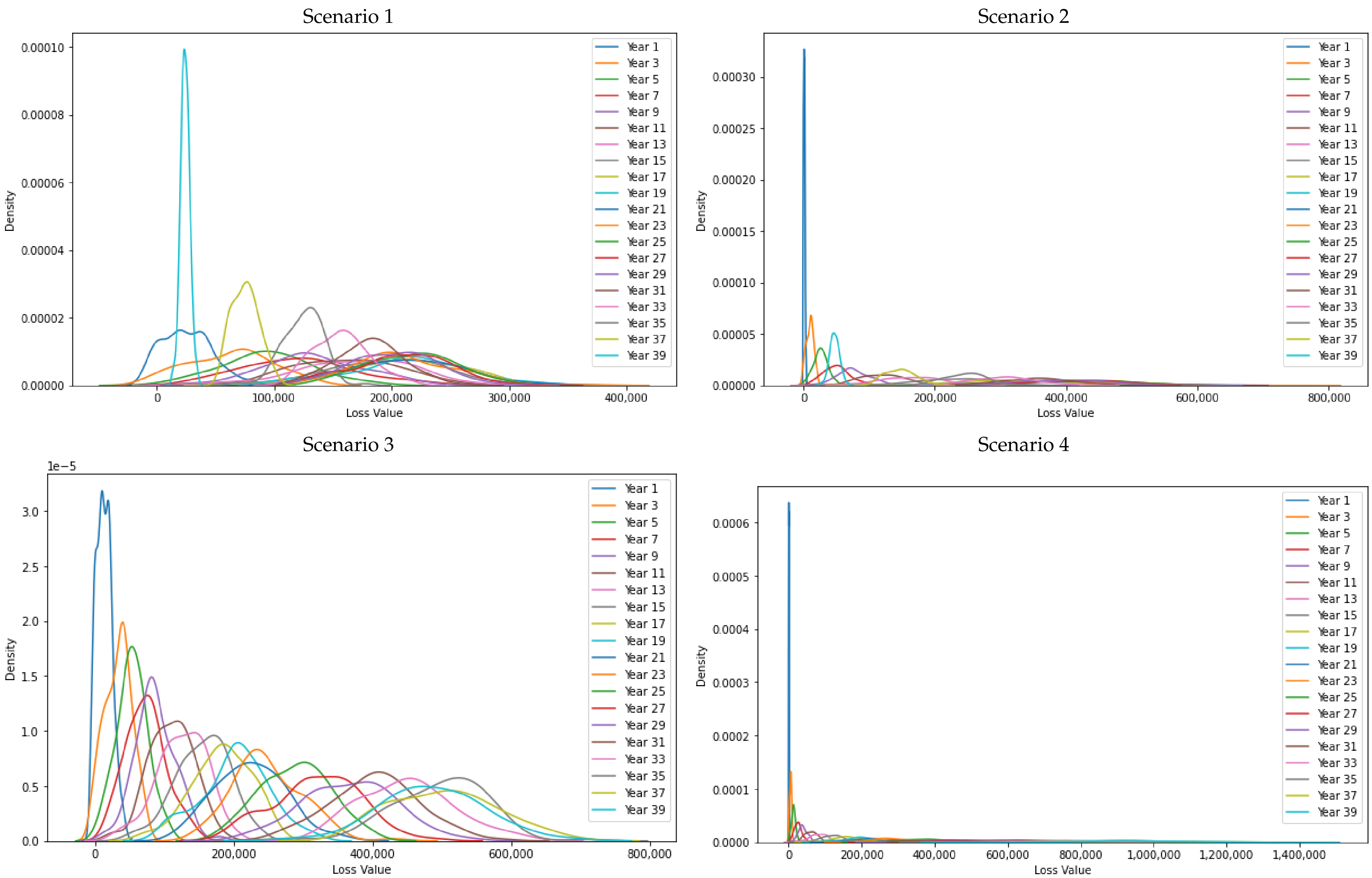
Having the densities, we can obtain VaR and ES which are age-specific statistics that differ according to the loans' age. Table 3 presents the numerical results for VaR and ES, while Figure 13 provides a visual illustration of the 95% VaR and ES across different scenarios. The figure clearly shows how both risk measures vary with the age of the loans, reflecting the influence of the underlying LGD profiles depicted in Figure 11. It also highlights how this age-sensitive approach delivers a far more rigorous and informative view than traditional, static risk metrics.

Figure 14 presents the density functions of losses obtained from the simulations across different scenarios and loan ages. Because the scales of the graphs differ, direct comparisons are not straightforward. Nonetheless, one clear pattern emerges: across all scenarios, the loss densities change with loan age, highlighting the dynamic nature of the risk profile over time. Note that losses under Scenario 1 exhibit the lowest values, whereas Scenario 4 shows the highest. This contrast directly reflects the structure of the underlying loss processes shown in Figure 11.



VaR (dashed) and ES (full) lines. Scenario 1 (orange), Scenario 2 (gray), Scenario 3 (yellow), Scenario 4 (blue).

Figure 13. 95% VaR and ES for the 4 LGD Scenarios, as a function of loan age.



To enhance readability, we only plot odd-numbered ages distributions, although a distinct density exists for every age.

Figure 14. Loss histograms per loan age for each scenario.

Table 3. VaR and ES per loans' age, for each loss scenario.

	Scenario 1				Scenario 2				Scenario 3				Scenario 4			
Age	**VaR 95**	**ES 95**	**VaR 99**	**ES 99**	**VaR 95**	**ES 95**	**VaR 99**	**ES 99**	**VaR 95**	**ES 95**	**VaR 99**	**ES 99**	**VaR 95**	**ES 95**	**VaR 99**	**ES 99**
1	58,500	78,000	78,000	78,500	2925	2925	2925	2925	30,000	30,500	40,000	40,500	1500	2000	2050	2150
2	95,000	114,000	114,000	114,500	9750	9750	9750	9750	50,000	60,000	60,500	70,000	5000	6000	6500	7000
3	111,925	129,500	129,500	129,550	17,696	17,843	17,843	20,475	60,500	70,000	72,500	75,000	9075	10,500	11,000	12,500
4	126,000	166,500	162,180	180,000	27,300	27,300	27,300	28,080	70,000	92,500	90,100	100,000	14,000	18,500	18,020	20,000
5	157,500	179,375	175,175	192,500	43,875	43,875	43,875	44,363	90,000	102,500	100,100	110,000	22,500	25,625	25,025	27,500
6	153,000	178,500	170,340	204,000	52,650	52,650	52,650	52,650	90,000	105,000	100,200	120,000	27,000	31,500	30,060	36,000
7	214,500	239,250	231,165	247,500	88,725	88,725	88,725	88,725	130,000	145,000	140,100	150,000	45,500	50,750	49,035	52,500
8	192,000	216,000	224,000	255,000	93,600	93,600	93,600	94,380	120,000	135,000	140,000	145,000	48,000	54,000	56,000	57,500
9	202,275	232,500	232,965	279,000	114,514	114,953	114,953	122,850	130,500	150,000	150,300	180,000	58,725	67,500	67,635	81,000
10	225,000	240,000	245,000	250,000	146,250	146,250	146,250	146,250	150,000	160,000	165,000	170,000	75,000	80,000	80,000	90,000
11	232,000	253,750	246,790	275,500	171,600	171,600	171,600	171,600	160,000	175,000	170,200	190,000	88,000	96,250	93,610	104,500
12	252,000	266,000	266,500	266,750	210,600	210,600	210,600	211,770	180,000	190,000	195,000	200,000	108,000	114,000	123,000	150,000
13	243,000	265,500	270,000	275,000	228,150	228,150	228,150	228,150	180,000	196,667	200,000	240,000	117,000	127,833	130,000	160,000
14	273,650	299,000	286,650	351,000	287,333	288,015	288,015	300,300	210,500	230,000	220,500	270,000	147,350	161,000	154,350	189,000
15	263,125	285,000	287,625	300,000	307,856	308,588	308,588	321,750	210,500	228,000	230,100	240,000	157,875	171,000	172,575	180,000
16	288,000	324,000	336,120	348,000	374,400	374,400	374,400	375,960	240,000	270,000	280,100	290,000	192,000	216,000	224,080	232,000
17	276,000	296,125	299,000	305,000	397,800	397,800	397,800	397,800	240,000	257,500	260,000	270,000	204,000	218,875	221,000	235,000
18	264,000	302,500	308,110	319,000	421,200	421,200	421,200	424,710	240,000	275,000	280,100	290,000	216,000	247,500	252,090	261,000
19	294,525	317,100	325,605	336,000	519,626	520,553	520,553	537,225	280,500	302,000	310,100	320,000	266,475	286,900	294,595	304,000
20	290,000	315,000	310,400	350,000	565,500	565,500	565,500	567,450	290,000	315,000	310,400	350,000	290,000	315,000	310,400	350,000
21	285,475	317,300	323,190	342,000	556,676	539,078	539,078	557,603	300,500	334,000	340,200	360,000	315,525	350,700	357,210	378,000
22	279,000	301,500	297,090	306,000	544,050	544,050	544,050	544,050	310,000	335,000	330,100	340,000	341,000	368,500	363,110	374,000
23	280,500	312,375	298,180	365,500	546,975	546,975	546,975	546,975	330,000	367,500	350,800	430,000	379,500	422,625	403,420	494,500
24	288,000	300,000	304,000	335,000	561,600	561,600	561,600	563,160	360,000	375,000	380,000	390,000	432,000	450,000	456,000	500,000
25	285,000	296,250	292,575	300,000	555,750	555,750	555,750	555,750	380,000	395,000	390,100	400,000	475,000	493,750	487,625	500,000
26	280,000	294,000	294,070	301,000	546,000	546,000	546,000	547,365	400,000	420,000	420,100	430,000	520,000	546,000	546,130	559,000
27	266,825	289,900	292,760	318,500	520,309	519,675	508,268	508,268	410,500	446,000	450,400	490,000	554,175	602,100	608,040	661,500
28	258,000	285,000	294,120	306,000	503,100	503,100	503,100	504,270	430,000	475,000	490,200	510,000	602,000	665,000	686,280	714,000
29	258,775	275,000	280,665	297,000	504,611	505,148	505,148	514,800	470,500	500,000	510,300	540,000	682,225	725,000	739,935	783,000
30	240,000	261,250	260,100	270,000	468,000	468,000	468,000	470,925	480,000	522,500	520,200	540,000	720,000	783,750	780,300	810,000
31	229,500	243,000	238,725	261,000	447,525	447,525	447,525	448,403	510,000	540,000	530,500	580,000	790,500	837,000	822,275	899,000
32	216,000	223,000	220,120	232,000	421,200	421,200	414,180	421,200	540,000	557,500	550,300	580,000	864,000	892,000	880,480	928,000
33	196,175	208,600	210,070	217,000	382,541	382,883	382,883	390,390	560,500	596,000	600,200	620,000	924,825	983,400	990,330	1,023,000
34	171,000	182,000	177,180	195,000	333,450	333,450	328,185	333,450	570,000	606,667	590,600	650,000	969,000	1,031,333	1,004,020	1,105,000
35	147,625	155,500	150,275	177,500	287,869	288,113	283,238	288,113	590,500	622,000	601,100	710,000	1,033,375	1,088,500	1,051,925	1,242,500
36	120,000	122,667	122,020	124,000	234,000	234,000	234,000	234,000	600,000	613,333	610,100	620,000	1,080,000	1,104,000	1,098,180	1,116,000
37	94,575	98,700	99,045	103,500	184,421	184,568	184,568	187,493	630,500	658,000	660,300	690,000	1,166,425	1,217,300	1,221,555	1,276,500
38	59,000	62,500	62,050	67,000	115,050	115,050	115,050	115,245	590,000	625,000	620,500	670,000	1,121,000	1,187,500	1,178,950	1,273,000
39	30,500	33,000	34,000	34,000	59,475	59,475	59,475	59,475	610,000	660,000	680,000	690,000	1,189,500	1,287,000	1,326,000	1,357,000

By explicitly accounting for loan age, this methodology reveals how credit risk evolves over time—something standard measures often obscure. Even when using familiar statistics like VaR and ES, the results differ meaningfully across loan age cohorts, underscoring the value of a dynamic perspective for risk assessment and portfolio management.

5. Managerial and Policy Implications

The survival analysis framework developed in this study carries significant practical implications for financial institutions and regulators, particularly in emerging markets where data limitations and dynamic risk profiles present unique challenges. For bank management, this approach enables dynamic risk monitoring through age-dependent metrics like VaR and ES, which evolve over a loan's life cycle. For instance, in the case of the Angolan banks we analyzed, our results show that the 95% VaR for Scenario 1 increases from 58,500 monetary units in Year 1 to 209,095 close to maturity, underscoring the need for progressive capital allocation and early interventions for aging portfolios.

For pricing and provisioning decisions, the Gompertz–Makeham hazard function reveals how default risk escalates over time. This insight allows banks to adjust credit spreads for longer-term loans and front-load loss provisions. The framework's adaptability to portfolio-specific data also supports targeted risk mitigation, such as restructuring high-risk cohorts before hazards peak.

From a regulatory perspective, this methodology directly supports Basel IRB compliance by providing time-to-default probabilities and age-sensitive LGD estimates. Central banks in emerging markets could adopt it as a standardized approach for institutions lacking ML capabilities, as advocated by recent studies on interpretable models. The model's extendibility to macroeconomic variables further enables regulators to stress-test portfolios under commodity price shocks or GDP contractions, informing countercyclical capital buffer policies.

However, broader implementation requires the addressing of two systemic gaps. First, the lack of LGD data in emerging markets remains a critical limitation, as our scenarios demonstrate VaR estimates vary by 110% across LGD assumptions. Second, the Poisson distribution may underestimate tail risk in volatile economies, suggesting a need for hybrid approaches. Policymakers can mitigate these gaps by incentivizing shared credit registries and standardizing loan age data collection, building on initiatives like the EU's AnaCredit system.

6. Conclusions

This paper demonstrates how survival analysis offers a robust, interpretable alternative to traditional credit risk models, particularly for emerging markets operating under the Basel IRB framework.

The proposed approach is grounded in survival analysis and centers on the calibration of key functions, such as the survival function and the number of loans at risk over time. It assumes a Poisson distribution for the number of defaults. We apply this methodology to a real dataset from 10 Angolan banks, comprising 10,479 loans. Based on this information, we fully characterize the probability of default—one of the core components of the IRB framework. Additionally, we simulate default paths conditional on the age of each loan.

Unfortunately, the dataset does not include information on loss given default (LGD), another essential element of the IRB framework, nor does it allow for holdout validation. To address the loss limitation, we explore several LGD scenarios and demonstrate that in shaping the final loss distribution and standard risk metrics—such as Value at Risk (VaR) and Expected Shortfall (ES)—the LGD profile is at least as influential as the default profile.

While LGD data was unavailable in our case, such information is typically accessible to banks through historical experience, making its calibration feasible in practice.

Building on the strengths of the survival-based approach, we also propose calculating risk measures conditional on loan age. These age-dependent metrics offer significantly more insight than traditional aggregated risk statistics. Our results reveal that loss densities vary meaningfully with loan age—an aspect that should not be overlooked, especially since such conditional information can be easily derived through simple simulations. To the best of our knowledge, this is the first study to compute VaR and ES as functions of loan age in the context of credit risk evaluation.

Throughout our application, we calibrated functional forms for key variables—such as the survival function and the number of loans at risk—to reflect the structure of our dataset. These calibrations achieved a strong empirical fit, though other datasets may require alternative specifications to accommodate different portfolio characteristics or risk dynamics. For modeling defaults, we relied on the Poisson distribution, a widely used and intuitive choice for count data. However, alternative distributions—such as the Negative Binomial or Binomial—may be more suitable in contexts where default counts exhibit greater variability or are constrained by the number of exposures.

Future research could focus on obtaining and incorporating empirical data on LGD, which would improve the calibration of loss distributions and enhance the accuracy of associated risk measures like VaR and ES. Moreover, integrating macroeconomic or borrower-specific covariates into the survival and LGD models could capture additional risk drivers and improve predictive power. Finally, linking the model's outputs to capital planning and stress testing processes would increase its practical relevance for both regulatory compliance and strategic risk management.

Author Contributions: Conceptualization, F.L.D., M.L.E. and R.M.G.; methodology, F.L.D., M.L.E. and R.M.G.; software, F.L.D. and R.M.G.; validation, M.L.E. and R.M.G.; formal analysis, F.L.D., M.L.E. and R.M.G.; investigation, F.L.D., M.L.E. and R.M.G.; resources, F.L.D., M.L.E. and R.M.G.; data curation, F.L.D.; writing—original draft preparation, R.M.G.; writing—review and editing, R.M.G.; visualization, F.L.D. and R.M.G.; supervision, M.L.E. and R.M.G.; project administration, R.M.G.; funding acquisition, M.L.E. and R.M.G. All authors have read and agreed to the published version of the manuscript.

Funding: MLE work was partially supported by national resources through the FCT—Fundação para a Ciência e a Tecnologia, I.P., under the scope of the project UIDB/00297/2020 and project UIDP/00297/2020. RMG work was partially supported by FCT, I.P., the Portuguese national funding agency for science, research and technology, under the Project UID06522.

Data Availability Statement: The empirical data used in this study was collected from 10 Angolan banks and is fully disclosed on Table 1. All remaining data is simulated using the methods described.

Conflicts of Interest: The views expressed in this paper are those of the authors and do not necessarily reflect those of the Banco Nacional de Angola; this study was conducted solely for academic purposes. The authors also declare no other conflicts of interest.

References

Abdalla, Hemn Barzan, Yulia Kumar, Jose Marchena, Stephany Guzman, Ardalan Awlla, Mehdi Gheisari, and Maryam Cheraghy. 2025. The future of artificial intelligence in the face of data scarcity. *Computers, Materials & Continua* 84: 1–27. [CrossRef]

Baesens, Bart, Tony Van Gestel, Maria Stepanova, Dirk Van den Poel, and Jan Vanthienen. 2005. Neural network survival analysis for personal loan data. *Journal of the Operational Research Society* 56: 1089–98. [CrossRef]

Banasik, John, Jonathan N. Crook, and Lyn C. Thomas. 1999. Not if but when will borrowers default. *Journal of the Operational Research Society* 50: 1185–90. [CrossRef]

BCBS. 2011. *Basel III: A Global Regulatory Framework for More Resilient Banks and Banking Systems Revised Version June 2011*. Basel: Basel Committee on Banking Supervision.

BIS. 2006. Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework-Comprehensive Version.
Bank for International Settlements. Available online: https://www.bis.org/publ/bcbs128.htm (accessed on 1 August 2025).

Bo, Wang. 2024. The impact of macroeconomic factors on credit risk management in developing country banks: An analysis based on basel III. *Financial Engineering and Risk Management* 7: 97–105. [CrossRef]

Borisov, Vadim, Tobias Leemann, Kathrin Seßler, Johannes Haug, Martin Pawelczyk, and Gjergji Kasneci. 2022. Deep neural networks and tabular data: A survey. *IEEE Transactions on Neural Networks and Learning Systems* 35: 7499–519. [CrossRef]

Brier, Glenn W. 1950. Verification of forecasts expressed in terms of probability. *Monthly Weather Review* 78: 1–3. [CrossRef]

Charpignon, Marie-Laure, Enguerrand Horel, and Flora Tixier. 2014. *Prediction of Consumer Credit Risk*. Stanford: Standford University.

Crosbie, Peter, and Jeff Bohn. 2003. *Modeling Default Risk: Modeling Methodology*. San Francisco: KMV Corporation.

Derbali, Abdelkader. 2018. How the Default Probability Is Defined by the Creditrisk+ Model? Available online: https://hal.science/hal-01696011v1 (accessed on 1 August 2025).

Dickson, David C. M., Mary Hardy, Mary R. Hardy, and Howard R. Waters. 2013. *Actuarial Mathematics for Life Contingent Risks*. Cambridge: Cambridge University Press.

Djeundje, Viani Biatat, and Jonathan Crook. 2019. Dynamic survival models with varying coefficients for credit risks. *European Journal of Operational Research* 275: 319–33. [CrossRef]

Erkkilä, Rami. 2025. Interpretable machine learning for credit risk predictions. *Economic Review* 105: 315–20.

Fejza, Doris, Dritan Nace, and Orjada Kulla. 2022. The credit risk problem—A developing country case study. *Risks* 10: 146. [CrossRef]

Gompertz, Benjamin. 1825. Xxiv. on the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. in a letter to francis baily, esq. frs &c. *Philosophical Transactions of the Royal Society of London* 115: 513–83.

Hassan, M. Kabir, Jennifer Brodmann, Blake Rayfield, and Makeen Huda. 2018. Modeling credit risk in credit unions using survival analysis. *International Journal of Bank Marketing* 36: 482–95. [CrossRef]

Hernes, Marcin, Jędrzej Adaszyński, and Piotr Tutak. 2023. Credit risk modeling using interpreted xgboost. *European Management Studies* 21: 46–70. [CrossRef]

Huang, Haonan, Jing Li, Chundan Zheng, Sikang Chen, Xuanyin Wang, and Xingyan Chen. 2025. Advanced default risk prediction in small and medum-sized enterprises using large language models. *Applied Sciences* 15: 2733. [CrossRef]

Hurlin, Christophe, and Christophe Pérignon. 2023. *Machine Learning and Irb Capital Requirements: Advantages, Risks, and Recommendations*. HEC Paris. Available online: https://ideas.repec.org/p/ebg/heccah/1480.html (accessed on 1 August 2025).

Kakadiya, Rushikesh, Tarannum Khan, Anjali Diwan, and Rajesh Mahadeva. 2024. Transformer models for predicting bank loan defaults a next-generation risk management. Paper presented at 2024 IEEE 6th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA), Hamburg, Germany, October 19–20, pp. 26–31.

Kaplan, Edward L., and Paul Meier. 1958. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 53: 457–81. [CrossRef]

Kleinbaum, David G., and Mitchel Klein. 2012a. *Survival Analysis: A Self-Learning Text*. Berlin and Heidelberg: Springer.

Kleinbaum, David G., and Mitchel Klein. 2012b. *Survival Analysis. Statistics for Biology and Health*, 3rd ed. New York: Springer.

Kragh Jørgensen, Rasmus Rask, Faartoft Jonas Jensen, Tarec El-Galaly, Martin Bøgsted, Rasmus Froberg Brøndum, Mikkel Runason Simonsen, and Lasse Hjort Jakobsen. 2025. Development of time to event prediction models using federated learning. *BMC Medical Research Methodology* 25: 143. [CrossRef]

Leo, Martin, Suneel Sharma, and Koilakuntla Maddulety. 2019. Machine learning in banking risk management: A literature review. *Risks* 7: 29. [CrossRef]

Mahbobi, Mohammad, Salman Kimiagari, and Marriappan Vasudevan. 2023. Credit risk classification: An integrated predictive accuracy algorithm using artificial and deep neural networks. *Annals of Operations Research* 330: 609–37. [CrossRef]

Makeham, William Matthew. 1860. On the law of mortality and the construction of annuity tables. *Journal of the Institute of Actuaries* 8: 301–10. [CrossRef]

McDonald, Ross A., Ania Matuszyk, and Lyn C. Thomas. 2010. Application of survival analysis to cash flow modelling for mortgage products. *OR Insight* 23: 1–14. [CrossRef]

Micocci, Marco. 2000. MARC: An actuarial model for credit risk. Paper presented at ASTIN Colloquium 2000, Porto Cervo, Italy, September 17–20, vol. 31.

Poudel, Ravi Prakash Sharma. 2013. Macroeconomic determinants of credit risk in nepalese banking industry. Paper presented at 21st International Business Research Conference, Toronto, ON, Canada, June 10–11, pp. 10–11.

Rocha, Cristina, and Ana Luísa Papoila. 2009. *Análise de Sobrevivência*. Portugal: Sociedade Portuguesa de Estatística.

Sharma, Alok Kumar, Li-Hua Li, and Ramli Ahmad. 2022. Default risk prediction using random forest and xgboosting classifier. In *2021 International Conference on Security and Information Technologies with AI, Internet Computing and Big-data Applications*. Berlin and Heidelberg: Springer, pp. 91–101.

Siphuma, Elekanyani , and Terence van Zyl. 2025. Enhancing credit risk assessment through transformer-based machine learning models. In *Artificial Intelligence Research. SACAIR 2024. Communications in Computer and Information Science*. Edited by Aurona Gerber, Jacques Maritz and Anban W. Pillay. Berlin and Heidelberg: Springer, vol. 2326.

Smith, Peter J. 2017. *Analysis of Failure and Survival Data*. Boca Raton: CRC Press.

Stepanova, Maria, and Lyn Thomas. 2002. Survival analysis methods for personal loan data. *Operations Research* 50: 277–89. [CrossRef]

Steyerberg, Ewout W., Andrew J. Vickers, Nancy R. Cook, Thomas Gerds, Mithat Gonen, Nancy Obuchowski, Michael J. Pencina, and Michael W. Kattan. 2010. Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology* 21: 128–38. [CrossRef] [PubMed]

Suisse Credit. 1997. *Creditrisk+, A Credit Risk Management Framework*. London: Credit Suisse Financial Products.

Zhang, Dongfang, Basu Bhandari, and Dennis Black. 2021. Covariate selection for mortgage default analysis using survival models. *Journal of Mathematical Finance* 11: 218–33. [CrossRef]

Zokirjonov, Mukhammadsodik. 2018. Methodology of creditmetrics for credit risk assessment. *International Finance and Accounting* 2018: 135.

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.