An investigation into Loss Given Default modelling and Economic Capital for Recovery Risk

J Larney

orcid.org 0000-0003-0091-9917

Thesis accepted in fulfilment of the requirements for the degree *Doctor of Philosophy in Science with Risk Analysis* at the North-West University

Promoter:	Prof GL Grobler
Co-promoter:	Prof JS Allison
Co-promoter:	Dr M Smuts

Graduation July 2023

20357273

Acknowledgements

As I reflect on the completion of this thesis, I am reminded of the African proverb that it takes a village to raise a child, and the same can be said for a thesis. I am incredibly grateful to the many individuals who have supported me throughout this journey.

I am forever indebted to my supervisor, Professor Gerrit Grobler, for his guidance, patience, encouragement, and endless cups of coffee throughout my PhD journey. Without his invaluable input and support, I would not have been able to achieve this milestone in my academic career.

In addition, I'm deeply grateful to my co-supervisors, Professor James Allison and Doctor Marius Smuts, for their generous contributions to my research. Their expertise and guidance have been invaluable, and I am fortunate to have had their support throughout my thesis journey.

Besides my supervisor and co-supervisors, this thesis would not have been possible without the fruitful collaboration with Professor Helgard Raubenheimer and Doctor Arno Botha, and I am sincerely thankful to them for sharing their knowledge so generously.

To my parents and sisters, I am forever grateful for your constant encouragement, love, and belief in me. I am especially grateful to my sister (and colleague), Mentje, whose eye for detail was a valuable asset to my thesis.

Lastly, but most importantly, I owe an enormous debt of gratitude to my husband, Marius, and children, Frederik, Annemarie and Thomas, for their unwavering support and sacrifices throughout my PhD journey. Their understanding, love, and encouragement carried me through the challenging moments, and I could not have done it without them.

Abstract

Risk management in the credit risk environment has advanced considerably over the last few decades. During the same time, statistical models developed for this purpose have grown in variety and sophistication. Initially, statistical models focused on the assessment of a counterparty's creditworthiness (i.e. default risk), but the modelling of the recovery process and outcome after default are receiving more and more attention. Loss Given Default (LGD) represents the loss suffered on a loan, expressed as a percentage of the exposure amount at default. The LGD is one of the key risk parameters used in determining a bank's regulatory capital under the Basel Capital Accords. This thesis comprises three articles, each making a unique contribution to a different aspect of the modelling of LGD.

In the first article, a methodology for determining an LGD discount rate is proposed. The methodology infers a discount rate from a market-consistent, or fair, price for defaulted exposures, inspired by the Solvency II regulatory regime's cost of capital approach. The main drivers of the LGD discount rate are found to be the mean and variance of recovery cash flows. The discount rate is also shown to be sensitive to the bank's cost of equity, particularly when recovery rates are low. It is demonstrated that this discount rate reflects the non-diversifiable risk inherent in the recovery cash flows and therefore satisfies the principles established by the banking regulator.

In the second article, two alternative models are proposed to better capture the bimodality of LGD data, and to reduce the difference between the empirical and distributional means, where the distributional mean is the expectation of the fitted parametric model. Both proposed models are mixtures of standard power distributions and outperform the benchmark beta distribution in terms of reducing bias in the estimation of the mean LGD. In this way, better alignment between a bank's Regulatory Capital and Economic Capital models can be achieved.

In the third article, an improved survival modelling approach for modelling the time to loan write-off is proposed. This approach includes a frailty component in the promotion time cure model. The frailty parameter accounts for common unobservable factors and possible observable covariates not in the model. The proposed models include gamma and inverse Gaussian distributed frailty parameters, as well as a shared frailty model. The finite sample performance of maximum likelihood estimation is analysed through a Monte Carlo simulation study, after which the models are applied to a simulated data set, as well as a real-world loan loss data set. The gamma frailty model produces the best fit to the data and allows for the modelling of a non-monotone shape of the hazard function. Additionally, the shared frailty term included accounts for the dependence between the failure times of individual subjects within a cluster, leading to more accurate estimates of hazard rates and write-off times. The proposed models offer an attractive alternative approach to modelling the probability of loan write-off in two-step LGD models.

Overall, this study should contribute to the development of more accurate and robust credit risk models, specifically in relation to LGD, which are essential for the effective management of credit risk in financial institutions.

Key words: *Loss Given Default, Recovery, Write-off, Economic capital, Discount rate, Bimodal distribution, Survival analysis, Frailty*

Preface

This thesis is written in article format and consists of two published articles and one unpublished manuscript (intended for publication). A general overview and background to the study are given in Chapter 1, along with the objectives of the thesis. The three articles can be found in Chapters 2, 3 and 4, respectively. In Chapter 5 we report the main findings of the three articles and provide some concluding remarks.

The first article, *A Cost of Capital Approach to determining the LGD Discount Rate*, is intended for submission to *ORiON*.

The second article, *Introducing Two Parsimonious Standard Power Mixture Models for Bimodal Proportional Data with Application to Loss Given Default*, has been published in a special issue of the journal *Mathematics*: *Mathematical Statistics and Modelling in Economics.*

The third article, *Modelling the Time to Write-Off of Non-Performing Loans using a Promotion Time Cure Model with Parametric Frailty*, has been published in a special issue of the journal *Mathematics*: *Application of Survival Analysis in Economics, Finance and Insurance.*

All co-authors have given consent for the use of these articles as part of the final thesis. The candidate was responsible for the formulation of the research questions, the conceptual design, the review of relevant literature, interpretation of results, and the initial drafting, as well as final editing, of the articles.

Table of Contents

Chapter 1

Introduction

A well-functioning banking system is a vital component of a country's economy and a key driver of prosperity. Banks play a crucial role in the economic development of a country by channeling funds from depositors to borrowers, and thereby facilitating investment, consumption and growth. According to Oort (1990), by extending credit, banks help individuals, businesses and governments to finance investments and manage cash flows.

Banks also have a crucial role to play in promoting financial stability by managing their exposure to risk. As financial institutions that intermediate between savers and borrowers, the majority of banks' earnings come from interest income levied on credit activities, such as loans and advances. However, extending credit is a risky business, as there is always a possibility that the borrower will default and fail to repay the loan. Suffering credit losses is therefore part and parcel of credit extension, and as such, part of the cost of doing business. Banks should manage this risk by carefully assessing the creditworthiness of borrowers, setting prudent loan terms, and by holding adequate provisions against expected credit losses. Banks must also hold sufficient capital to absorb losses in the event of worse than expected loan defaults, which, as argued by Kaufman (1996), helps to ensure that depositors' funds are protected, and that the financial system remains stable.

All commercial enterprises will hold capital to fund their operations and growth, as well as to provide a cushion for unexpected events. Elizalde and Repullo (2007) attest that, even in the absence of regulation, a bank's owners will mandate that the bank holds capital as a means of protecting itself against unexpected losses, and to maintain investor confidence. However, the role of banks in financial stability, and in the protection of depositors' money, necessitates intervention by the Regulator as to a minimum amount of capital required in order to reduce systemic risk in the financial system (Anginer and Demirgüç-Kunt, 2014). In many countries, including South Africa, the regulatory capital requirements for banks are governed by the the Basel Capital Accords, which are considered in Section 1.1.2.

The use of probability theory and statistical analysis to model the distribution of credit losses is a common practice in the field of finance (see e.g., the textbook by McNeil et al. (2015)), and this enables lenders and investors to quantify the credit risk of a portfolio of loans, and estimate potential losses. The distribution of credit losses can then also be used to determine the amount of provisions and capital needed to absorb expected and unexpected credit losses.

1.1 The determination of bank risk capital

1.1.1 Credit losses as a random variable

The credit losses suffered on a portfolio of loans can be regarded as a random variable, which can be defined by its probability density function (PDF). By representing credit losses as a PDF, banks can obtain a clearer picture of the potential range and likelihood of losses that they may experience from their credit exposures. From the analysis of such a loss distribution, a financial institution can estimate both the expected loss (EL) and the unexpected loss (UL) on its credit portfolio. The EL equals the (unconditional) mean of the loss distribution; it represents the loss that a bank can predict with reasonable certainty will realise over a specific time period (usually 1 year). The EL is part of doing business and should, firstly, be adequately priced for, and secondly, provisioned for in the bank's balance sheet. With adequate diversification and appropriate pricing, the expected net interest income from a portfolio of loans should exceed the portfolio's EL.

However, as the EL represents the mean value of the credit loss distribution, this amount can easily be exceeded. Besides provisioning for expected credit losses, banks are therefore also expected to hold risk capital against UL. UL represents the unanticipated losses suffered when banks and financial institutions function outside their normal operating environment. UL can be seen as the difference between a sufficiently high quantile (e.g. 99%) of the credit risk distribution, the so called Value-at-Risk (VaR), and the EL, as illustrated in Figure 1.1. According to Athanasoglou et al. (2008) banks should hold enough risk capital in order to fully cover the unexpected losses it may face.

The potential losses a bank may suffer can be described by three key risk parameters, and these parameters also drive the amount of risk capital a bank must hold. As explained by Genest and Brie (2013) these risk parameters include: Probability of Default (PD), which is the likelihood that a borrower will default on a loan or other credit exposure over a given period of time. Exposure at Default (EAD), which is the amount of exposure that a bank has to a borrower at the time of default. It is typically calculated as the outstanding balance of the loan plus any undrawn commitments that the bank has to the borrower. Finally, Loss Given Default (LGD), being the focus of this thesis, represents the amount of loss that a bank expects to incur if a borrower defaults on a loan. LGD is typically expressed as a percentage of

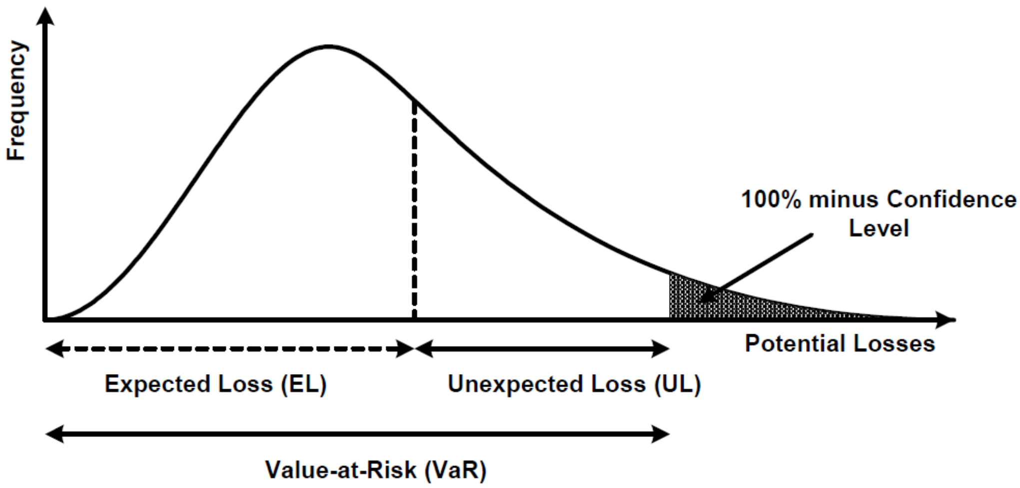

Fig. 1.1: The probability density function of credit losses with the EL, UL and Value-at-Risk (VaR) indicated. Source:(BCBS, 2004)

the EAD and is influenced by factors such as the characteristics of the borrower, the type of collateral or security that is held against the loan, and the legal and regulatory framework in which the bank operates.

The EL of a homogeneous loan portfolio is then simply calculated as $EL = EAD \cdot PD \cdot LGD$, where EAD is the total portfolio exposure amount, and PD and LGD are assumed to be independent, and identical for all loans in the portfolio (see BCBS (2006, pp. 86-88)).

1.1.2 Regulatory and Economic capital

We can distinguish between two types of risk capital, namely Regulatory capital and Economic capital. The amount of risk capital held by a bank will be the greater of these two amounts.

Regulatory capital (RC) is the amount of capital a bank is required to hold as per the stipulations prescribed by the banking regulator of the country. By requiring banks to hold sufficient risk capital, regulatory authorities aim to reduce the likelihood of bank failures and minimise the impact of such failures on the financial system and the economy as a whole, as argued by Kaufman (1996).

Economic capital (EC), on the other hand, refers to the amount of capital that a bank deems it should hold to run its business and remain solvent at a given confidence level and over a certain time horizon. It takes into account the bank's specific risk profile, business model and risk appetite, and Saidenberg and Schuermann (2003) argue that it may be more reflective of the bank's actual risk exposure than the regulatory minimum capital requirement.

The Basel Capital Accords

The Basel Capital Accords (the Accords, see BCBS (2003)) provide a framework for the calculation of minimum capital requirements that banks must hold in order to ensure their stability and solvency. The Accords have been widely adopted by countries around the world, including South Africa, and have been incorporated into the national banking regulations of the subscribing countries.

The Accords were developed by the Basel Committee on Banking Supervision (BCBS), which was established in 1974 by the Group of Ten (G-10) countries with the goal of promoting international cooperation and coordination among national banking supervisory and monetary authorities (Goodhart, 2011). Over the years, the BCBS has developed several versions of the Accords, including Basel I, Basel II, and Basel III, which have become the cornerstone of the global regulatory framework for banks.

Basel I, first introduced in 1988, was a relatively simple framework that established minimum capital requirements for banks based on the risk weightings of their assets. The risk weightings were predetermined and based on broad categories of assets, such as government securities, loans to commercial banks and residential mortgages. Under Basel I, all loans in the same category were assigned the same risk weight, regardless of the credit quality of the borrower, which led to criticisms (see e.g., Gehrig (1995)) that the capital requirements did not accurately reflect the risks involved.

Basel II, which was introduced in 2004, made substantial changes to the capital adequacy requirements in respect of credit risk. It recognised that the credit risk of loans can vary depending on the creditworthiness of the borrower, the characteristics of the loan, and the prevailing economic conditions. Therefore, it provided more risk-sensitive risk weightings for assets, which allowed banks to hold less RC for less risky loans, and more RC for riskier loans.

Basel II introduced two approaches to credit risk measurement: the standardised approach and the advanced internal ratings-based (AIRB) approach. Under the standardised approach, which is conceptually similar to Basel I, the risk weights depend on ratings provided by external credit rating agencies. The AIRB approach, on the other hand, is a more sophisticated method that allows banks to use their internal rating systems to estimate the PD, LGD and EAD for each borrower. A major advantage of the AIRB approach is the fact that it is more risk-sensitive, which in turn enables a bank to have a better understanding of the relationship between risk and return (van Laere and Baesens, 2009).

RC under the AIRB approach is calculated using the Asymptotic Single Risk Factor (ASRF) model, first introduced by Gordy (2003). Under the ASRF model, the loss rate (i.e. the total loss amount expressed as a percentage of the exposure) in respect of a well diversified portfolio of loans depends only on a (single) systematic risk factor, and not on idiosyncratic risk factors associated with individual exposures. The single systematic factor allowed for is widely interpreted as reflecting the state of the global economy. The ASRF model takes the average PD over an economic cycle, a conservative estimate of LGD, known as "downturn LGD", and EAD, as inputs. The entire distribution of the (annual) PD is derived from a supervisory mapping function that transforms the average PD into one that is conditional on the state of the global economy. By combining the distribution of the conditional PD with the EAD and LGD (both assumed to be constants), and incorporating a correlation factor that reflects the dependency between individual obligors, a credit loss distribution is produced. The UL (and RC) of the portfolio is then determined as the difference between the Value-at-Risk, at a 99.9% confidence interval, of the loss distribution and the portfolio EL, as described in Section 1.1.1.

Unlike the PD, regulators do not specify a formula for calculating a conditional LGD. It is up to the bank to provide estimates of LGD that are sufficiently conservative, and reflective of economic downturn conditions. Consequently, the ASRF model does not allow for variation in the LGD, and the use of downturn LGD can be seen as compensation for this simplifying assumption. A further limitation of the ASRF model is that independence between PD, LGD and EAD is assumed. Several studies have identified volatility in LGD, as well as dependence between PD and LGD, as potentially important sources of unexpected credit losses for banks; see e.g., Frye et al. (2000), Düllmann and Gehde-Trapp (2004) and Caselli et al. (2008). Therefore, despite the Basel AIRB approach allowing for the estimation of a fairly risk-sensitive capital requirement, most banks still find it necessary to build their own internal capital models to determine EC, as discussed in the next section.

Economic capital

EC, as defined in the third paragraph of Section 1.1.2, is widely considered a more accurate reflection of a financial institution's risk profile than RC. Hence, the closer the RC requirement of a bank is to its EC, the higher its return on equity will be, because the excess RC requirement represents a "superficial" burden on the bank's balance sheet as argued by Mérő (2021). Consequently, banks have a strong incentive to obtain regulatory permission to use their own internal models to determine their RC requirement, and achieve alignment between RC and EC.

Financial institutions will typically apply internal credit risk models to evaluate the EC necessary to face the risk associated with their credit portfolios. EC is widely used for product pricing, internal capital adequacy assessments and risk-adjusted performance measures (Burns, 2004).

The relevance and usefulness of calculating EC do, however, depend on the extent to which it is incorporated into strategic decision making by management (BCBS, 2008). The credit risk EC framework should ideally recognise concentration risks and diversification benefits that may arise from, e.g., geographical and industry exposure, and as a result, as asserted by van Laere and Baesens (2009), EC may better reflect the actual risks embedded in a credit portfolio than RC.

In practice, there are generally two ways to generate a loss distribution for EC modelling purposes. The first method is based on approximating the credit loss PDF analytically, by adopting various simplifying assumptions. From the approximated PDF, the EL and UL, using the VaR_α (or alternative risk measure), evaluated at a suitable confidence interval, $1 - \alpha$, can be determined. The approach is similar to the ASRF model discussed in Section 1.1.2, but with some of the limiting assumptions of the model relaxed, as illustrated by Pykhtin (2003) and Tasche (2004).

The second, and more widely used, approach relies on Monte Carlo (MC) simulation to generate a portfolio-specific loss distribution. The studies by Rösch and Scheule (2011), Farinelli and Shkolnikov (2012) and Kaposty et al. (2017) provide examples of how such a MC simulation approach may be implemented. MC simulation allows for not only the PD, but also the LGD, and potentially the EAD, to be modelled as random variables.

Besides allowing for variability in both PD and LGD, the EC model can also explicitly allow for dependencies between these risk parameters. This should provide better insight to the bank on the drivers of unexpected credit losses.

1.2 The Loss Given Default risk parameter

Early credit risk models focused primarily on assessing PD as the main determinant of credit risk. However, despite its importance, the PD does not account for the potential impact of default on a lender's portfolio. In recent years, more and more research effort has been focused on the modelling of the outcome after default, to provide a more comprehensive assessment of credit risk. Accurate estimation of post-default recovery rates does not only aid more accurate loan pricing and provisioning, but, as argued by Wang et al. (2020), results in a more efficient use of a bank's capital.

LGD, as defined in Section 1.1.1 refers to the fraction of a loan that is not recovered following a default event. LGD typically falls in the range $[0, 1]$, where 0 represents a full recovery and 1 a total loss of the outstanding amount. In some models, such as Basel II's ASRF model, LGD is treated as deterministic and known in advance. In others, e.g., most EC models, it is more common for LGD to be treated as a random variable, see e.g., Pykhtin (2003) and Heid and Krüger (2011). This approach, arguably, provides a more accurate representation of reality.

LGD is formally defined as the "economic loss" on a financial instrument, which means that both material discount effects and material direct and indirect collection costs are included (BCBS, 2006). Several approaches to estimate LGD exist, with the workout method being the most widely used and regarded as suitable for both retail and corporate exposures (Baesens et al., 2016). The workout method considers the collection process after default, and entails discounting recovery cash flows (net of costs) to the date of default and comparing the resultant net present value to the EAD amount. The fraction of the EAD not recovered is the LGD.

In many cases, the workout period for recoveries may span several years and, hence, the rate used to discount recovery cash flows has a substantial effect on the LGD. Regulatory guidance requires that the discount rate used include a risk premium reflective of the "undiversifiable risk" of the post recovery cash flows. However, despite clear guidance on the principles that should be adhered to in setting the risk premium, no firm guidance on the methodology that should be followed has been issued.

Very few studies offer proposals on how to derive a risk premium that reflects the uncertainty of post default recoveries, in the absence of reliable market price data. The proposals of Maclachlan (2005), Gibilaro and Mattarocci (2007) and Witzany (2009) all make use of the *capital asset pricing model* (CAPM) introduced by Sharpe (1964). Consequently, they all rely on a market risk premium (MRP), despite there being very little agreement amongst practitioners and academics alike on what constitutes an appropriate MRP (Soenen and Johnson, 2008). There remains therefore ample scope for investigating alternative ways of determining the LGD discount rate.

1.2.1 Stochastic LGD for economic capital modelling

LGD, being subject to random fluctuations (and affected by a variety of uncertain factors) can be regarded as a random variable. It is therefore possible to estimate expected LGD for performing loans by considering the observed realised workout recoveries on defaulted accounts. Estimation of a portfolio's LGD should be based on the bank's "own loss and recovery experience, as it is reflected in historical data on defaulted exposures" (European Banking Authority, 2017). Having discounted the recovery cash flows to the date of default through the workout approach, the individual realised loss l_i in respect of individual loan i, $i = 1, 2, ..., n$, can be calculated, and the average and variance of the portfolio level

LGD, $\bar{L}$ and $\hat{\sigma}^2$, can be estimated as follows:

$$\bar{L} = \frac{1}{n}\sum_{i}^{n} l_i$$

and,

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(l_i - \bar{L}\right)^2.$$

Recognising the stochastic nature of LGD is specifically important for EC modelling, where MC simulation approaches are employed to derive an empirical distribution of potential credit losses. In most simulation models it is assumed that LGD takes a specific parametric form, such as the beta distribution. Because of the highly regulated nature of the banking industry, parametric models are often preferred due to their analytical tractability and interpretability (Gordy, 2003). The BCBS, having surveyed practices in credit risk modelling, reports that most banks make use of the beta distribution to model LGD within their MC simulation frameworks (BCBS, 1999). Furthermore, many commercial credit models, including Creditmetrics^{TM} (Gupton et al., 1997) and KMV CreditPortfolioManager^{TM} (Vasicek, 1984), use the beta distribution to represent the LGD variable when simulating credit losses for EC calculation purposes.

The popularity of the beta distribution can mainly be attributed to it having support on the unit interval, $[0, 1]$, and the flexibility to represent a variety of shapes. This flexibility includes the ability to represent a bimodal shape, which is commonly reported for LGD data; see e.g. Memmel et al. (2012).

Despite its popularity and convenient properties, several studies have shown that the beta distribution often fails to provide a good fit for LGD data; see e.g., the studies by Calabrese and Zenga (2010) and Huang and Liang Xie (2012). In some cases this can be ascribed to U-shaped beta densities having asymptotes at 0 and 1, which leads to an overestimation of the masses near the endpoints. The masses at the endpoints then have an undue impact on the estimated parameters of the fitted beta distribution when using maximum likelihood estimation (MLE), leading to adverse effects in the estimation of the moments of the fitted distribution, where material bias could be present.

An important consideration for the choice of a parametric function to represent the LGD random variable is the fact that there exists a linear relationship between the average LGD and a bank's regulatory capital requirement. In the interest of alignment between a bank's RC and EC, as motivated in Section 1.1.2, the parametric distribution chosen to represent LGD for EC purposes should therefore be evaluated not only on goodness-of-fit, but also on the extent to which bias in the estimation of the mean is avoided.

This presents an opportunity to explore alternative distributions that provide a better fit than the beta distribution, specifically at the end points. Such a distribution should then also reduce the bias in the estimation of the mean of LGD.

Banks will typically fit parametric PDFs to historical loss data at segment level. This means that loan data has been subdivided into different risk segments, with the assumption that all obligors within a segment are statistically identical. Despite its important role in simulation-based EC modelling, the segment-based approach does not provide the bank with real insights into the relationship between LGD and the underlying risk factors. Models that are able to capture these relationships are invaluable to a bank's risk management toolkit and are considered in the next section.

1.2.2 LGD prediction models

Being able to estimate the impact of a specific client or loan factor on the loan-level LGD, enables the bank to evaluate the impact that these explanatory variables may have on the LGD of a performing loan. LGD prediction models are therefore enjoying growing interest from academics and bank practitioners alike.

The growth in the availability, quality and granularity of data is allowing for the development of models that can incorporate a wider variety of factors that may impact on LGD. Furthermore, diverse modelling techniques, including machine learning (ML) algorithms, have emerged that can handle complex datasets and capture nonlinear relationships between variables (see e.g., Olson et al. (2021)).

Although there are many ways to classify LGD prediction models, one perspective is to distinguish between "direct" and "indirect" (or "two-step") approaches. With the direct approach, LGD is estimated directly from a model that will take various factors, such as loan and borrower characteristics, as well as macroeconomic conditions, as inputs. In this regard. regression and transformation regression models have dominated; see e.g., Bastos (2010), Jacobs and Karagozoglu (2011), and Krüger and Rösch (2017), but the use of ML algorithms are rapidly gaining traction (Leo et al., 2019).

With a two-step model the probability and loss severity components of LGD are modelled separately, and the estimates of the components are then combined to arrive at an estimation of the LGD. Estimation techniques for the loss severity component broadly mirror those of the direct approach.

The purpose of the probability component is to predict the outcome of the loan after default. Two main outcomes after default are possible: the outstanding amount is fully recovered (resulting in an LGD of 0), or the loan is written off, and a partial ($0 < \text{LGD} < 1$), or full ($\text{LGD} = 1$) loss is suffered. For the estimation of the probabilities of these possible outcomes, logistic regression, or logit models, have enjoyed a strong foothold; see e.g. Matuszyk et al. (2010), Leow and Mues (2012), and Tong et al. (2013).

However, notwithstanding the success of logit models, both ML models and survival analysis are enjoying emerging popularity for modelling the outcome after default.

Survival analysis is a statistical method used to analyse and predict the time until an event of interest occurs. The event of interest is often referred to as the “failure” or “death” event, and in our context it is the write-off of a defaulted loan. Besides being able to model the time until write-off, a survival model is also able to estimate the probability of write-off within a certain time frame.

Survival analysis is particularly suitable for modelling the outcome of a defaulted loan, due to the likely presence of censoring in the data. Due to the potentially long resolution times of defaulted loans, a loan’s status at the end of the observation period is often recorded as “not resolved”. Furthermore, Hibbeln and Gürtler (2013) show that write-offs are typically associated with longer resolution times than full recoveries, and therefore, the number of write-offs is usually underrepresented in samples of defaulted loans.

In credit risk the Cox proportional hazards (CPH) model, proposed in the seminal paper of Cox (1972), has dominated, mainly due to the model’s ability to allow for loan- and borrower-specific covariates. Several authors, including Betz et al. (2017), Malwandla et al. (2017) and Joubert et al. (2019), have made use of the CPH to model the outcome after default of non-performing loans.

Alternatives to the CPH model include cure models, but these have received limited attention in the credit risk environment to date. Cure models, as all survival models, had their origination in medicine, where Berkson and Gage (1952) first propositioned that there exists a certain proportion of cancer patients who will never experience the event of interest, i.e. they can be considered cured. There are two primary groups of cure models, namely the mixture cure model and the promotion time cure (PTC) model (Amico and van Keilegom, 2018). Although only to a limited extend, both types of cure models have found applications in the credit risk environment, and specifically in the modelling of the probability of default, which in many high-quality loan portfolios represents a rare event. These studies include those by Tong et al. (2012), Dirick et al. (2019), de Oliveira et al. (2017), and Smuts and Allison (2021).

As far as application to the post default outcome of loans is concerned, cure models have received limited attention, and the literature is restricted to a single study by de Oliveira and Louzada (2014). In their study, the PTC model is used to model the probability of full loan repayment for portfolios of retail bank loans. The data is segmented according to risk score, and separate models are fitted for each segment.

There therefore exists much potential for the use of cure models in the modelling of post-default loan outcomes. Besides their ability to account for censored data, and to support a cure fraction in the data, cure models are also able to accommodate various borrower-specific and economic covariates. This makes them ideally suited for modelling the probability of write-off within two-step LGD models.

1.3 Thesis objectives

The general objective of this thesis is to contribute to the literature on LGD modelling, specifically in the context of EC models.

Having identified the paucity of approaches that satisfies the regulatory principles for the LGD discount rate, the first specific objective is to propose an alternative approach to determine such a rate. In the first article, *A Cost of Capital Approach to determining the LGD Discount Rate*, a methodology to derive a market consistent risk discount rate is described. Departing from the premise of Fama (1970) that the market knows best, an approach to estimate a market consistent price for a portfolio of defaulted loans is proposed, and from this price, a market consistent discount rate is inferred. It is demonstrated that this discount rate reflects the non-diversifiable risk inherent in the recovery cash flows and therefore satisfies the principles established by the banking regulator.

The second objective is to propose an alternative distribution to characterise the LGD random variable. Considering the bimodal nature of LGD, the second article, *Introducing Two Parsimonious Standard Power Mixture Models for Bimodal Proportional Data with Application to Loss Given Default*, aims to improve on the commonly used beta distribution to better capture the bimodality of LGD data. Besides parsimony, the objective is to formulate a model that reduces the difference between the empirical and distributional means. In this way, better alignment between the RC and EC models of a bank can be achieved.

The third thesis objective is to contribute to the literature on the use of survival analysis to model the outcome after loan default, for use, *inter alia*, in two-step LGD prediction models. In the third article, *Modelling the Time to Write-Off of Non-Performing Loans using a Promotion Time Cure Model with Parametric Frailty*, an improvement to existing survival modelling approaches is proposed by including a frailty component in the PTC model. We proposition that the outcome after default can, besides latent competing risks, also be influenced by common, unobservable drivers, such as the state of the economy. The ability of a univariate, as well as a shared, frailty parameter to control for this heterogeneity is therefore investigated. Besides being able to predict the time to write-off, the model can also output the probability of a loan to be written off or fully repaid at a specific point in time.

Bibliography

Amico, M. and van Keilegom, I. (2018). Cure models in survival analysis. *Annual Review of Statistics and its Application*, 5:311–342.

Anginer, D. and Demirgüç-Kunt, A. (2014). Bank capital and systemic stability. *World Bank Policy Research Working Paper; Number 6948.*

Athanasoglou, P. P., Brissimis, S. N., and Delis, M. D. (2008). Bank-specific, industry-specific and macroeconomic determinants of bank profitability. *Journal of International Financial Markets, Institutions and Money*, 18(2):121–136.

Baesens, B., Roesch, D., and Scheule, H. (2016). *Credit Risk Analytics: Measurement Techniques, Applications, and Examples in SAS*. John Wiley & Sons.

Bastos, J. A. (2010). Forecasting bank loans loss-given-default. *Journal of Banking & Finance*, 34(10):2510–2517.

BCBS (1999). Credit risk modelling: Current practices and applications. Available at `https://www.bis.org/publ/bcbs49.pdf`. Accessed on 8 May 2023.

BCBS (2003). The New Basel Capital Accord. Available at `http://www.bis.org/bcbs/irbriskweight.htm`. Accessed on 8 May 2023.

BCBS (2004). An Explanatory Note on the Basel II IRB Risk Weight Functions. Available at `http://www.bis.org/bcbs/irbriskweight.htm`. Accessed on 8 May 2023.

BCBS (2006). International convergence of capital measurement and capital standards: A revised framework, comprehensive version. Available at `https://www.bis.org/publ/bcbs128.htm`. Accessed on 8 May 2023.

BCBS (2008). Range of practices and issues in economic capital frameworks. Available at `https://www.bis.org/publ/bcbs152.pdf`. Accessed on 8 May 2023.

Berkson, J. and Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, 47(259):501–515.

Betz, J., Krüger, S., Kellner, R., and Rösch, D. (2017). Macroeconomic effects and frailties in the resolution of non-performing loans. *Journal of Banking and Finance*, 112:105212.

Burns, R. L. (2004). Economic capital and the assessment of capital adequacy. *Supervisory Insights*, 2(2):5–11.

Calabrese, R. and Zenga, M. (2010). Bank loan recovery rates: Measuring and nonparametric density estimation. *Journal of Banking and Finance*, 34:903–911.

Caselli, S., Gatti, S., and Querci, F. (2008). The sensitivity of the loss given default rate to systematic risk: new empirical evidence on bank loans. *Journal of Financial Services Research*, 34:1–34.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202.

de Oliveira, Jr, M. and Louzada, F. (2014). Recovery risk: Application of the latent competing risks model to non-performing loans. *Tecnologia de Crédito*, 88:44–53.

de Oliveira, Jr, M., Moreira, F., and Louzada, F. (2017). The zero-inflated promotion cure rate model applied to financial data on time-to-default. *Cogent Economics & Finance*, 5(1):1–14.

Dirick, L., Bellotti, T., Claeskens, G., and Baesens, B. (2019). Macro-economic factors in credit risk calculations: Including time-varying covariates in mixture cure models. *Journal of Business & Economic Statistics*, 37:40 – 53.

Düllmann, K. and Gehde-Trapp, M. (2004). Systematic risk in recovery rates: An empirical analysis of US corporate credit exposures. Available at `https://ssrn.com/abstract=2793954`.

Elizalde, A. and Repullo, R. (2007). Economic and regulatory capital in banking: What is the difference? *International Journal of Central Banking*, 3(3):87–117.

European Banking Authority (2017). Guidelines on PD estimation, LGD estimation and the treatment of defaulted exposures. Available at `https://tinyurl.com/45bv8v3d`. Accessed on 8 May 2023.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417.

Farinelli, S. and Shkolnikov, M. (2012). Two models of stochastic loss given default. *The Journal of Credit Risk*, 8(2):3.

Frye, J., Ashley, L., Bliss, R., Cahill, R., Calem, P., Foss, M., Gordy, M., Jones, D., Lemieux, C., Lesiak, M., Lobbes, C., McGrew, L., Mehta, P., Nelson, J., and Waggoner, E. (2000). Collateral damage: A source of systematic credit risk. *Risk*, 13(4):91–94.

Gehrig, T. (1995). Capital adequacy rules: Implications for banks' risk-taking1. *Swiss Journal of Economics and Statistics*, 131(4/2):747–764.

Genest, B. and Brie, L. (2013). Basel II IRB risk weight functions. *Demonstration and Analysis, Global Research and Analytics*, page 5.

Gibilaro, L. and Mattarocci, G. (2007). The selection of the discount rate in estimating loss given default. *Global Journal of Business Research*, 1(1):15–33.

Goodhart, C. (2011). *The Basel Committee on Banking Supervision: A history of the early years 1974–1997*. Cambridge University Press.

Gordy, M. B. (2003). A risk-factor model foundation for ratings-based bank capital rules. *Journal of Financial Intermediation*, 12(3):199–232.

Gupton, G. M., Finger, C. C., and Bhatia, M. (1997). CreditMetrics: The Benchmark for Understanding Credit Risk, Technical Document. *New York, NY: JP Morgan Inc.*

Heid, F. and Krüger, U. (2011). Do capital buffers mitigate volatility of bank lending? A simulation study. Available at `https://ssrn.com/abstract=2794054`.

Hibbeln, M. and Gürtler, M. (2013). Pitfalls in modeling loss given default of bank loans. *Journal of Banking & Finance*, 37:2354–2366.

Huang, C. and Liang Xie, S. (2012). Modeling Loss Given Default (LGD) by Finite Mixture Model. In *Proceedings of the Northeast SAS User Group 2012 Conference*.

Jacobs, M. and Karagozoglu, A. K. (2011). Modeling ultimate loss given default on corporate debt. *The Journal of Fixed Income*, 21(1):6–20.

Joubert, M., Verster, T., and Raubenheimer, H. G. (2019). Making use of survival analysis to indirectly model loss given default. *ORiON*, 34:107 – 132.

Kaposty, F., Löderbusch, M., and Maciag, J. (2017). Stochastic loss given default and exposure at default in a structural model of portfolio credit risk. *Journal of Credit Risk*, 13(1).

Kaufman, G. G. (1996). Bank failures, systemic risk, and bank regulation. *Cato Journal*, 16:17.

Krüger, S. and Rösch, D. (2017). Downturn LGD modeling using quantile regression. *Journal of Banking & Finance*, 79:42–56.

Leo, M., Sharma, S., and Maddulety, K. (2019). Machine learning in banking risk management: A literature review. *Risks*, 7(1):29.

Leow, M. and Mues, C. (2012). Predicting loss given default (LGD) for residential mortgage loans: A two-stage model and empirical evidence for UK bank data. *International Journal of Forecasting*, 28(1):183–195.

Maclachlan, I. (2005). Choosing the discount factor for estimating economic LGD. In Altman, E. I., Resti, A., and Sironi, A., editors, *Recovery Risk: The Next Challenge in Credit Risk Management*. Risk Books.

Malwandla, M. C., Marimo, M., and Breed, D. G. (2017). A cross-sectional survival analysis regression model with applications to consumer credit risk. *South African Statistical Journal*, 51(1):217–234.

Matuszyk, A., Mues, C., and Thomas, L. C. (2010). Modelling lgd for unsecured personal loans: Decision tree approach. *Journal of the Operational Research Society*, 61(3):393–398.

McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools-revised edition*. Princeton University Press, Princeton, New Jersey.

Memmel, C., Sachs, A., and Stein, I. (2012). Contagion at the Interbank Market with Stochastic LGD. *International Journal of Central Banking*, September 2012.

Mérő, K. (2021). The ascent and descent of banks' risk-based capital regulation. *Journal of Banking Regulation*, 22(4):308–318.

Olson, L. M., Qi, M., Zhang, X., and Zhao, X. (2021). Machine learning loss given default for corporate debt. *Journal of Empirical Finance*, 64:144–159.

Oort, C. J. (1990). Banks and the stability of the international financial system. *De Economist*, 138(4):451–463.

Pykhtin, M. (2003). Recovery rates: Unexpected recovery risk. *Risk*, 16(8):74–79.

Rösch, D. and Scheule, H. (2011). A multi-factor approach for systematic default and recovery risk. In *The Basel II Risk Parameters*, pages 117–135. Springer.

Saidenberg, M. and Schuermann, T. (2003). *The new Basel capital accord and questions for research*. Research Papers in Economics, Wharton Financial Institutions Center.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442.

Smuts, M. and Allison, J. S. (2021). An overview of survival analysis with an application in the credit risk environment. *ORiON*, 36:89–110.

Soenen, L. and Johnson, R. (2008). The equity market risk premium and the valuation of overseas investments. *Journal of Applied Corporate Finance*, 20(2):113–121.

Tasche, D. (2004). The single risk factor approach to capital charges in case of correlated loss given default rates. Available at https://ssrn.com/abstract=510982.

Tong, E. N., Mues, C., and Thomas, L. (2013). A zero-adjusted gamma model for mortgage loan loss given default. *International Journal of Forecasting*, 29(4):548–562.

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

van Laere, E. and Baesens, B. (2009). Regulatory and economic capital: theory and practice, evidence from the field. In *International Risk Management Conference, June*, pages 22–24.

Vasicek, O. A. (1984). Credit valuation. Available at http://www.ressources-actuarielles.net/EXT/ISFA/1226.nsf/0/c181fb77ee99d464c125757a00505078/$FILE/Credit_Valuation.pdf. KMV Corporation.

Wang, H., Forbes, C. S., Fenech, J.-P., and Vaz, J. (2020). The determinants of bank loan recovery rates in good times and bad–new evidence. *Journal of Economic Behavior & Organization*, 177:875–897.

Witzany, J. (2009). Unexpected recovery risk and LGD discount rate determination. *European Financial and Accounting Journal*, 2009(1):61–84.

Chapter 2

Article 1: A Cost of Capital Approach to determining the LGD Discount Rate

The first article, *A Cost of Capital Approach to determining the LGD Discount Rate*, is intended for submission to *ORiON*. A summary of the guidelines to authors from the journal is presented below.

Manuscript	The manuscript should be of suitable length. No page limit is specified.
Title	A suitable title should be provided. No word limit is specified.
Abstract and keywords	Papers should be preceded by an abstract not exceeding 300 words in length. A list of suitable key words should be listed directly after the abstract. Authors should consult the list of official ORiON keywords, and preferably only use words in this list.
Figures and Tables	Figures and Tables should be numbered consecutively, using separate numbering sequences. Tables and figures should be accompanied by detailed captions and should be included in the main body of the text (not on separate pages at the end of the manuscript).
References	Authors have a choice whether to follow the Harvard (author date) standard or the Vancouver (numerical) standard for literature citations. One of these standards should be applied consistently.
General formatting	A LaTeXtemplate is provided for submission.
Additional information	https://orion.journals.ac.za/pub/information/authors

Operations Research Society of South Africa

Submitted for publication in ORiON

Operasionele Navorsings-vereniging van Suid-Afrika

A cost of capital approach to determining the LGD discount rate

Janette Larney[a], Gerrit Lodewicus Grobler[b], Helgard Raubenheimer[a], and Arno Botha[a]

[a] *Centre for Business Mathematics and Informatics, North-West University, South Africa*
[b] *Department of Statistics, North-West University, South Africa*

Abstract

Loss Given Default (LGD) is one of the key risk parameters in determining a bank's regulatory capital. With the popular workout method to estimate LGD, realised recovery cash flows are discounted at an appropriate rate, and compared with the exposure at default. Regulatory guidance mandates that this discount rate allows for the time value of money, and include a risk premium that reflects the "undiversifiable risk" of the recovery cash flows. In this study we propose an approach to determine the LGD discount rate, inspired by the Solvency II regulatory regime's cost of capital approach. The method involves estimating a market-consistent price for a portfolio of defaulted loans, from which a market-consistent discount rate may be inferred. We apply the methodology to mortgage and personal loans data of a large South African bank. The results reveal the main drivers of the discount rate as the mean and variance of recovery cash flows, and the bank's cost of capital (in excess of the risk-free rate). The proposed methodology satisfies the regulatory principles that it should reflect the undiversifiable risk of recovery cash flows, and therefore, represents a viable option for both regulatory and economic capital modelling purposes.

Key words: loss given default, discount rate, risk premium, cost of capital, Solvency II

1 Introduction

One of the key risk parameters used in determining a bank's regulatory capital under the Basel II Capital Accord of the Basel Committee on Banking Supervision (BCBS) [6] is the *loss given default* (LGD). Denoted as the random variable $L \in [0, 1]$, a loan's LGD is the fraction of its balance B_{τ_d} (or the *exposure at default*, EAD) that was eventually lost on a defaulted account, as at the time of default τ_d. The LGD is often described as the "economic loss" on a financial instrument, which includes both material discount effects and material direct/indirect collection costs; see BCBS [6, §460], itself corroborated by the European Banking Authority (EBA) in [15].

In nursing the strained relationship between bank and borrower back to health, collection staff can pursue a variety of remedial actions during this so-called *resolution/workout* *period*. From Van Gestel & Baesens [46, pp. 26-28], Finlay [20, pp. 11-13], and Botha [8, §2.2], these remedial actions commonly include: 1) temporarily lowering instalments and/or zeroing interest rates; 2) extending a payment 'holiday' during which instalments are suspended; and 3) restructuring the loan in the borrower's favour. Successful collection efforts imply the distressed borrower repaying a newly-negotiated stream of cash flows over time and/or lump sums, both of which are called *receipts*. These receipts should diminish the arrears balance and decrease the delinquency, ultimately 'curing' the loan back to health; resulting in a zero credit loss $L = 0\%$. Conversely, collection efforts may fail and erode the credit relationship beyond repair. The bank will now initiate legal proceedings and/or foreclose on any available collateral, in recovering as much of the defaulted debt in record time. Any last receipts resulting from debt recovery are offset against the loan's last-known balance, with the non-zero remainder being written (or charged) off as a credit loss, i.e., $L > 0\%$. Accordingly, a defaulted loan will resolve into either a *cured* or a *written-off* outcome, with the remaining unresolved cases being right-censored (pending an outcome). These ideas are illustrated in Fig. 1.

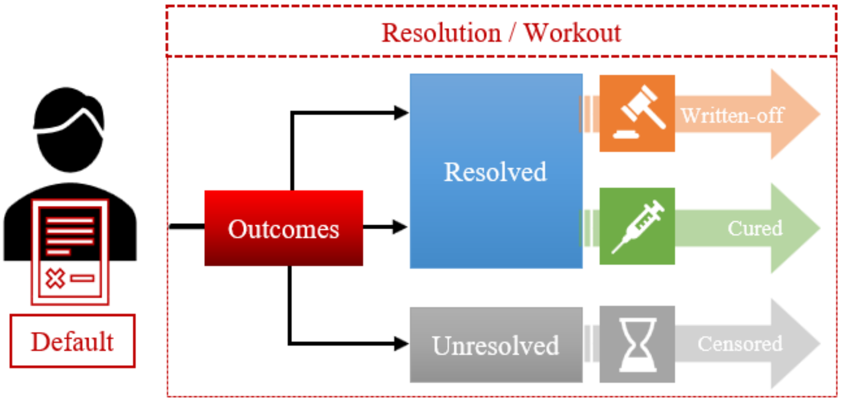

Figure 1: *Illustrating the typical process of resolving a defaulted loan into either a written-off or cured outcome during the workout period. The remaining defaults are considered as right-censored. From Botha [8, pp. 59].*

Under the advanced *internal rating based* (IRB) approach of Basel, a bank can produce its own LGD-estimates. From Schuermann [40], Van Gestel & Baesens [46, pp. 217-222], and Baesens et al. [2, §10], there are four common methods for estimating the LGD of either retail or corporate exposures. Of these methods, we shall focus on the more popular *workout* method, which caters to both retail and corporate exposures. At each month-end time period t, let X_{it} denote the net cash flow (or receipt minus all costs & benefits) of account $i = 1, 2, ..., n$, as observed from data. Similarly, let B_{it} record the month-end outstanding balance or utilised amount of a credit limit, as observed from data. For a written-off loan i, the realised loss $l_i(u)$ is then calculated at (and from) a given time u as the proportion of B_{iu} that was ultimately lost. This given time u can range from the moment of default τ_d of loan i, up until its resolution time $\tau_r \geq \tau_d$ (write-off); see Gürtler & Hibbeln [24] and Baesens et al. [2, §10]. Using an appropriate factor v_m that discounts a quantity m periods backwards, this realised loss is expressed more formally as

$$l_i(u) = \frac{B_{iu} - \sum_{t=u}^{\tau_r} X_{it} v_{t-u}}{B_{iu}} = 1 - \frac{\sum_{t=u}^{\tau_r} X_{it} v_{t-u}}{B_{iu}} \quad \text{for } u = \tau_d, \ldots, \tau_r \,. \tag{1}$$

In preparing the modelling dataset, Eq. 1 is typically evaluated only once at τ_d of each resolved defaulted loan that was eventually written-off. By implication, these $l_i(\tau_d)$-values are considered as realisations from L in respect of the defaulted loan i.

In Eq. 1, selecting the appropriate discount rate in v_m can be non-trivial and even contentious in correctly accounting for the time value of money. Moreover, the cumulative effect of any discount rate on $l_i(\tau_d)$ becomes particularly pronounced for longer resolution periods, e.g., as in residential mortgages. Van Gestel & Baesens [46, pp. 220-221] note that the most appropriate rate ought to be the "risk-free" rate r_f. However, using any rate other than the contract (or client) rate may produce illogical $l_i(\tau_d)$-values, particularly for defaulted exposures that dutifully repaid instalments on time, as the discounted present value of the instalments will differ from the loan amount.

From a regulatory perspective, BCBS [5] offers principled (but limited) guidance on selecting an appropriate discount rate, which is arguably inspired by the seminal work of Markowitz [33]. In short, discount rates must reflect the time value of money (r_f) and include an appropriate risk premium δ (for "undiversifiable risk"), which compensates for the uncertainty of receiving X_{it}. Accordingly, we rewrite the discount factor v_m in Eq. 1 as

$$v_m = \frac{1}{(1 + r_f + \delta)^m} \,. \tag{2}$$

Some regulators have adopted this principle *verbatim* in selecting a discount rate, e.g., see the retail credit template of the South African regulator, the South African Reserve Bank [43, p. 20]. Others have digressed somewhat, e.g., the EU-regulator (European Banking Authority) proposed using a discount rate comprised of the primary interbank rate plus 5%; see European Banking Authority [14]. Though simple, this proposal deliberately ignores a bank's funding costs when assessing the uncertainty of receiving X_{it}; itself affected by competent collection staff and associated indirect costs. Similarly, the UK-regulator (Prudential Regulation Authority) suggests a discount rate that consists of the *Sterling Overnight Index Average* (SONIA) as the interbank rate, plus 5%; see PRA [36] and PRA [37]. In addition, the PRA imposes a minimum of 9% for discount rates used when estimating downturn LGD-values.

As an alternative, we propose a so-called *cost of capital* (CoC) approach for inferring δ based on a market-consistent (or fair) price for defaulted exposures. Inspired by the Solvency II (SII) regulatory regime for insurers [16], our CoC-approach can produce a δ that reflects the undiversifiable risk within X_{it}. Our work is closest to that of Altman & Pompeii [1], Brady et al. [9], and Jacobs [29], wherein the discount rate was similarly inferred from market prices of defaulted debt. These studies rely on the *"informational efficient market"* proposition of Fama [17], which assumes that market prices incorporate all available information about future values. However, these procedures fail in the absence of such market prices, which is often the case for retail loan portfolios. Our CoC-approach can generate such market-consistent (but synthetic) prices, itself driven by the cost of holding capital against uncertain cash flows.

The remainder of this paper is organised as follows. In section 2 an overview of the most relevant literature, as it pertains to the determination of an appropriate LGD risk premium, is provided. In section 3 we introduce the SII regime, with a discussion of the CoC-approach to determine a market-consistent price (MCP) for a stream of uncertain cash flows. The application of the CoC-methodology to determine a MCP for a portfolio of defaulted loans is then described, and we show how a market-consistent discount rate can be inferred from this MCP. In section 4 we describe the two data sets, each relating to a different retail portfolio of a large South African bank, used to illustrate our proposed methodology. The results of our application is reported in section 5, and in section 6 we provide some concluding remarks.

2 Earlier methods for determining the risk premium δ

Much of the literature on determining an appropriate LGD discount rate attempts to infer a market-related discount rate from the observed market prices of defaulted debt (typically corporate bonds and traded bank loans). These studies include those by Brady et al. [9], Altman & Pompeii [1] and Jacobs [29]. In theory, provided a deep and liquid market for the debt instrument in question exists, this approach should yield an objective and market consistent rate that reflects the market's view of the risk inherent in the recovery cash flows. In many cases, however, secondary markets for bank debt do not exist and in other cases the market may be highly illiquid [30]. In such cases, a different approach is required.

A variety of approaches used in practice to determine the LGD discount rate have been identified and evaluated by e.g., Gibilaro & Mattarocci [22], Maclachlan [32] and Witzany [48]. Scheule & Jortzik [39] go further to benchmark five of these methods with reference to their economic robustness and empirical implications. The authors confirm that the contract or "effective loan interest rate" remains the most popular in practice. The authors also conclude (in line with others - see e.g., Maclachlan [32] and Chalupka & Kopecsni [12]), that the use of this rate contradicts the guiding principles of the Basel requirements. A coherent formulation of this argument is provided by Jacobs [29]: *"The bank is an investor in a new financial claim post default, which is no longer a financial claim having promised payments subject to default risk, but rather an asset dependent upon the recoverability of a defaulted firm's assets or of collateral values. The appropriate discount rate depends upon the degree of undiversifiable risk inherent in this new asset, and it may be less or more (and by no means necessarily equal to) than that on the pre-existing claim."*

Very few studies offer proposals on how to derive a risk premium, δ, that reflects the uncertainty of post default recoveries in the absence of reliable market price data. Maclachlan [32] sets out a method, based on the *capital asset pricing model* (CAPM) introduced by Sharpe [41], that determines an asset class risk premium based on the Basel asset correlation factor, the asset volatility, market volatility and a market risk premium (MRP). Gibilaro & Mattarocci [22] extend Maclachlan [32]'s approach by proposing a multi-factor model based on systematic and specific factors that explain LGD volatility and in this manner improves on the broad brush approach of applying the Basel prescribed asset correlation factors. In both studies the risk premium is a function of the MRP despite very little agreement amongst practitioners and academics alike on what constitutes an appropriate MRP [42].

Also making use of the CAPM, Witzany [48] estimates the market implied "cost of risk capital" from the relationship between the MRP and the Basel equity portfolio risk measure. The author then determines δ by multiplying the cost of risk capital with economic capital (EC) in respect of "unexpected undiversifiable Loss Given Default" risk.

Our proposed methodology shares this premise: post default loss risk must be backed by EC, and δ should in some way reflect the cost of holding EC against this risk. Our proposition is that this cost of capital should be reflected in the fair price an investor, accepting the loss risk, would be willing to pay for a portfolio of defaulted loans. This is the principle underlying the valuation approach of the SII risk-based capital regime (discussed below) and the basis for our CoC-methodology to determine an LGD discount rate.

3 Solvency II and the cost of capital approach for market consistent valuation

SII is a regulatory regime aimed at insurance and reinsurance undertakings in the European Union (EU) [16]. Several non-EU countries, including South Africa[1], Australia and Singapore have adopted SII-equivalent risk-based capital regimes [19]. Similar to the Basel regulatory regime for banks, SII is based on the principle that institutions should hold enough capital to cover the potential (unexpected) future losses that could result from their risky activities. SII and Basel do however differ in the determination of capital requirements, and also in the specific rules and regulations they impose.

One of the key components of SII is the *economic balance sheet* approach, where capital requirements are based on a balance sheet that reflects the economic value of a firm's assets and liabilities. Both assets and liabilities must be valued at *fair value*, i.e., "the price that would be received to sell an asset or paid to transfer a liability in an orderly transaction between market participants at the measurement date", as defined by IASB [27]. In most cases, the fair value is simply the current market value of an asset or liability, or alternatively, the market value of a *replicating portfolio* [7].

Where no deep, liquid, and transparent market exists, SII requires an asset or liability to be valued in a market consistent manner. For this, the *Best Estimate plus Risk Margin* approach has been adopted, where the best estimate present value of future cash flows is adjusted by a *Risk Margin*[2] (RM). The "best estimate present value" is defined by SII as "the probability weighted average of future cash flows, taking account of the time value of money", i.e., it is the expected present value where cash flows are discounted at the risk-free rate.

The idea with the RM is that a potential buyer of liabilities will need to be compensated for the cost of holding assets in excess of the value of the best estimate cash flows, to protect against the risk that the best estimate value was too low. Similarly, the buyer of an asset will need to be compensated for the uncertainty surrounding the best estimate value of the asset's cash inflows. The RM is therefore needed to compensate the providers of capital, whose funds are used to absorb the risk assumed. The RM must then be subtracted from (or added to) the best estimate value of assets (or liabilities) to determine the market consistent price (MCP) of the cash flows.

The RM is calculated as the cost of holding the required amount of capital over the

[1]Solvency Assessment and Management (SAM) represents a "Solvency II equivalent" risk-based supervisory approach for the prudential regulation of South African long-term and short-term insurers. See https://tinyurl.com/tawdn8mz

[2]Also referred to as the Market Value Margin.

run-off period of the risky cash flows. This is the so-called *cost of capital* (CoC) approach prescribed by the SII regime [35]. It is based on the principle that the price of an asset or liability should reflect the cost of holding risk capital against the risk inherent in that asset or liability over its lifetime.

To determine a MCP for a portfolio of defaulted debt, the price that a potential investor would be willing to pay for the risky recovery cash flows needs to be calculated. We do this using the CoC-approach described above: The risky cash flows, being the expected recoveries from the defaulted loans, are discounted at the risk-free rate, and from this best estimate present value, the cost of holding capital (against the risk of recovering less than the expected amount) is deducted. The capital required at any future time t is discounted at the risk-free rate back to the transaction date to reflect the fact that the capital buffer will be invested in assets earning the risk-free rate.

Two inputs are required to determine the RM: The capital requirement as well as the *cost of capital rate*. EC in respect of post default loss risk is discussed in subsection 3.1 below. As far as the *cost of capital rate* is concerned, the net cost of holding risk capital must be considered. This is the marginal cost of raising (and holding) risk capital, less interest that can be earned at the risk-free rate. For the purpose of this study we regard a bank's own cost of equity (in excess of the risk-free rate) to represent a realistic view of the cost of raising and holding the necessary risk capital. In an ideal world such a rate could be prescribed by the regulator, in a similar way that SII prescribes a CoC-rate of 6% for insurance and reinsurance undertakings.

3.1 Regulatory and economic capital for post default credit risk

Regulatory capital (RC) under the IRB approach of the Basel capital adequacy regime is based on the Asymptotic Single Risk Factor (ASRF) model, first introduced by Gordy [23]. Under the ASRF model, the losses of a well diversified portfolio of loans depend only on the (single) systematic risk factor (widely interpreted as reflecting the state of the global economy) and not on idiosyncratic risk factors associated with individual exposures. The model takes the bank's estimate of average probability of default (PD) as input, as well as as an estimate of "downturn LGD", which must reflect losses under economic downturn conditions. LGD and EAD are treated as constants in the model, but the entire distribution of the annual PD is derived through a supervisory mapping function that transforms the average PD into a conditional PD. The mapping function, which is essentially a mathematical formula or algorithm, takes the average PD and applies a transformation to derive the conditional PD. The conditional PD represents the estimated probability of default under specific conditions or scenarios, such as an economic downturn. The transformation helps adjust the average PD to reflect the potential impact of different factors that may affect the likelihood of default. From the distribution of the conditional PD, a credit loss distribution is derived and the "unexpected loss" (UL) of the portfolio is determined as the difference between the Value-at-Risk, at a 99.9% confidence interval, of the loss distribution. and the expected loss ($EL = EAD \cdot PD \cdot LGD$) of the loan portfolio. The RC amount is then equal to UL .

Application of the ASRF Basel model to a portfolio of defaulted loans (i.e., where the $PD = 1$) results in a zero RC requirement. This is because, under the model's simplifying assumptions, when $PD = 1$ the "worst-case" loss is equal to the EL, and no capital charge is therefore required to cover UL. Notwithstanding the zero RC requirement, a significant body of academic and practitioner research has identified systematic volatility in recovery rates as a potentially important source of unexpected credit losses; see e.g., Frye et al. [21], Düllmann & Gehde-Trapp [13] and Caselli et al. [11]. The BCBS also deems a capital charge for defaulted assets desirable "in order to cover systematic uncertainty in realised recovery rates for these exposures" [3]. It is therefore expected from banks to determine an appropriate amount of EC in respect of post default loss risk. Most studies to assess EC in respect of loss risk rely on simulation approaches and attempt to capture the dependence between the default event and the subsequent loss risk. These include the studies by Heid & Krüger [25], Rösch & Scheule [38] and Kaposty et al. [31].

Closed-form expressions for EC in respect of post default loss risk have, in turn, been proposed by Weißbach & von Lieres und Wilkau [47] and Tasche [45]. The latter's model is, akin to the ASRF model, a "single factor" model, but unlike in the case of the ASRF model, LGD is treated as a random variable, and allowance is made for PD and LGD to be correlated. The model uses a parametric beta distribution to represent the loss random variable, L, and incorporates expected LGD and the LGD variance as parameters. It calculates the EC amount as the difference between the Value-at-Risk (VaR_α) of the loss distribution, determined at a suitably high confidence level, $1-\alpha$, and the expected loss of the loan portfolio. Despite its shortcomings[3], the beta distribution, with support on $[0,1]$, and able to represent a variety of shapes, remains a popular choice for characterising LGD. It is also fairly straightforward to estimate the two parameters of the distribution, a and b, by moment matching, given the historical average and variance of realised loss data.

A critical review of EC models for loss risk is beyond the scope of this paper. Due to its simplicity, and ease of implementation, we employ the single risk factor model proposed by Tasche [45] to determine the EC amount, and to then use as input in our CoC-methodology. In line with the Basel requirements (see BCBS [4, pp. 11-12]) we apply a confidence interval of 99.9% over a one-year horizon. However, considering the fact that the capital amount used represents EC, rather than RC, a bank should consider its own strategic objectives and risk appetite in deciding on the appropriate risk measure, confidence level, and time horizon to use.

3.2 Determination of a market consistent discount rate

Given realisations $l_i(\tau_d), i = 1, 2, ..., n$ from L, where τ_d is the time of default, the expected LGD, $\mathbb{E}[L]$, can be estimated by

$$\bar{L} = \frac{1}{n}\sum_{i}^{n} l_i(\tau_d) \tag{3}$$

and the unbiased estimator of the LGD variance, $\mathbb{V}[L]$, by

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(l_i(\tau_d) - \bar{L}\right)^2. \tag{4}$$

Our aim is to determine a MCP for a portfolio of defaulted debt, and from this, a market consistent risk discount rate, $r = r_f + \delta$. Given an appropriate amount of EC (as

[3]Several studies have shown that the beta distribution often does not provide a good fit for LGD data, especially if it exhibits a bimodal-shape (see e.g., Calabrese & Zenga [10] and Huang & Liang Xie [26])

discussed in subsection 3.1, and which is a function of $\bar{L}$ and σ^2) at time t, denoted by C_t, and an appropriate cost of capital rate, c, the estimated RM for the portfolio of defaulted loans can be calculated as

$$R(r_f) = \sum_{t=\tau_d}^{\tau_R} \frac{c\, C_t}{(1+r_f)^{t-\tau_d}}, \tag{5}$$

where $\tau_R = \max\limits_{1 \le i \le n} \tau_{ir}$, where τ_{ir} denotes the resolution time of the i^{th} loan.

If we denote the portfolio level recoveries by $X_t = \sum_{i=1}^{n} X_{it}$, then the best estimate present value of all recovery cash flows, Y, is given by

$$Y(r_f) = \sum_{t=\tau_d}^{\tau_R} \frac{X_t}{(1+r_f)^{t-\tau_d}}, \tag{6}$$

and the estimated MCP for the portfolio at time τ_d is then

$$P(r_f) = Y(r_f) - R(r_f). \tag{7}$$

An investor purchasing the portfolio of defaulted loans will therefore pay P, and receive the recovery cash flows, X_t, at times $t = \tau_d, ..., \tau_R$, in return. The expected rate of return, r, on the transaction can be determined by solving for r in:

$$P(r_f) = Y(r), \tag{8}$$

The expected rate of return, $r = r_f + \delta$, calculated from Eq. 8 represents the fair return that an investment in a portfolio of defaulted loans is expected to yield over the term of the recoveries. It incorporates a risk premium, δ, to compensate the investor for the risk of purchasing an uncertain income stream. The risk premium is driven by the cost of raising and holding the required amount of EC in respect of the uncertain cash flows. As such, we argue that this expected rate of return represents an appropriate LGD discount rate for the portfolio of loans in question.

4 Data and model calibration

4.1 Data

We illustrate our methodology by applying it to two datasets provided by a large South African bank. The first dataset comprises recovery data of an unsecured personal loans (PL) portfolio, whereas the second dataset relates to that of a portfolio of secured residential mortgage loans (ML).

Our data includes cash flows (both positive and negative) received on accounts that defaulted in the period October 2005 to September 2011, i.e., six years in total. Included in this period is a two year period that has been identified as representing a "downturn period" for each portfolio. We define the downturn period as the 24 months that gave rise to the highest rates of default in an economic cycle. For the PL portfolio this period spans 1/12/2007 to 30/11/2009 and for the ML portfolio it is from 1/9/2007 until 31/8/2009. Unsurprisingly these downturn periods correspond (roughly) with the global financial crisis (GFC) that gained traction in the second half of 2007 [34]. We consider both the full, and applicable downturn period, for each portfolio separately.

Table 1: *Summary statistics for PL and ML portfolios over the full and downturn periods.*

	$\bar{L}$	$\hat{\sigma}$	Term (in months)	
			Maximum	Average
Personal Loans				
Full	0.748	0.291	87	4.67
Downturn	0.760	0.273	80	4.57
Mortgage Loans				
Full	0.256	0.366	113	25.64
Downturn	0.293	0.372	90	33.59

The summary statistics for the two portfolios are shown in Table 1. Besides the difference in average LGD ($\bar{L}$), the significantly different average terms of the recoveries for the two portfolios are noteworthy, albeit not surprising. Given the lack of collateral, banks are typically agile in recognising a futile recovery process in the case of unsecured loans, and will normally choose to avoid long and expensive workout processes. In contrast, most owner-occupier mortgage borrowers will engage with the bank to arrange a payment holiday or a more affordable repayment schedule when faced with the prospect of losing their home. Even if the property is eventually foreclosed on, the legal and operational processes attached to this action typically take years, rather than months, to resolve. Consequently, the average LGD for the PL portfolio is also much higher than that for the mortgage portfolio. The lower average LGD for the mortgage portfolio is however accompanied with a higher standard deviation ($\hat{\sigma}$). Interestingly the means and standard deviations for the PL portfolio are very similar for the two periods under consideration, whereas the average LGD in respect of ML in the downturn period is somewhat higher at 29.3% compared to 25.6% over the full period.

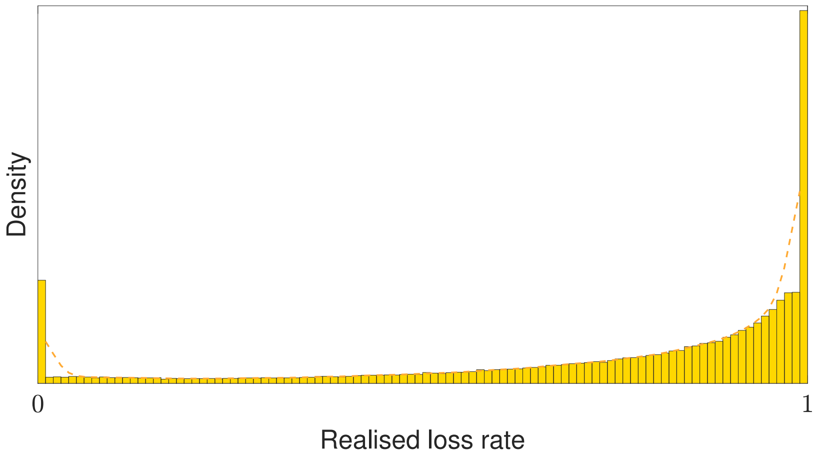

Figure 2: *Histogram of realised loss rates for the personal loans portfolio over the full reference period. Due to confidentiality requirements, no actual values are shown.*

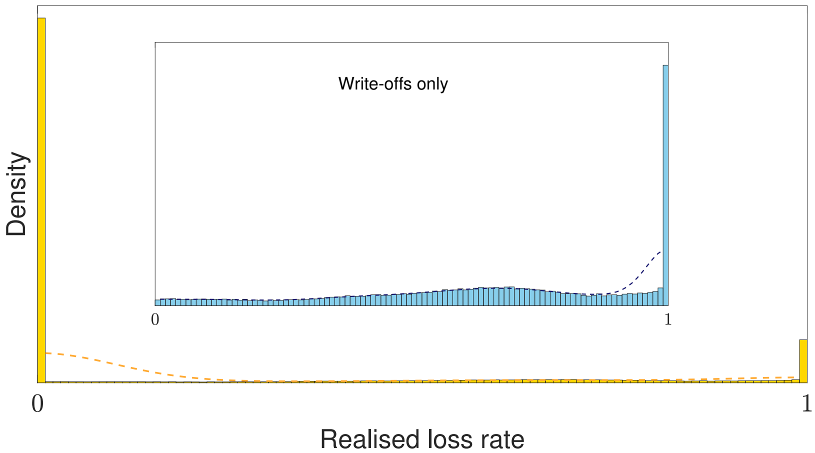

Figure 3: *Histogram of realised loss rates for the mortgage loan portfolio over the full reference period. The outer graph is in respect of all data (including "cures"), while the insert depicts loans that were written off. Due to confidentiality requirements, no actual values are shown*

Fig. 2 and Fig. 3 show the distribution of the aggregated LGD data over the full period for the PL and ML portfolios, respectively. For the ML portfolio Fig. 3 shows both the full dataset as well as that for non-zero losses. As expected, considering its secured nature, the ML portfolio has a substantial proportion of accounts resulting in a "no loss", known as "cured" accounts, whereas these constitute a minor proportion for the unsecured PL portfolio. The higher standard deviation of the ML portfolio can mainly be attributed to the high proportion of these "no loss" accounts.

In Fig. 4 the proportions of LGD observations at 0 (dark blue bars), on the interval $(0, 1)$ (mid blue bars) and at 1 (light blue bars) are shown for each month over the reference period. The areas shaded in red denote the downturn periods for each portfolio. We can observe that the ML portfolio experienced a marked drop in the proportion of "cures"

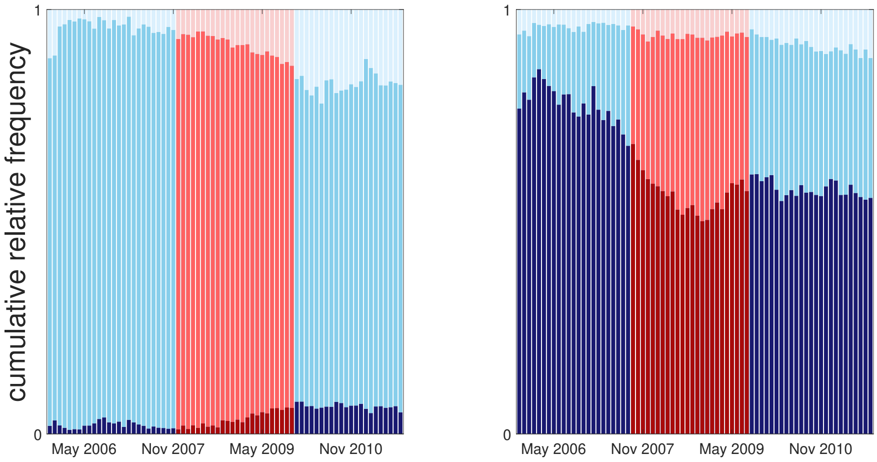

Figure 4: *Proportion of LGD observations for each month at* 0 *(dark blue), on the interval* $(0,1)$ *(mid blue), and at* 1 *(light blue) in respect of Personal Loans (left pane) and Mortgage Loans (right pane). Downturn months are shaded in red.*

during the downturn period, and this proportion did not return to pre-downturn levels before the end of the reference period. A similar pattern emerges in the PL portfolio, where the proportion of total losses continue to increase well beyond the downturn period and then stabilises. This suggests that the effects of economic downturn continued well into 2011. The ML data suggests that the recovery of residential property prices (collateral) possibly lagged economic recovery, resulting in greater loan-to-value ratios and fewer defaulted accounts with full recoveries. For both PL and ML the trends could also be indicative of a change in the bank's underwriting standards. Following the severe losses suffered by banks during the GFC, many banks attempted to increase earnings by relaxing underwriting standards in pursuit of market share, as reported by the Federal Advisory Council and Board of Governors [18]. This factor might also be at play in our dataset.

4.2 Model calibration

In line with the Basel requirements, we determine $C_t(r_f, \delta)$ at the start of each of the run-off years for our portfolios of defaulted loans at the 99.9% confidence interval. We implement the model of Tasche [45], using moment matching to estimate the parameters of the beta distribution.

In order to calculate C_t the market consistent discount rate, $r = r_f + \delta$, is used. This complicates the method to obtain r in Eq. 8. We solve this circularity problem using an iterative approach by minimising the distance between the discounted recoveries $Y(r_f + \delta)$ and the market consistent price $P(r_f)$ by applying the following algorithm:

1. Set an initial value for $\delta = \delta_0$ and calculate C_t.

2. For a given cost of capital rate c, portfolio level recoveries $X_t, t = \tau_d, ..., \tau_R$ and risk free rate r_f, calculate P in (*Eq. 7*).

3. Determine δ_i by minimising:

$$\delta_i = \min_{\delta} \{Y(r_f + \delta) - P\}^2 .$$

4. Update $\delta = \delta_i$ and repeat steps 1 to 3 above until δ_i converges (according to a pre-determined convergence criterium.)

The average risk-free rate (r_f) for each portfolio (over the full and downturn periods) has been calculated by first interpolating between the closest available tenors on the South African yield curve for each trading day within the relevant period, and then calculating the average of these interpolated rates.

5 Results

The results from our implementation (with the cost of capital rate, c, chosen[4] at 7%) are shown in Table 2. Besides the calculated risk-free rate, r_f, the discount rate, r, and resultant risk premium, δ, the total EC amount as a proportion of the price paid, P, for each portfolio and reference period, is shown. The much higher risk-free rates of the downturn periods reflect the turbulent market conditions over this period, and the associated surging bond yields. The effect is less pronounced for the ML portfolio, due to the longer averaging period.

Table 2: *Results from implementation with the cost of capital rate, c, chosen as* 7%.

	c	r_f	r	δ	EC/P	$\bar{L}$	$\hat{\sigma}$
			Personal Loans				
Full	7%	6.08%	14.91%	8.83%	1.452	0.748	0.291
Down	7%	8.89%	17.01%	8.12%	1.431	0.760	0.273
			Mortgage Loans				
Full	7%	6.96%	12.36%	5.4%	1.468	0.256	0.366
Down	7%	8.49%	13.99%	5.5%	1.653	0.290	0.370

It can be seen from Eq. 5 and Eq. 7 that besides the RM, R, there are two further drivers of the MCP, and therefore the discount rate, r. These are the expected recoveries, Y, and the risk free rate, r_f. R, in turn, has c, as well as $\bar{L}$ and $\hat{\sigma}$, as drivers.

Considering the full period, it is interesting to note that the EC amounts for the PL and ML portfolios are almost identical, despite having very different values for $\bar{L}$. This is due to the significantly longer term (average of 25.64 months vs. 4.67 months) of the ML portfolio compared to the PL portfolio. On average EC therefore needs to be held for more than five times as long in respect of a defaulted mortgage loan compared to a personal loan.

Despite the very similar EC amounts, and the lower average risk free rate, r_f, the discount rate, r, is considerably higher for the PL portfolio than for the ML portfolio

[4]The choice of $c = 7\%$ is based on the Standard Bank Group's reported average cost of equity of 14% over the five years from 2007 to 2011, with an average risk-free rate of 7.2% over the same period [44].

over the full period. This is attributable to the much lower average recovery rate (1 - $\bar{L}$) for PL's, resulting in the risk margin, R carrying a significantly higher weight than the recoveries in the determination of r.

When comparing the ML portfolio over the downturn period with that over the full period, the EC as a proportion of P is considerably higher (1.653 vs 1.468). This is attributable to both the higher $\bar{L}$ (29% vs 25.6%) and the longer average recovery period (33.59 vs. 25.64 months). The lower recovery rate and higher EC amount result in a higher discount rate, r. However, because of the higher risk free rate, r_f, over the downturn period, the risk premiums are in fact very similar (5.4% vs 5.5%).

For the PL portfolio it is notable that both the variance and the EC amount is slightly higher for the full period than for the downturn period. Considering that the full period encompasses the downturn period, and then also include divergent (prosperous) months with much lower losses, this result seems sensible. When comparing the PL discount rates, it is evident that the downturn period produces a much higher rate than the full period, despite fairly similar values for $\bar{L}$ and $\hat{\sigma}^2$. This can almost entirely be ascribed to the higher risk free rate, r_f, that applied over the downturn period. The risk premium, δ, for PL is slightly lower than that for the full period due to the considerably higher risk free rate, over that period. For ML the downturn risk premium is only marginally higher than for the full period, due to the longer recovery term. This result is consistent with that reported by [39]. The authors find lower risk premiums associated with periods of higher risk free rates when applying a variety of market equilibrium methods to determine the LGD discount rate risk premium.

We now turn our attention to the effect of the cost of capital rate, c, on the discount rate. In Table 3 the discount rate, r, and resultant risk premium, δ, are shown for each loan portfolio, and for each reference period for $c = 6\%, 7\%$ and 8%.

Table 3: *Results from implementation for different values of $c = 6\%, 7\%, 8\%$ in respect of both the full period as well as the downturn period, for personal loans and mortgage loans.*

	c	r_f	r	δ	EC/P	$\bar{L}$	$\hat{\sigma}$
			Personal Loans				
Full	6%	6.08%	13.6%	7.52%	1.445	0.746	0.293
	7%	6.08%	14.91%	8.83%	1.452	0.748	0.291
	8%	6.08%	16.22%	10.14%	1.459	0.750	0.288
Down	6%	8.89%	15.81%	6.92%	1.424	0.758	0.275
	7%	8.89%	17.01%	8.12%	1.431	0.760	0.273
	8%	8.89%	18.22%	9.33%	1.436	0.762	0.270
			Mortgage Loans				
Full	6%	6.96%	11.53%	4.57%	1.445	0.253	0.364
	7%	6.96%	12.36%	5.4%	1.468	0.256	0.366
	8%	6.96%	13.22%	6.26%	1.492	0.258	0.367
Down	6%	8.49%	13.13%	4.64%	1.624	0.288	0.369
	7%	8.49%	13.99%	5.5%	1.653	0.290	0.370
	8%	8.49%	14.86%	6.37%	1.684	0.293	0.372

With its low recovery rate (approximately 25%, given $\bar{L}$ of between 74.6% and 76.1%), a higher value of c has, as expected, a much more pronounced effect on the discount rate of the PL portfolio, than on that of the ML portfolio. As can be observed from Table 3, a 1% increase in c results in an approximate increase of 1.2%-1.3% in r for the PL portfolio (over the full and downturn periods), versus a 0.83%-0.88% increase in r for the ML portfolio. The more pronounced impact on the PL portfolio is, as previously highlighted, due to the cost of capital amount having a more pronounced effect (due to its higher weight) in the determination of r, when recoveries are lower.

6 Concluding remarks

We have demonstrated how, by calculating a market consistent price for a portfolio of defaulted debt, a market consistent discount rate can be inferred. The determination of a market consistent price relies on the cost of capital approach of the Solvency II regulatory regime. For our implementation we used the single factor model proposed by Tasche [45] to calculate economic capital in respect of post default loss risk, but we anticipate that banks will make use of their own internal capital models for this purpose.

The main drivers of the LGD discount rate are the mean and variation of recovery cash flows, the average risk free rate applicable over the reference period, the average term of the cash flows, and the cost of capital rate. As can be expected, higher volatility (relative to the expected value) of recovery cash flows is associated with more risk, and hence, a higher discount rate. The same holds for longer recovery terms. Similar to the findings of Scheule & Jortzik [39], we find that a higher risk free rate is, ceteris paribus, associated with a lower risk premium, leaving the overall discount rate with surprising little variability over time.

Regardless of other drivers, the discount rate is fairly sensitive to the cost of capital rate, more so when recovery rates are low. This seems sensible, considering the cost of raising and servicing risk capital, and the requirement for economic capital to guard against variability in post default recoveries. Also, given the fact that the cost of equity (in excess of the risk-free rate) should reflect an individual bank's overall riskiness and its cost of funding, this sensitivity seems justified.

As reported by Scheule & Jortzik [39] the contract rate remains the most popular for use in practice, presumably in part due to its prescription for provisioning purposes by the IFRS Accounting Standard [28]. Due to data limitations, a comparative study between the proposed CoC-discount rate and the average contract rate on the ML and PL portfolios was excluded from the scope of our study. To provide a more comprehensive analysis of the CoC-discount rate, future research could consider such a comparative study to fully explore the difference between the CoC-risk premium and that embedded in the loan contract rate.

The proposed methodology satisfies the regulatory requirement for the LGD discount rate to account for the time value of money and to include a risk premium that reflects the undiversifiable risk of recovery cash flows. The methodology can easily be applied to through-the-cycle and downturn conditions and, as such, could be regarded as appropriate for the determination of an LGD discount rate for both regulatory and economic capital purposes.

References

[1] Altman, E. I. & Pompeii, J. (2003). Market size and investment performance of defaulted bonds and bank loans: 1987–2001. *Economic Notes*, 32(2), 147–176.

[2] Baesens, B., Rösch, D., & Scheule, H. (2016). *Credit Risk Analytics: Measurement Techniques, Applications, and Examples in SAS*. Hoboken, New Jersey: John Wiley & Sons.

[3] BCBS (2003). *The New Basel Capital Accord*. Technical report, Bank for International Settlements, Switzerland.

[4] BCBS (2004). *An Explanatory Note on the Basel II IRB Risk Weight Functions*. Technical report, Bank for International Settlements, Switzerland.

[5] BCBS (2005). *Guidance on paragraph 468 of the framework document*. Bank for International Settlements, Switzerland.

[6] BCBS (2006). *International Convergence of Capital Measurement and Capital Standards: A Revised Framework, Comprehensive Version*. Bank for International Settlements, Switzerland.

[7] Boekel, P., van Delft, L., Hoshino, T., Ino, R., Reynolds, C., & Verheugen, H. (2009). Replicating portfolios. an introduction: Analysis and illustrations. https://tinyurl.com/4kw8c7mp. Milliman Research Report.

[8] Botha, A. (2021). *A procedure for loss-optimising the timing of loan recovery under uncertainty*. PhD thesis, University of Pretoria.

[9] Brady, B., Chang, P., Miu, P., Ozdemir, B., & Schwartz, D. (2006). Discount rate for workout recoveries: An empirical study. Available at https://ssrn.com/abstract=907073.

[10] Calabrese, R. & Zenga, M. (2010). Bank loan recovery rates: Measuring and nonparametric density estimation. *Journal of Banking and Finance*, 34, 903–911.

[11] Caselli, S., Gatti, S., & Querci, F. (2008). The sensitivity of the loss given default rate to systematic risk: new empirical evidence on bank loans. *Journal of Financial Services Research*, 34, 1–34.

[12] Chalupka, R. & Kopecsni, J. (2008). *Modelling bank loan LGD of corporate and SME segments: A case study*. Technical report, IES Working Paper.

[13] Düllmann, K. & Gehde-Trapp, M. (2004). Systematic risk in recovery rates: An empirical analysis of US corporate credit exposures. Available at https://ssrn.com/abstract=2793954.

[14] European Banking Authority (2017). *GL/2017/16: Guidelines on PD estimation, LGD estimation and the treatment of defaulted exposures*. European Banking Authority (EBA), France.

[15] European Parliament (2006). Directive 2006/48/EC of the European Parliament and of the Council of 14 June 2006 relating to the taking up and pursuit of the business of credit institutions. *Official Journal of the European Union*, 177, 1–200.

[16] European Parliament (2009). Directive 2009/138/EC of the European Parliament and of the Council of 25 November 2009 on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II). *Official Journal of the European Union L*, 335, 1–155.

[17] Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383–417.

[18] Federal Advisory Council and Board of Governors (2015). FAC Meeting Minutes - May 8, 2015. https://www.federalreserve.gov/aboutthefed/fac-20150512.pdf. Accessed: 2023-01-16.

[19] Ferguson, W. L. (2009). Solvency: Models, assessment and regulation. *Journal of Risk and Insurance*, 76(4), 960.

[20] Finlay, S. (2010). *The Management of Consumer Credit: Theory and Practice*. Hampshire, UK: Palgrave Macmillan, Second edition.

[21] Frye, J., Ashley, L., Bliss, R., Cahill, R., Calem, P., Foss, M., Gordy, M., Jones, D., Lemieux, C., Lesiak, M., Lobbes, C., McGrew, L., Mehta, P., Nelson, J., & Waggoner, E. (2000). Collateral damage: A source of systematic credit risk. *Risk*, 13(4), 91–94.

[22] Gibilaro, L. & Mattarocci, G. (2007). The selection of the discount rate in estimating loss given default. *Global Journal of Business Research*, 1(1), 15–33.

[23] Gordy, M. B. (2003). A risk-factor model foundation for ratings-based bank capital rules. *Journal of Financial Intermediation*, 12(3), 199–232.

[24] Gürtler, M. & Hibbeln, M. (2013). Improvements in loss given default forecasts for bank loans. *Journal of Banking & Finance*, 37(7), 2354–2366.

[25] Heid, F. & Krüger, U. (2011). Do capital buffers mitigate volatility of bank lending? A simulation study. Available at `https://ssrn.com/abstract=2794054`.

[26] Huang, C. & Liang Xie, S. (2012). Modeling Loss Given Default (LGD) by Finite Mixture Model. In *Proceedings of the Northeast SAS User Group 2012 Conference*.

[27] IASB (2011). *International Financial Reporting Standard (IFRS) 13: Fair Value Measurement*. IFRS Foundation: International Accounting Standards Board (IASB), London.

[28] IASB (2014). *International Financial Reporting Standard (IFRS) 9: Financial Instruments*. IFRS Foundation: International Accounting Standards Board (IASB), London.

[29] Jacobs, Jr., M. (2012). An empirical study of the returns on defaulted debt. *Applied Financial Economics*, 22(7), 563–579.

[30] Jankowitsch, R., Nagler, F., & Subrahmanyam, M. G. (2014). The determinants of recovery rates in the US corporate bond market. *Journal of Financial Economics*, 114(1), 155–177.

[31] Kaposty, F., Löderbusch, M., & Maciag, J. (2017). Stochastic loss given default and exposure at default in a structural model of portfolio credit risk. *Journal of Credit Risk*, 13(1).

[32] Maclachlan, I. (2005). Choosing the discount factor for estimating economic LGD. In E. I. Altman, A. Resti, & A. Sironi (Eds.), *Recovery Risk: The Next Challenge in Credit Risk Management*: Risk Books.

[33] Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77–99.

[34] Olbrys, J. (2021). The Global Financial Crisis 2007-2009: A Survey. Available at https://ssrn.com/abstract=3872477. SSRN Working Paper.

[35] Pelkiewicz, A., Ahmed, S., Fulcher, P., Johnson, K., Reynolds, S., Schneider, R., & Scott, A. (2020). A review of the risk margin - Solvency II and beyond. *British Actuarial Journal*, 25, e1.

[36] PRA (2019). *CP21/19: Credit risk: Probability of Default and Loss Given Default*. Bank of England, Prudential Regulation Authority (PRA), United Kingdom.

[37] PRA (2020). *PS11/20: Credit risk: Probability of Default and Loss Given Default*. Bank of England, Prudential Regulation Authority (PRA), United Kingdom.

[38] Rösch, D. & Scheule, H. (2011). A multi-factor approach for systematic default and recovery risk. In *The Basel II Risk Parameters* (pp. 117–135). Heidelberg: Springer.

[39] Scheule, H. H. & Jortzik, S. (2020). Benchmarking loss given default discount rates. *Journal of Risk Model Validation*, 14(3).

[40] Schuermann, T. (2004). What do we know about loss given default? Available at https://doi.org/10.2139/ssrn.525702. Wharton Financial Institutions Center Working Paper.

[41] Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3), 425–442.

[42] Soenen, L. & Johnson, R. (2008). The equity market risk premium and the valuation of overseas investments. *Journal of Applied Corporate Finance*, 20(2), 113–121.

[43] South African Reserve Bank (2018). Long Form Credit Report Template (Retail). Available at https://tinyurl.com/2a32bcbt. SARB, South Africa.

[44] Standard Bank Group (2012). Analysis of financial results, June 2012. Available at https://tinyurl.com/364ayufz. Accessed on 2022-05-16.

[45] Tasche, D. (2004). The single risk factor approach to capital charges in case of correlated loss given default rates. Available at https://ssrn.com/abstract=510982.

[46] Van Gestel, T. & Baesens, B. (2009). *Credit Risk Management: Basic Concepts*. Oxford University Press.

[47] Weißbach, R. & von Lieres und Wilkau, C. (2010). Economic capital for nonperforming loans. *Financial Markets and Portfolio Management*, 24(1), 67–85.

[48] Witzany, J. (2009). Unexpected recovery risk and LGD discount rate determination. *European Financial and Accounting Journal*, 2009(1), 61–84.

Chapter 3

Article 2: Introducing Two Parsimonious Standard Power Mixture Models for Bimodal Proportional Data with Application to Loss Given Default

The second article, *Introducing Two Parsimonious Standard Power Mixture Models for Bimodal Proportional Data with Application to Loss Given Default*, has been published in *Mathematics, 10, 4520. https://doi.org/10.3390/math10234520.* A summary of the guidelines to authors from the journal is presented below.

Manuscript	No restrictions on the length of manuscripts, provided that the text is concise and comprehensive.
Title	The title should be concise, specific and relevant.
Abstract and keywords	An abstract (maximum length of 200 words) in a single paragraph and three to ten pertinent keywords must be included.
Figures, Schemes and Tables	All Figures, Schemes and Tables should be inserted into the main text close to their first citation and must be numbered following their number of appearance. All Figures, Schemes and Tables should have a short explanatory title and caption. All table columns should have an explanatory heading.
References	References must be numbered in order of appearance in the text (including table captions and figure legends) and listed individually at the end of the manuscript.The reference list should include the full title, as recommended by the ACS style guide. In the text, reference numbers should be placed in square brackets [ ], and placed before the punctuation.
General formatting	A LaTeXtemplate is provided for submission.
Additional information	https://www.mdpi.com/journal/mathematics /instructionspreparation

mathematics

MDPI

Article

Introducing Two Parsimonious Standard Power Mixture Models for Bimodal Proportional Data with Application to Loss Given Default

Janette Larney [1,†], Gerrit Lodewicus Grobler [2,*,†] and James Samuel Allison [2,†]

1 Centre for Business Mathematics and Informatics, North-West University, Potchefstroom 2531, South Africa
2 School of Mathematical and Statistical Sciences, North-West University, Potchefstroom 2531, South Africa
* Correspondence: gerrit.grobler@nwu.ac.za
† These authors contributed equally to this work.

Abstract: The need to model proportional data is common in a range of disciplines however, due to its bimodal nature, U- or J-shaped data present a particular challenge. In this study, two parsimonious mixture models are proposed to accurately characterise this proportional U- and J-shaped data. The proposed models are applied to loss given default data, an application area where specific importance is attached to the accuracy with which the mean is estimated, due to its linear relationship with a bank's regulatory capital. In addition to using standard information criteria, the degree to which bias reduction in the estimation of the distributional mean can be achieved is used as a measure of model performance. The proposed models outperform the benchmark model with reference to the information criteria and yield a reduction in the distance between the empirical and distributional means. Given the special characteristics of the dataset, where a high proportion of observations are close to zero, a methodology for choosing a rounding threshold in an objective manner is developed as part of the data preparation stage. It is shown how the application of this rounding threshold can reduce bias in moment estimation regardless of the model choice.

Keywords: proportional bimodal data; parsimonious model; mixture model; rounding threshold; standard power distribution

MSC: 62P05

Citation: Larney, J.; Grobler, G.L.; Allison, J.S. Introducing Two Parsimonious Standard Power Mixture Models for Bimodal Proportional Data with Application to Loss Given Default. *Mathematics* **2022**, *10*, 4520. https://doi.org/10.3390/math10234520

Academic Editor: Andrej Srakar

Received: 26 September 2022
Accepted: 17 November 2022
Published: 30 November 2022

1. Introduction

The need to model proportional data is common in a wide range of disciplines, including the fields of economics, biometrics, as well as social and political sciences [1]. With proportional (or fractional) data, the variable of interest takes on values in the range $[0, 1]$ and bi- or multi-modality is commonly observed.

Given its characteristics and its flexibility to represent a variety of shapes, the beta distribution has been widely used for the modelling of continuous, proportional data, bounded between 0 and 1. For example, in 1974, Falls [2] reinforced this notion in a study related to global cloud cover, by stating that "the beta distribution possesses the versatile statistical characteristics necessary to assume the wide variety of shapes exhibited by global cloud cover". In a biological study, Kousathanas and Keightley [3] also proposed the use of beta distribution to model the bimodal nature of protein-coding loci, where amino-acid changing mutations are either neutral or strongly deleterious. In [4], the beta distribution is used to describe the distribution of consumer probabilities to purchase a brand in a consumer market. In credit risk, beta distribution has long been a popular choice for the modelling of loss given default (LGD) data, as can be seen, e.g., in [5–7]. LGD refers to the fraction of a loan that is not recovered in the case of default and it is usual for LGD to fall in the range $[0, 1]$, where 0 represents a full recovery and 1 represents a total loss of the outstanding amount. Finite mixtures of discretised beta distributions have also been proposed in the literature to deal with data that converge towards one of two opposing sides (this is referred to as polarisation in the literature, as can be seen, e.g., in [8]. Furthermore, Simone and Tutz [9], for example, proposed a mixture of binomial and discretised beta distributions to model uncertainty and response styles in ordinal data.

The ability of the beta distribution to provide a good fit to bimodal data mainly emerges in applications where frequency distributions exhibit either a U-shape or a J-shape, as can be seen, e.g., in [3,4,10,11]. However, beta densities that exhibit U-shapes have asymptotes at 0 and 1, potentially leading to an overestimation of the masses near the endpoints. Related to this property, the masses at the endpoints have an undue impact on the estimated parameters of the fitted beta distribution using maximum likelihood estimation (MLE), leading to adverse effects in the estimation of the moments of the fitted distribution, and most notably the mean and skewness where significant bias could be present. An unbiased mean is of particular importance for our application area, namely LGD modelling, because of the linear relationship between the LGD parameter and a bank's regulatory capital, as prescribed by the Basel Committee on Banking Supervision ([12]). To overcome this problem, alternative distributions that provide a better fit, specifically at the end points, can be explored. Cognisant of the importance of an unbiased mean, we entertain a slight but necessary diversion in our study to address a practical problem commonly encountered in the modelling of LGD data. This relates to the undue impact that a high concentration of values very close to zero has on the estimation of the mean. As a secondary objective to our study, we therefore investigate a data preparation approach to choose a threshold below which values may be rounded to zero. We first demonstrate that rounding greatly reduces bias in mean estimation, and thereafter demonstrate an approach to choose such a rounding threshold in an objective manner.

Having addressed this practical data preparation challenge, we propose two distributions to more accurately model the U- and J-shaped LGD data while favouring parsimony and simplicity. In addition to an improved fit, our proposed models further reduce the bias in the estimation of the mean.

The remainder of the article is organised as follows. In Section 2, we introduce our benchmark model, the zero-one inflated beta distribution, and describe the characteristics of our data, before evaluating the performance of the fitted benchmark model. We also discuss an approach to handle trivial observations close to the boundaries and propose an objective approach to select an appropriate rounding threshold. Alternative models are then proposed in Section 3, the performance of which is evaluated using traditional information criteria both for our observed sample as well as in a bootstrap simulation study for small samples. We also propose and motivate an alternative approach to measure model performance, which has specific relevance for our application to LGD data. Section 4 concludes the paper and outlines some ideas for future work.

2. Data and Benchmark Model

In this section, we introduce our benchmark model, which we will refer to as the zero-one inflated beta distribution, before briefly describing the dataset used in this study. We then demonstrate the importance of rounding very small observations to zero during the data preparation phase in order to reduce bias and propose an objective approach for choosing the rounding threshold. Section concludes with an evaluation of the fit of our benchmark model.

2.1. Zero-One Inflated Beta Distribution

For LGD data, a high concentration of observations is typically observed at the boundaries, i.e., at total recovery and total loss. Zero-one inflated models are ideally suited to handle this phenomenon, as proposed by Ospina and Ferrari [13], who made use of the beta distribution to capture the distribution in the open interval $(0,1)$. The authors motivate the use of the beta distribution based on its ability to model a variety of distributional shapes. Calabrese and Zenga [14] used Ospina and Ferrari [13]'s zero-one inflated beta (ZOIB) distribution to model the recovery rate of loans as the mixture of a Bernoulli random variable and a beta random variable.

In general, if a random variable X with density f is distributed as a zero-one inflated distribution, then

$$f(x) = wf_1(x) + (1-w)f_2(x),\ x \in [0,1], \tag{1}$$

where f_1 is a mass function, f_2 is a density function, and $w \in [0,1]$ is the mixture parameter. The mass function f_1 is that of a Bernoulli random variable, given by

$$f_1(x) = p^x(1-p)^{1-x},\ x = 0,1, \tag{2}$$

where p represents the proportion of discrete observations (i.e., 0 or 1) with a value of 1.

In what follows, we will investigate a number of different choices for the density f_2, which is used to model observations in the interval $(0,1)$. The benchmark distribution is obtained when f_2 is the beta distribution with density function

$$f_2(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1},\ x \in (0,1), \tag{3}$$

with shape parameters $\alpha > 0,\ \beta > 0$ and where Γ is the gamma function, defined by

$$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\mathrm{d}t,\ z > 0.$$

2.2. Data

The LGD data (i.e., the proportion of an individual loan account not recovered after default) used in this study to illustrate the methods proposed were obtained from an unsecured retail portfolio of a large South African bank. The period from January 2008 to April 2009 (16 months) was selected, resulting in more than 140,000 defaulted loan accounts. The choice of dates was motivated by default rates peaking over the reference period, which means that the LGD data reflect downturn conditions. The Basel Capital Accord mandates the use of LGD parameters that "reflect economic downturn conditions" for the determination of regulatory capital, as can be seen in [15]. Figure 1 below shows the distribution of LGD over the observation period, with the light blue bars representing a boxplot of data for each month. Due to confidentiality requirements, no actual values are shown. All figures and numerical results were developed in MATLAB version R2020a.

The solid dark blue line and lighter blue dashed line in Figure 1 were fitted through the mean and median LGD for each month, respectively, and indicate that LGD remained fairly stable over the reference period. The considerable distance between the plotted median and mean values is indicative of a left-skewed distribution, explained by a high concentration of observations at and close to 1. This is also evident from the proximity of the 75th percentile to 1.

The high concentration of values at 1, suggested by Figure 1, is confirmed in Figure 2, where the proportions of LGD observations at 1, 0, and on the interval (0, 1) are shown.

For illustrative purposes, the observations on (0, 1) are separated into observations in the (0, 0.01) and [0.01, 1) intervals, as indicated by the diamond markers within the sky blue bar. As can be observed, the proportion of observations between 0 and 0.01 exceeds that at 0 in each month considered.

In addition to having high concentrations of observations at the boundaries, it is common for proportional data to also exhibit high concentrations of values close to the boundary values. The substantial proportion of observations close to 0 and 1 in LGD data can in most cases be attributed to the effects of rounding during the collection and accounting processes. Specifically, accounting practices often result in trivial loss amounts being recorded for fully settled accounts.

In practice, values close to 0 and 1 are normally rounded during the data preparation stage, and the thresholds below or above which values are rounded are mostly chosen in an arbitrary manner. In the next section, we demonstrate the importance of rounding and show how a rounding threshold can be chosen in a more objective manner.

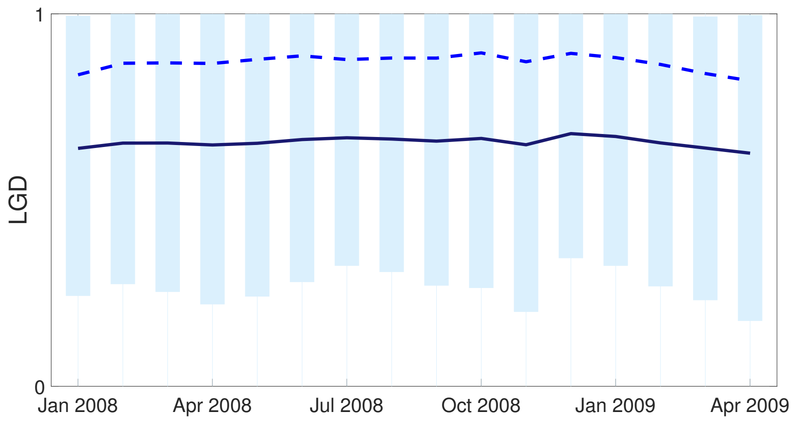

Figure 1. Distribution of LGD data with the mean (solid line). median (dashed line) and boxplot of data for each month over the observation period shown.

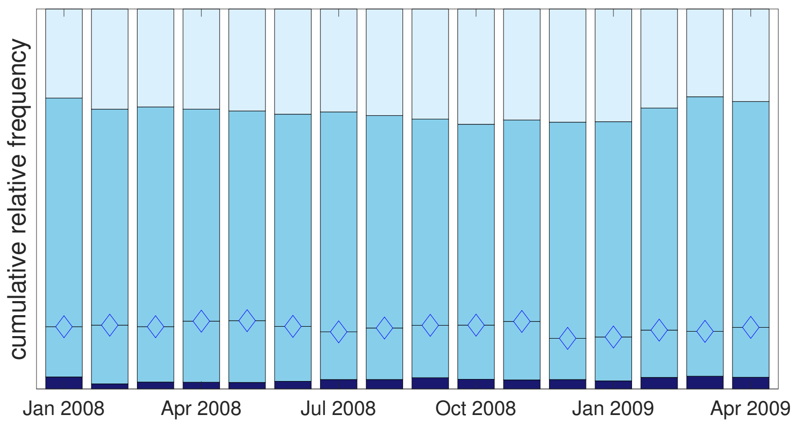

Figure 2. Proportion of LGD observations for each month at 1 (dark blue bar), 0 (light blue bar), and on the intervals $(0, 0.01)$ and $[0.01, 1)$ (medium blue bar with diamonds at 0.01).

2.3. Selection of a Rounding Threshold

The benchmark ZOIB model defined in (1) to (3) was fitted to the LGD data using maximum likelihood estimation by applying MATLAB's "betafit" function. The parameter estimates $\hat{w}$, $\hat{p}$, $\hat{\alpha}$, and $\hat{\beta}$ of the parameters w, p, α, and β are shown in Table 1 (recall that α and β is the shape parameters of the beta distribution, p is the probability of success in the Bernoulli distribution, and w is the mixing parameter). The stability of parameter estimates from month to month confirms our initial observation of homogeneity between the months within our reference period. The values of $\hat{p}$ close to one emphasise the low concentration of observations at 0 compared to those at 1.

Figure 3 shows the kernel density of the raw LGD data on (0,1) compared to the fitted beta density, after observations at 0 and 1 were removed.

It is evident how a high concentration of observations very close to zero leads to an overestimation of the density in the left tail, with an opposite effect in the right tail. This introduces bias into the estimation of moments, including higher-order moments. This is further shown in Figure 4, where a significant difference between the fitted skewness and sample skewness can be observed.

Table 1. Parameter estimates of benchmark model.

	Bernoulli		Beta	
Month	$\hat{w}$	$\hat{p}$	$\hat{\alpha}$	$\hat{\beta}$
200801	0.266	0.881	0.303	0.653
200802	0.277	0.952	0.27	0.658
200803	0.276	0.934	0.277	0.646
200804	0.281	0.938	0.26	0.66
200805	0.286	0.939	0.255	0.643
200806	0.297	0.932	0.271	0.63
200807	0.296	0.916	0.298	0.626
200808	0.305	0.92	0.281	0.636
200809	0.319	0.908	0.277	0.655
200810	0.329	0.922	0.266	0.65
200811	0.316	0.925	0.256	0.682
200812	0.323	0.922	0.317	0.634
200901	0.318	0.933	0.311	0.672
200902	0.291	0.896	0.302	0.643
200903	0.265	0.872	0.319	0.651
200904	0.274	0.889	0.297	0.698

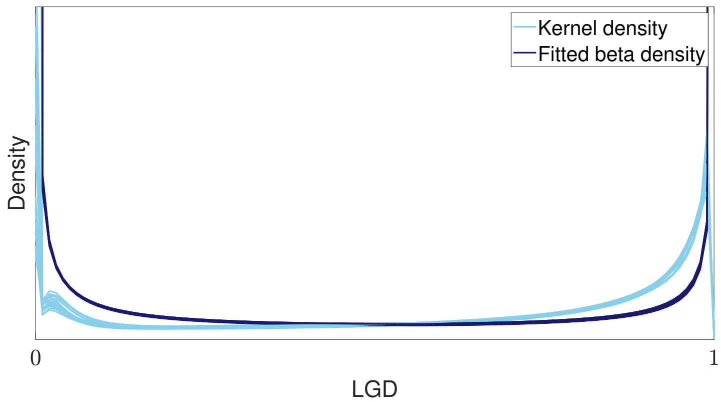

Figure 3. The kernel density of the raw LGD data on (0, 1) compared to the fitted beta density for each month from January 2008 to April 2009.

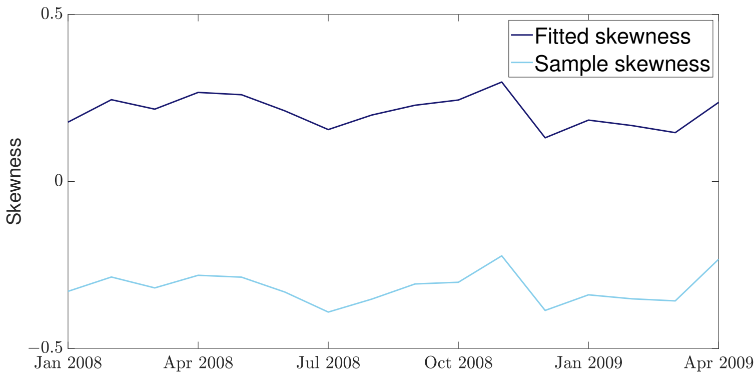

Figure 4. Comparison of sample skewness for each month with that of the fitted beta distribution in the interval $(0, 1)$.

Regardless of the model fitted, the rounding of very small LGD values to 0 during the data preparation phase is imperative to reduce the bias caused by these observations. As mentioned earlier, an unbiased mean is of specific importance in the modelling of LGD, because there exists a direct, linear relationship between the parameter and the regulatory capital that a bank needs to hold. In practice, the threshold below which observations are rounded to zero is mostly chosen in an arbitrary manner. Below, we outline how the rounding threshold can be selected in a more objective manner.

Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the ZOIB model. The estimated mean of the fitted model is given by

$$\hat{\mu}_X = \hat{w}\hat{p} + (1 - \hat{w})\frac{\hat{\alpha}}{\hat{\alpha} + \hat{\beta}},$$

where $\hat{w}$, $\hat{p}$, $\hat{\alpha}$, and $\hat{\beta}$ are the maximum likelihood estimators (MLEs) of the parameters w, p, α and β, respectively.

The mean $\hat{\mu}_X$ will, in general, not be equal to the sample mean $\bar{X}$ and hence an empirical bias, defined by

$$e_X = \hat{\mu}_X - \bar{X}, \tag{4}$$

exists.

Figure 5 shows the calculated empirical bias for the LGD data in the interval $(0, 1)$ as well as the resulting empirical bias if the LGD data in the interval $(0, \epsilon)$ are rounded to zero for chosen values of $\epsilon = \{0.0001, 0.001, 0.01\}$.

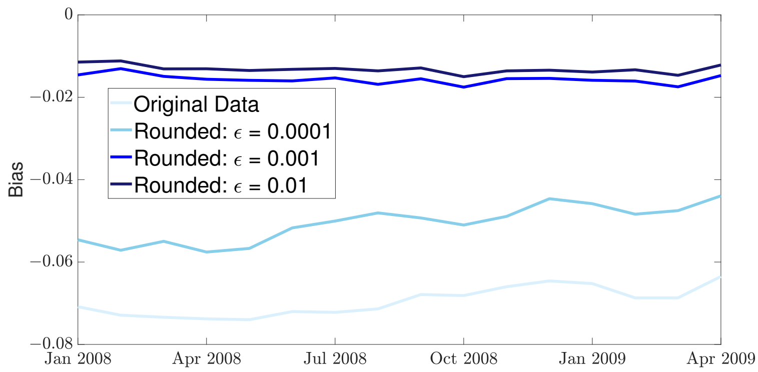

Figure 5. The calculated empirical bias compared to the empirical bias after rounding for chosen values of ϵ.

From Figure 5, it is clear that the empirical bias can be significantly reduced by increasing the rounding threshold ϵ. However, this comes at the risk of changing the properties of the observed sample. Therefore, we aimed to choose the value of ϵ to make the bias as small as possible, while still retaining the properties of the observed sample as far as possible.

The rounding of small observations to 0 is given by

$$Y_{i,\epsilon} = X_i \mathbb{1}_{\{X_i \geq \epsilon\}},$$

where $\mathbb{1}_{\{X_i \geq \epsilon\}}$ is an indicator function taking on the value 1 when $X_i \geq \epsilon$ and 0 otherwise.

This then leads to the following decomposition of the empirical bias in (4):

$$e_X = (\hat{\mu}_X - \hat{\mu}_{Y,\epsilon}) + e_{Y,\epsilon} + (\bar{Y}_\epsilon - \bar{X}),$$

where $\hat{\mu}_{Y,\epsilon}$ and $\bar{Y}_\epsilon$ represent the mean of the fitted model and the sample mean, respectively, both calculated on the data after observations below ϵ were rounded to 0. The resulting empirical bias is then denoted by $e_{Y,\epsilon}$.

Rewriting this, we have that

$$\hat{\mu}_{Y,\epsilon} - \hat{\mu}_X = (e_{Y,\epsilon} - e_X) - (\bar{X} - \bar{Y}_\epsilon).$$

Therefore, the change in the fitted mean can be written as the difference between the increased accuracy obtained by rounding and the cost of rounding bias given by $\bar{X} - \bar{Y}_\epsilon \geq 0$.

The reduction in empirical bias obtained by rounding is relatively large for small values of ϵ, however, as can be seen from Figure 5, it seems to decelerate as ϵ increases. In contrast, as ϵ increases, the rounding bias will increase.

To find a value of ϵ in an objective manner, we choose ϵ as

$$\epsilon^* = \inf\{\epsilon > 0 : \left|\frac{\mathrm{d}e_{Y,\epsilon}}{\mathrm{d}\epsilon}\right| \leq 1\}, \tag{5}$$

where $\frac{\mathrm{d}e_{Y,\epsilon}}{\mathrm{d}\epsilon}$ represents the derivative of $e_{Y,\epsilon}$ with respect to ϵ, noting that $\mathrm{d}(e_{Y,\epsilon} - e_X) = \mathrm{d}e_{Y,\epsilon}$. The threshold ϵ^* can be interpreted as a conservative estimate of the rounding bias, since from the definition of Y_i, we have that

$$0 \leq \bar{X} - \bar{Y}_\epsilon = \frac{1}{n}\sum_{i=1}^{n} X_i \mathbb{1}_{\{X_i<\epsilon\}} < \epsilon.$$

Therefore, ϵ^* in (5) is the smallest value of ϵ where the rate of improvement in empirical bias exceeds the rate of increase in rounding bias.

We implement this method by plotting $\frac{1}{\Delta}(e_{Y,\epsilon+\Delta} - e_{Y,\epsilon})$ against a grid of values of ϵ with increment size Δ. The smallest value of ϵ where this function is not greater than one is our objective choice of ϵ^*. Since we implement this for several months, we take the average across all the months of these functions for each value of ϵ. The result is shown in Figure 6, where the value of $\epsilon^* = 0.001$ was chosen for our data.

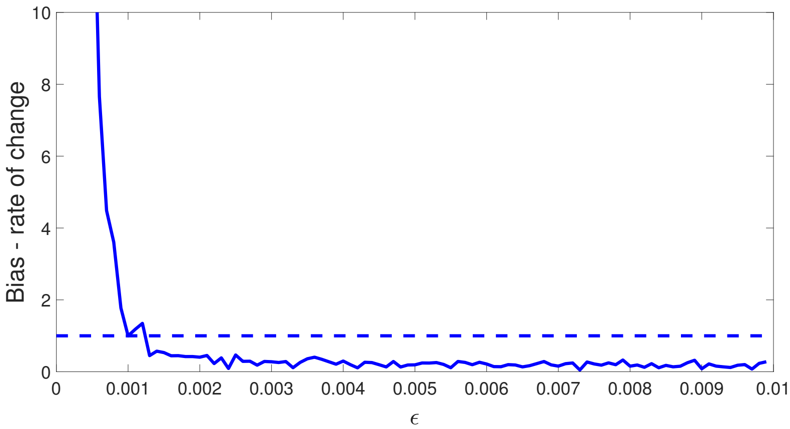

Figure 6. The rate of change in the rounding bias $\frac{1}{\Delta}(e_{Y,\epsilon+\Delta} - e_{Y,\epsilon})$ plotted against the rounding threshold ϵ.

2.4. Fitting the Benchmark Model

Having rounded observations between 0 and ϵ^* (chosen as $\epsilon^* = 0.001$ for the given data), as per the approach described in the previous section, we set out to evaluate the fit of the benchmark model to our data.

In the second and third columns of Table 2, the parameter estimates for the fitted ZOIB model described in Section 2.1 are shown.

The stability of parameter estimates from month to month confirms our initial observation of homogeneity between months within our reference period. The effect of the rounding procedure (as described in Section 2.3) is evident in the estimates of $\hat{p}$, where a substantial increase in the proportion of 0 s compared to Figure 1 can be observed.

We also note that the estimates for $\hat{\alpha}$ are relatively close to one for most months under observation. It therefore seems worthwhile to investigate whether the assumption that $\alpha = 1$ can be supported, as fixing the α parameter at 1 reduces the number of parameters associated with our benchmark model from 3 to 2. This is important considering our desire for parsimony.

Table 2. Parameter estimates with estimated standard errors in parenthesis of the ZOIB and ZOISP models on rounded LGD data.

	Bernoulli		**Beta**		**Standard Power**
Month	$\hat{w}$	$\hat{p}$	$\hat{\alpha}$	$\hat{\beta}$	$\tilde{\beta}$
200801	0.393	0.596	0.94 (0.019)	0.525 (0.009)	0.54 (0.008)
200802	0.428	0.616	1.003 (0.02)	0.521 (0.009)	0.52 (0.007)
200803	0.418	0.617	0.962 (0.02)	0.51 (0.009)	0.519 (0.008)
200804	0.438	0.602	0.972 (0.021)	0.509 (0.009)	0.516 (0.008)
200805	0.444	0.605	0.966 (0.021)	0.493 (0.009)	0.501 (0.008)
200806	0.437	0.634	0.931 (0.019)	0.488 (0.009)	0.505 (0.018)
200807	0.417	0.65	0.959 (0.018)	0.508 (0.008)	0.518 (0.007)
200808	0.435	0.645	0.916 (0.018)	0.496 (0.008)	0.518 (0.007)
200809	0.452	0.641	0.918 (0.019)	0.506 (0.009)	0.527 (0.008)
200810	0.466	0.651	0.891 (0.018)	0.489 (0.009)	0.517 (0.008)
200811	0.466	0.627	0.887 (0.019)	0.503 (0.009)	0.532 (0.008)
200812	0.427	0.698	0.891 (0.018)	0.508 (0.009)	0.537 (0.008)
200901	0.429	0.692	0.897 (0.016)	0.533 (0.009)	0.562 (0.008)
200902	0.41	0.636	0.898 (0.016)	0.507 (0.007)	0.534 (0.007)
200903	0.377	0.613	0.867 (0.014)	0.514 (0.008)	0.552 (0.006)
200904	0.4	0.609	0.832 (0.014)	0.527 (0.008)	0.578 (0.007)

Setting $\alpha = 1$ in our benchmark model, we have the continuous part represented by the $Beta(1, \beta)$ distribution. If $X \sim Beta(1, \beta)$, then $1 - X \sim Beta(\beta, 1)$ with density

$$f_2(x) = \beta(1 - x)^{\beta - 1},\ x \in (0, 1). \tag{6}$$

This distribution is not only a special case of the beta distribution, but also the Kumaraswamy distribution introduced by [16] based on [17] and extended to include zeros and ones in [18]. This special case is known as the standard power (SP) distribution [19].

Suppose that $1 - X_1, 1 - X_2, \ldots, 1 - X_n$ is an i.i.d. sample from the SP distribution, and then the MLE of β is given by

$$\tilde{\beta} = \frac{-n}{\sum_{i=1}^{n} \log(1 - X_i)}.$$

The estimates for $\tilde{\beta}$ for the SP distribution fitted on $(0, 1)$ are shown in the last column of Table 2. Since $\hat{\alpha}$ is relatively close to one for all months, the difference between the estimators $\tilde{\beta}$ and $\hat{\beta}$ is small. In Figure 7, the fitted means of the beta and standard power distributions are shown. It is evident that the impact of these differences in estimates can be meaningful. Shown also in Figure 7 is the sample mean for observations between 0 and 1, which indicates that both fitted models underestimate the mean, with the standard power distribution to a lesser degree. The scale of the vertical axis is 0.1 with the actual LGD values not shown due to the confidential nature of the data.

The standard errors of the estimators $\hat{\alpha}$, $\hat{\beta}$ and $\tilde{\alpha}$ for the parameters of the beta and standard power distributions are also shown in Table 2. These standard errors were estimated by using a parametric bootstrap method, in which random variates were simulated from the fitted beta and standard power models. Beta random variates were generated using Matlab's built-in "betarnd" function, while random variates from the standard power distribution in (6) were generated using its inverse cumulative distribution function (cdf), given by

$$F^{-1}(p) = 1 - (p - 1)^{\frac{1}{\beta}},\ 0 \leq p \leq 1. \tag{7}$$

The calculated standard errors show that the SP model can be considered for several months, since $\alpha = 1$ is within the 95% confidence interval for α for each month from February 2008 until May 2008. The standard errors of $\hat{w}$, the estimator for the mixture parameter, and $\hat{p}$, the estimator for the Bernoulli parameter, are not reported in the tables as the focus of the study is to compare different models with density f_2 (as can be seen in (1)).

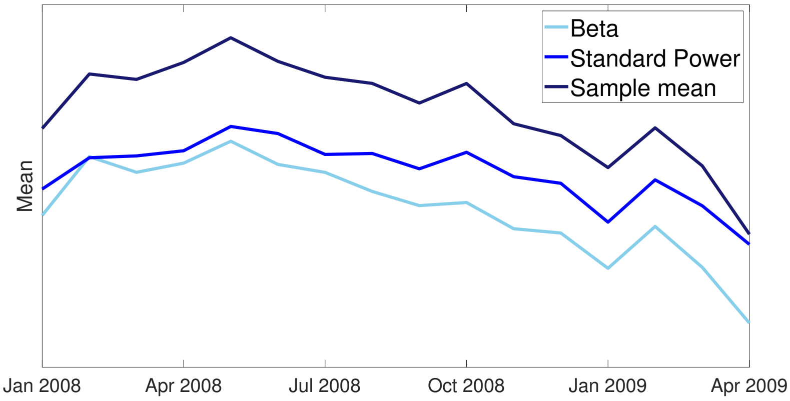

Figure 7. The fitted mean of the beta and standard power distributions excluding observations between 0 and 1. The scale on the vertical axis is 0.1.

To determine whether it is feasible to simplify the continuous part of our benchmark model to the SP distribution, we make use of two frequently used selection metrics. These are the Akaike information criteria (AIC) and Bayesian information criteria (BIC). Information criteria such as the AIC and BIC are specifically appropriate when evaluating model fit when parsimony is the main objective, since these measures incorporate a penalty for additional model parameters.

For any fitted model, the AIC can be expressed as

$$\text{AIC} = -2\ell(\hat{\theta_k} \mid X_1, X_2, \ldots, X_n) + 2k, \tag{8}$$

while similarly, the BIC is given by

$$\text{BIC} = -2\ell(\hat{\theta_k} \mid X_1, X_2, \ldots, X_n) + k\log(n), \tag{9}$$

where $\hat{\theta}_k$ is the vector of MLEs, $X_1, X_2, \ldots, X_n$ are random variables from a distribution F with a k-dimensional vector of parameters θ_k, and $\ell(\theta_k \mid \cdot)$ is the log likelihood function.

For the fitted ZOIB model, we have that $\hat{\theta}_k = (\hat{w}, \hat{p}, \hat{\alpha}, \hat{\beta})$ with $k = 4$, while for the fitted zero-one inflated standard power (ZOISP) distribution model, $\hat{\theta}_k = (\hat{w}, \hat{p}, \tilde{\beta})$ with $k = 3$.

Table 3 shows the AIC and BIC for the ZOIB and ZOISP models fitted to observations between 0 and 1. It is evident that the ZOISP distribution rarely provides a better fit than the ZOIB distribution, even when using the BIC as a metric. Given that the shape of the SP distribution is strictly increasing and unimodal, we suspect that the ZOIB distribution outperforms the ZOISP distribution for most months due to the beta distribution's ability to model bimodality. However, the beta distribution underestimates the masses close to zero on the interval $(0, 1)$ (after rounding), as can be observed from the QQ-plot in Figure 8.

In Figure 8, the empirical quantiles of the February 2008 data between 0 and 1 are plotted against the quantiles of the fitted beta distribution. This month has been chosen because the fitted beta distribution is close to the fitted standard power distribution with estimated parameter 0.52. The notable distance between the two lines in the lower quadrant demonstrates the underestimation by the beta distribution of the masses close to zero. This underestimation is likely to contribute to the difference between the fitted and sample means, as can be observed from the remaining empirical bias (after rounding) in Figure 5. In the next section, we set out to find an alternative distribution that is able to capture the bimodality of our data while simultaneously looking to reduce the difference between the fitted and sample means.

Table 3. Information criteria for the fitted ZOIB and ZOISP distributions.

	ZOIB		ZOISP	
Month	**AIC**	**BIC**	**AIC**	**BIC**
200801	13240	13268	13248 (0.06%)	13269 (0.01%)
200802	13418	13446	13416 (−0.01%)	13437 (−0.07%)
200803	12535	12563	12537 (0.02%)	12558 (−0.04%)
200804	12852	12880	12852 (0%)	12873 (−0.05%)
200805	13174	13202	13174 (0%)	13195 (−0.05%)
200806	12263	12291	12272 (0.07%)	12293 (0.02%)
200807	13817	13846	13820 (0.02%)	13841 (−0.04%)
200808	13900	13929	13918 (0.13%)	13939 (0.07%)
200809	14171	14199	14186 (0.11%)	14207 (0.06%)
200810	14699	14727	14730 (0.21%)	14751 (0.16%)
200811	15578	15606	15613 (0.22%)	15634 (0.18%)
200812	13102	13130	13133 (0.24%)	13154 (0.18%)
200901	15142	15170	15172 (0.2%)	15194 (0.16%)
200902	18403	18432	18443 (0.22%)	18465 (0.18%)
200903	17309	17338	17384 (0.43%)	17406 (0.39%)
200904	18401	18431	18526 (0.68%)	18548 (0.63%)

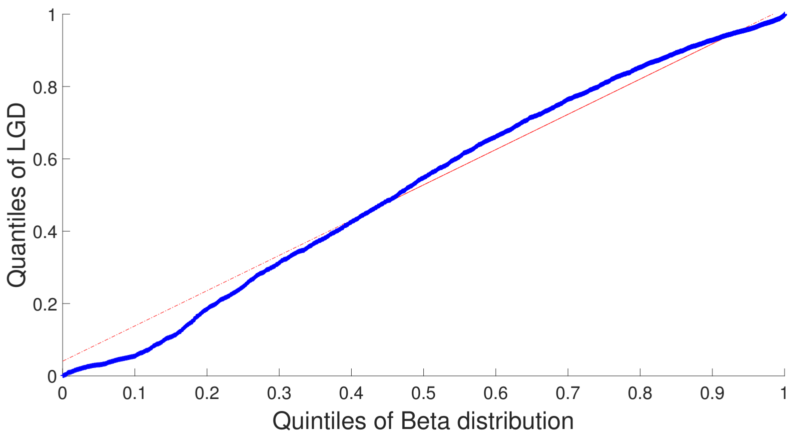

Figure 8. QQ-plot of February 2008 LGD data against fitted beta distribution.

3. New Proposed Models

Several authors have observed that LGD data exhibit bi- or multi-modality even when the discrete endpoints are removed. This can also be observed in our data, as illustrated by the bimodal kernel density curve in Figure 3. The ZOIB distribution is generally well-suited to deal with the excess zeros and ones in most LGD datasets, however, it is unable to accurately model bimodality in the continuous portion of the empirical distribution. In fact, Ref. [14] shows that when a large dataset of Italian bank loan recovery rates is considered, the beta function cannot adequately describe the estimated density function on $(0, 1)$.

To address bi- and multimodality in LGD data, Ref. [20] proposed an extension of [13]'s ZOIB model by formulating a mixture model of two beta probability distributions, inflated at zero and one. Ref. [21] expanded on this by fitting a number of mixture distributions to account for the multimodality on the unit interval, concluding that the inflated mixture of beta distributions provides the optimal representation, especially when considering the need for conservatism (i.e., not underestimating the mean). However, Ref. [22] demonstrated that beta mixtures "do more harm than they help" for bimodal U-shaped data. This, and the fact that mixing two beta distributions requires three additional parameters (when compared to our benchmark model), creates the opportunity to explore alternative mixture distributions for our non-binary data.

We address the matter of parsimony by proposing a mix of SP distributions to model the data on $(0, 1)$. Our aim was to cut down on the number of parameters (compared to [20]'s inflated mixture model), but attain the adequate characterisation of the bimodal nature of the empirical distribution. Our secondary but equally important objective relates to the alignment of the empirical and distributional means.

3.1. Inflated Mixed Standard Power Distribution

Suppose that X has density

$$f_2(x) = \pi a x^{a-1} + (1-\pi) b (1-x)^{b-1}, x \in (0,1), \tag{10}$$

with shape parameters $a, b > 0$ and mixture weight $\pi \in [0,1]$.

This distribution is a mixture of two SP distributions and will be called the mixture of standard power (MSP) distribution. The mixture of the MSP distribution and the Bernoulli distribution is henceforth referred to as the zero one inflated mixture of the standard power (ZOIMSP) model.

Even though the MSP distribution is a mixture of two distributions, maximum likelihood estimators of this model exist. This is due to a log likelihood function that is bounded from above, in contrast with other mixture models, such as a mixture of two normal distributions (see Section 9.2.1 of [23]).

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from the MSP distribution. The log likelihood function is given by

$$\ell(a, b, \pi \mid X_1, X_2, ..., X_n) = \sum_{i=1}^{n} \log\left(\pi a X_i^{a-1} + (1-\pi) b (1-X_i)^{b-1}\right). \tag{11}$$

Maximum likelihood estimators, denoted by $\hat{a}$, $\hat{b}$, and $\hat{\pi}$, can then be calculated by solving the following set of equations for a, b, and π:

$$\begin{aligned}
\frac{\partial \ell}{\partial a} &= \sum_{i=1}^{n} \frac{\pi X_i^{a-1}(1 + a \log X_i)}{\pi a X_i^{a-1} + (1-\pi) b (1-X_i)^{b-1}} = 0; \\
\frac{\partial \ell}{\partial b} &= \sum_{i=1}^{n} \frac{(1-\pi)(1-X_i)^{b-1}(1 + b \log(1-X_i))}{\pi a X_i^{a-1} + (1-\pi) b (1-X_i)^{b-1}} = 0; \\
\frac{\partial \ell}{\partial \pi} &= \sum_{i=1}^{n} \frac{a X_i^{a-1} - b(1-X_i)^{b-1}}{\pi a X_i^{a-1} + (1-\pi) b (1-X_i)^{b-1}} = 0.
\end{aligned}$$

To obtain these estimates, MATLAB's "fsolve" function was used, which numerically solves a system of equations by a trust-region-dogleg algorithm. Using the LGD data described in the previous section, the estimated values for this model were calculated and are shown in the first three columns of Table 4. It is interesting to note that $\hat{b}$ is close to one for several months, suggesting that the LGD distribution for these months may be unimodal. By setting $b = 1$ in (10), we can obtain a model with the same number of parameters as in our benchmark model. This restricted model is in effect a mixture of a uniform distribution and a standard power (MUSP) distribution and its log likelihood function can be obtained by setting $b = 1$ in (11).

Estimators $\hat{a}$ and $\hat{\pi}$ can then be obtained by solving the following two equations for a and b.

$$\begin{aligned}
\frac{\partial \ell}{\partial a} &= \sum_{i=1}^{n} \frac{\pi X_i^{a-1}(1 + a \log X_i)}{\pi a X_i^{a-1} + (1-\pi)} = 0; \\
\frac{\partial \ell}{\partial \pi} &= \sum_{i=1}^{n} \frac{a X_i^{a-1} - 1}{\pi a X_i^{a-1} + (1-\pi)} = 0.
\end{aligned}$$

The estimates obtained for LGD data using the MUSP model are shown in the final two columns of Table 4 and it is notable that these exhibit more stability over different months than those of the MSP model. In this regard, both the SP and beta models have better stability in terms of their parameter estimates over the reference period. However, this stability comes at the cost of assigning too little weight to the masses at the lower mode. The MSP's flexibility and ability to more accurately capture the lower mode may therefore make it the preferred model for bimodal U- or J-shaped data.

Table 4. Parameter estimates with estimated standard errors in parenthesis for MSP and MUSP models.

	MSP			MUSP	
Month	$\hat{a}$	$\hat{b}$	$\hat{\pi}$	$\hat{a}$	$\hat{\pi}$
200801	14.727 (0.722)	0.892 (0.05)	0.338 (0.02)	13.618 (0.518)	0.378 (0.007)
200802	14.945 (0.753)	0.798 (0.04)	0.324 (0.02)	13.047 (0.527)	0.407 (0.007)
200803	12.81 (0.601)	0.879 (0.057)	0.382 (0.023)	11.983 (0.463)	0.428 (0.008)
200804	12.609 (0.615)	0.82 (0.051)	0.367 (0.024)	11.575 (0.437)	0.438 (0.008)
200805	13.932 (0.676)	0.769 (0.041)	0.352 (0.021)	12.447 (0.476)	0.447 (0.008)
200806	13.755 (0.717)	0.739 (0.038)	0.327 (0.021)	12.329 (0.473)	0.438 (0.008)
200807	14.305 (0.726)	0.768 (0.036)	0.315 (0.019)	12.466 (0.463)	0.415 (0.007)
200808	14.586 (0.747)	0.761 (0.036)	0.308 (0.02)	12.522 (0.457)	0.414 (0.007)
200809	13.918 (0.713)	0.816 (0.046)	0.329 (0.022)	12.318 (0.492)	0.406 (0.008)
200810	11.369 (0.561)	1.041 (0.076)	0.451 (0.024)	11.522 (0.427)	0.439 (0.008)
200811	12.026 (0.581)	0.974 (0.066)	0.397 (0.023)	11.827 (0.452)	0.407 (0.008)
200812	9.182 (0.439)	1.411 (0.107)	0.515 (0.02)	11.651 (0.48)	0.403 (0.008)
200901	7.001 (0.415)	1.873 (0.329)	0.584 (0.063)	10.77 (0.428)	0.391 (0.008)
200902	13.489 (0.551)	0.911 (0.046)	0.36 (0.019)	12.641 (0.402)	0.394 (0.006)
200903	9.698 (0.392)	1.375 (0.078)	0.488 (0.015)	12.241 (0.421)	0.383 (0.006)
200904	8.164 (0.32)	1.708 (0.087)	0.515 (0.012)	12.455 (0.471)	0.344 (0.007)

The standard errors shown in Table 4 were again estimated using a parametric bootstrap approach. To simulate random variables from a mixture of two distributions, πn random variates were simulated from the one distribution and $(1-\pi)n$ from the other distribution. For the MSP model, random variates were simulated from two standard power distributions by making use of the inverse cumulative distribution function (cdf) of the two distributions, given by

$$F^{-1}(p) = p^{\frac{1}{\alpha}},\ 0 \leq p \leq 1, \tag{12}$$

and (7), respectively. For the MUSP model, the inverse cdf in (12) was used to generate πn random variates from the standard power distribution added to the $(1-\pi)n$ random variates from the standard uniform distribution.

The instability of the parameter estimates in the MSP model are emphasized by analysing the standard errors shown in Table 4. This is especially the case for the data relating to January 2009. An in-depth analysis of the bootstrap sample distribution of $\hat{b}$ for this specific month shows a distribution with two peaks, one of which is close to the estimated value of 1.873, while the other peak is at a value inferior to 1. This may be due to one of many possible factors that warrants further investigation but is beyond the scope of this study. Two possible factors are the numerical optimisation method used to find the estimates, or it can be due to the identifiability of these mixture models. Although the identifiability of the MSP model is not further investigated, another factor that may lead to the instability of the calculated estimates is considered.

In Figure 9, the different shapes that the MSP model allows are shown. In this figure, the kernel densities as well as the fitted densities for the LGD data of February 2008–March 2009 are plotted. The densities of the two fitted standard power distributions are plotted on the right. The J-shaped kernel densities were plotted using MATLAB's "ksdensity" function with a reflection boundary correction method applied. Even though the two fitted densities are very close to each other, the density for February 2008 is monotonically increasing, while the density for September 2009 is J-shaped. This corresponds to values of b in (10) that are less than 1 or greater than 1. Therefore, the difficulty in modelling J-shaped data that are close to unimodal may lead to estimators of b that are inconsistent, and consequently to the inconsistency of the estimators of a and π.

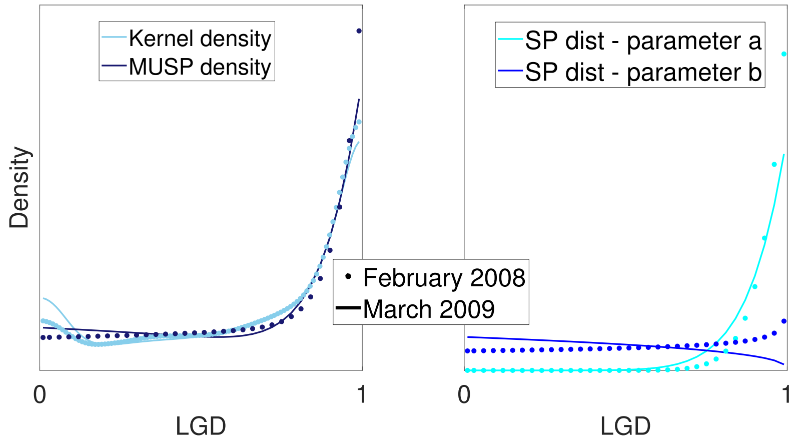

Figure 9. Densities for February 2008 and March 2009 LGD data. On the left, the kernel density is compared to the fitted MSP model, while on the right, the two contributing SP fitted densities of the MSP model are shown, driven by $\hat{a}$ and $\hat{b}$, respectively.

3.2. Model Performance Based on Information Criteria

When comparing the information criteria for the ZOIMSP and ZOIMUSP models (shown in the last four columns of Table 5), there is little to choose between the two models based on the AIC and BIC. This is also evident from Figure 10 below.

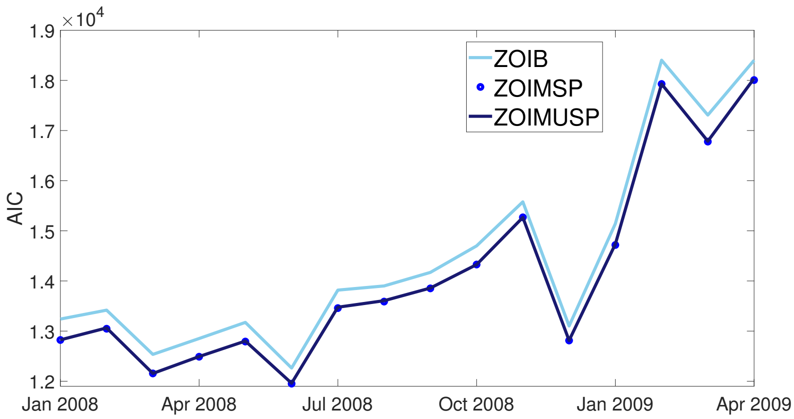

Figure 10. The AIC of the ZOIB as well as ZOIMSP and ZOIMUSP models, plotted for each month.

It is interesting to note that the ZOIMUSP outperforms the ZOIMSP for several months, suggesting that the lower mode is of less significance for the data relating to these months. This corroborates the earlier conclusion that a unimodal distribution could provide an adequate characterisation of the data in certain cases. It is noteworthy to observe that both the ZOIMSP and ZOIMUSP markedly outperform the benchmark ZOIB model, regardless of information criteria utilised. From Figure 10, it is also evident that the improvement in AIC is of similar magnitude for all months during the observation period. The results for the BIC are similar.

Table 5. Information criteria for the fitted ZOIMSP and ZOIMUSP models compared to the ZOIB model.

	ZOIB		ZOIMSP		ZOIMUSP	
Month	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**
200801	13240	13268	12822 (−3.2%)	12858 (−3.1%)	12825 (−3.1%)	12853 (-3.1%)
200802	13418	13446	13047 (−2.8%)	13082 (−2.7%)	13063 (−2.6%)	13091 (−2.6%)
200803	12535	12563	12156 (−3%)	12190 (−3%)	12155 (−3%)	12183 (−3%)
200804	12852	12880	12489 (−2.8%)	12524 (−2.8%)	12492 (−2.8%)	12520 (−2.8%)
200805	13174	13202	12791 (−2.9%)	12826 (−2.8%)	12801 (−2.8%)	12829 (−2.8%)
200806	12263	12291	11952 (−2.5%)	11987 (−2.5%)	11961 (−2.5%)	11988 (−2.5%)
200807	13817	13846	13458 (−2.6%)	13494 (−2.5%)	13476 (−2.5%)	13505 (−2.5%)
200808	13900	13929	13588 (−2.2%)	13624 (−2.2%)	13606 (−2.1%)	13634 (−2.1%)
200809	14171	14199	13852 (−2.2%)	13887 (−2.2%)	13858 (−2.2%)	13887 (−2.2%)
200810	14699	14727	14328 (−2.5%)	14363 (−2.5%)	14326 (−2.5%)	14355 (−2.5%)
200811	15578	15606	15269 (−2%)	15304 (−1.9%)	15267 (−2%)	15295 (−2%)
200812	13102	13130	12812 (−2.2%)	12847 (−2.2%)	12813 (−2.2%)	12841 (−2.2%)
200901	15142	15170	14717 (−2.8%)	14753 (−2.8%)	14735 (−2.7%)	14763 (−2.7%)
200902	18403	18432	17928 (−2.6%)	17965 (−2.5%)	17929 (−2.6%)	17958 (−2.6%)
200903	17309	17338	16779 (−3.1%)	16816 (−3%)	16787 (−3%)	16817 (−3%)
200904	18401	18431	18007 (−2.1%)	18043 (−2.1%)	18039 (−2%)	18068 (−2%)

3.3. Model Performance Based on Empirical Bias

Having evaluated the performance of the ZOIMSP and ZOIMUSP against the benchmark ZOIB model, we now set out to assess the relative alignment between the empirical and distributional means of the two models, both inflated at zero and one to account for discrete observations.

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from the ZOIMSP model. The fitted mean is given by

$$\hat{\mu}_X = \hat{w}\hat{p} + (1-\hat{w})\left(\hat{\pi}\frac{\hat{a}}{\hat{a}+1} + (1-\hat{\pi})\frac{1}{1+\hat{b}}\right),$$

where $\hat{w}$, $\hat{p}$, $\hat{b}$, $\hat{b}$, and $\hat{\pi}$ are the maximum likelihood estimators (MLEs) of the parameters w, p, a, b, and π while for the ZOIMUSP model the value of $\hat{b} = 1$.

In Figure 11, the empirical bias, as defined in (4), is shown with respect to each month in the reference period for the ZOIMSP, ZOIMUSP, as well as the benchmark ZOIB model. A significant reduction in bias can be observed for both the ZOIMSP and ZOIMUSP when compared to the benchmark model, with the ZOIMSP exhibiting better stability than ZOIMUSP. This can most likely be attributed to the ZOIMSP providing a better fit to the "hump" close to 0, evident from the plotted kernel density in Figure 3.

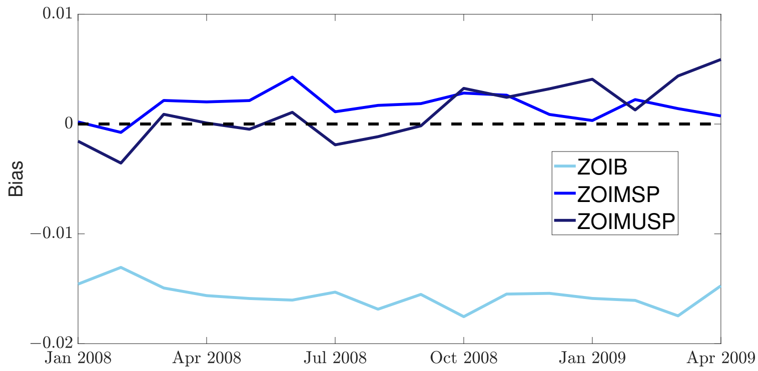

Figure 11. The empirical bias of the fitted ZOIMSP, ZOIMUSP, and benchmark ZOIB models plotted for each month.

The ZOIMSP exhibits more consistent bias-reduction than the ZOIMUSP. This is an important model attribute with regard to LGD data, where specific importance is attached to the mean of the fitted distribution. Besides the accuracy of the distributional mean, conservatism is desirable due to the relationship between this parameter and a bank's regulatory capital. Regulatory capital has as a primary aim financial protection against unexpected losses, and therefore, a prudent assessment of this amount is crucial. Conservatism in our context therefore means avoiding the underestimation of the distributional mean as much as possible to ensure that the LGD parameter satisfies the prudence requirement.

As can be observed from Figure 11 the empirical bias of the ZOIMSP rarely falls below 0, whereas the empirical bias of the ZOIMUSP is negative for several months within our reference period. In the context of our application to LGD data, the ZOIMSP therefore outperforms the ZOIMUSP both in terms of empirical bias and conservatism.

3.4. Model Performance for Small Samples

We conclude the paper with a mini simulation study to evaluate the small sample performance of the newly proposed ZOIMSP and ZOIMUSP models. In our current dataset, we have more than 140,000 observations spread across 16 months (an average of more than 8750 observations per month). In general, suppose that for month j, $j = 1, \ldots, 16$, we have m_j LGD observations denoted by $x_1, x_2, \ldots, x_{m_j}$. The idea is then to obtain the estimated distributional properties of the AIC and BIC statistics for each of the models for samples of size $n = 100, 200$, and 500. We view the m_j observations for month j as the population of values and apply the following algorithm to obtain the samples from this "population".

1. For month j, obtain a sample $X_1^*, X_2^*, \ldots, X_n^*$ by sampling $n < m_j$ observations with a replacement from $x_1, x_2, \ldots, x_{m_j}$.
2. Fit a model on $X_1^*, X_2^*, \ldots, X_n^*$ to obtain the estimator $\hat{\theta}^* = \hat{\theta}(X_1^*, X_2^*, \ldots, X_n^*)$.
3. Calculate the statistics AIC^* and BIC^* from Equations (8) and (9) with inputs $X_1^*, X_2^*, \ldots, X_n^*, \hat{\theta}^*, n$ and the number of parameters k.
4. Repeat steps 1–3 above B times to obtain $AIC_1^*, AIC_2^*, \ldots, AIC_B^*$, and $BIC_1^*, BIC_2^*, \ldots, BIC_B^*$

This algorithm is applied on the newly proposed ZOIMSP and ZOIMUSP models as well as the benchmark ZOIB model for each month during the observation period. For large values of B, the sets $\{AIC_i^*, 1 \leq i \leq B\}$ and $\{BIC_i^*, 1 \leq i \leq B\}$ represent the sample distributions of the AIC and BIC for a specific model and statistics calculated on these sets can be used to estimate the properties of these distributions.

In Figure 12, the bootstrap estimates of the expectation of the AIC and BIC for the ZOIMSP, ZOIMUSP, and ZOIB models are compared with each other when $n = 100$ and with $B = 1000$ bootstrap replications. The shaded areas in this figure show that the bootstrap estimated 5% and 95% quantiles of the AIC and BIC for the ZOIMUSP model, which was identified as the best-performing model.

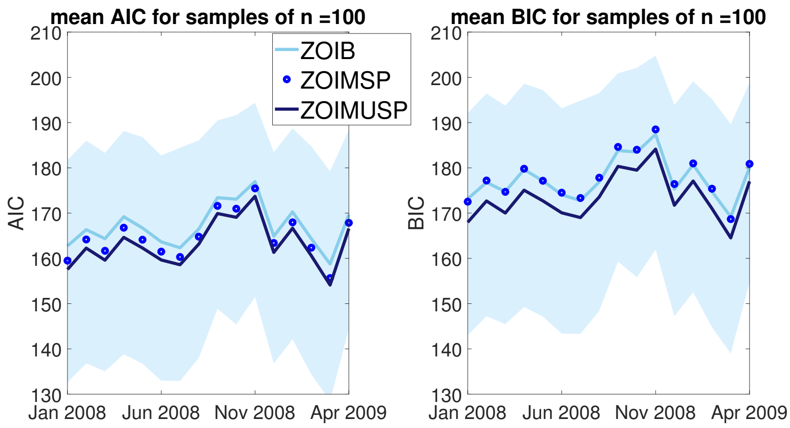

Figure 12. The bootstrap estimated mean of the AIC and BIC for ZOIB, ZOIMSP, and ZOIMUSP models for a sample size of $n = 100$. The shaded area represents a 90% confidence interval for the mean of the ZOIMUSP model.

From Figure 12, it is clear that, when $n = 100$, the ZOIMUSP model outperforms the less parsimonious ZOIMSP model with regard to both information criteria, while the benchmark ZOIB model compares reasonably well with the ZOIMSP model. Moreover, the bootstrap estimated expected that AIC and BIC for all three models are within the 90% confidence interval for the mean AIC and BIC of the ZOIMUSP model.

In Figures 13 and 14, the results are shown when the sample size increases to $n = 200$ and $n = 500$, respectively. The results are similar to those observed when $n = 100$, except that the bootstrap estimated expected AIC and BIC for the ZOIMSP model approach and the bootstrap estimated expected AIC and BIC for the ZOIMUSP model. This suggests that the degree to which the less parsimonious ZOIMSP model is penalised becomes negligible as the sample size increases.

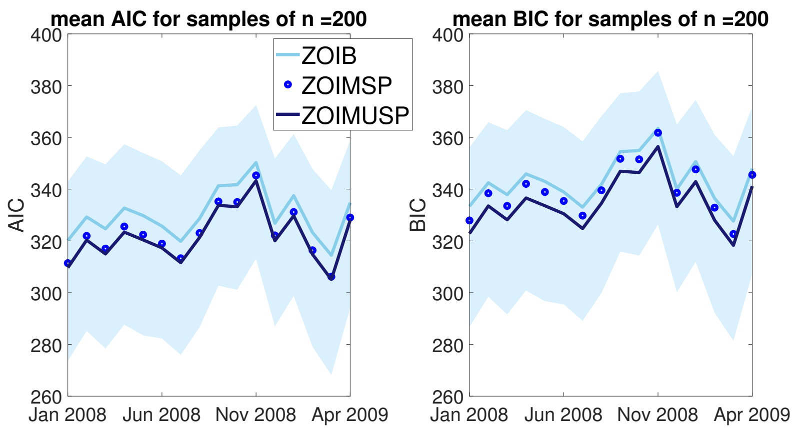

Figure 13. Bootstrap estimated mean of the AIC and BIC for the ZOIB, ZOIMSP, and ZOIMUSP models for a sample size of $n = 200$. The shaded area represents a 90% confidence interval for the mean of the ZOIMUSP model.

The results from Figures 12–14 suggest that the expected AIC and BIC of the ZOIMUSP model are less than the corresponding expectations of the ZOIB model. However, this analysis does not take the dependence between the AIC (or BIC) of the three models into account. Therefore, to analyse the relationship between the sample distributions of the AIC (or BIC), bootstrap estimates are obtained for the probabilities that the AIC (or BIC) for one model is greater than the AIC (or BIC) of another model. For example, comparing the AIC of the ZOIB and ZOIMSP models, we estimate the probability

$$P(AIC_{ZOIMSP} < AIC_{ZOIB})$$

by

$$\frac{1}{B}\sum_{i=1}^{B} \mathbb{1}(AIC^*_i{}_{ZOIMSP} < AIC^*_i{}_{ZOIB}), \tag{13}$$

where $\mathbb{1}(AIC^*_i{}_{ZOIMSP} < AIC^*_i{}_{ZOIB})$ is an indicator function taking on the value 1 when the *i*th bootstrap replication of AIC calculated for the ZOIMSP model is less than the AIC calculated for the ZOIB model.

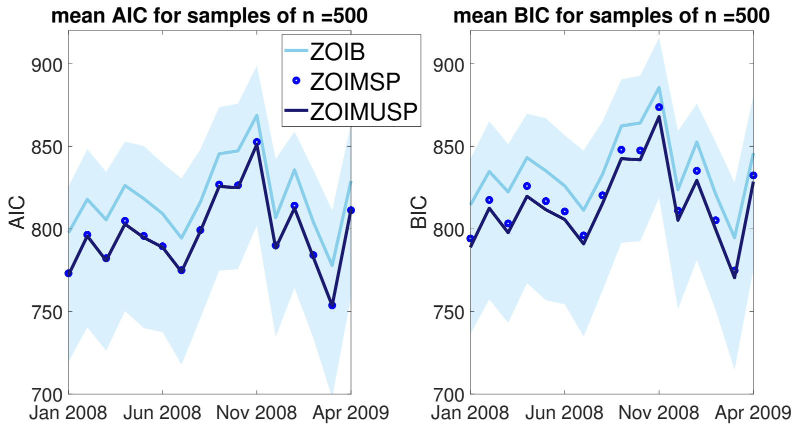

Figure 14. Bootstrap estimated mean of the AIC and BIC for the ZOIB, ZOIMSP, and ZOIMUSP models for a sample size of $n = 500$. The shaded area represents a 90% confidence interval for the mean of the ZOIMUSP model.

In Figure 15, the estimated probabilities in (13) are calculated for the AIC and BIC for both the ZOIMSP and ZOIMUSP models compared to the ZOIB model.

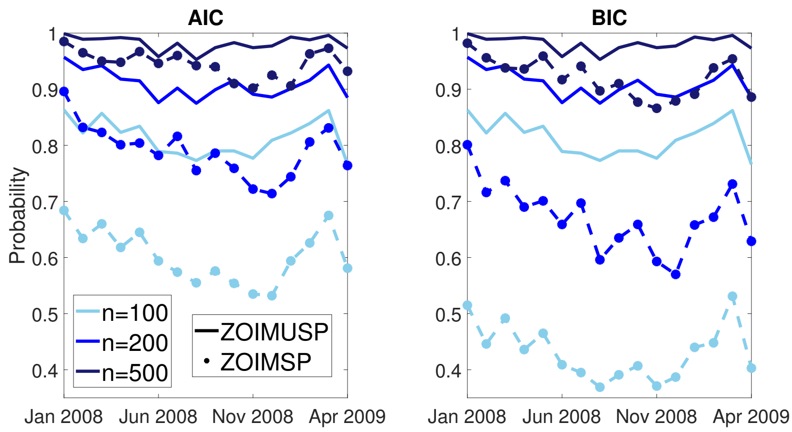

Figure 15. Bootstrap estimates of the probability that the AIC (or BIC) for the fitted ZOIMSP and ZOIMUSP models is less than the AIC (or BIC) of the fitted ZOIB model over several months and for sample sizes of $n = 100, 200$, and 500.

From Figure 15, it is evident that the ZOIMUSP outperforms the ZOIB for all sample sizes, with a clear improvement in relative performance as the sample size increases. The results for the ZOIMSP model again show the effect of an additional parameter on the AIC and BIC compared to the more parsimonious ZOIB and ZOIMSP models. In fact, in terms of the BIC, the results are slightly in favour of the ZOIB model for $n = 100$. However, the results show that as the sample size n increases, it is increasingly likely that the AIC (or BIC) for both the fitted ZOIMSP and ZOIMUSP models will be less than the AIC (or BIC) of the fitted ZOIB model.

4. Conclusions and Future Research

In this study, we propose two distributions to more accurately model U- and J-shaped fractional data, with parsimony and simplicity as priorities. We fit the models, both mixtures of standard power distributions, to the continuous part of our LGD data and evaluate model performance using traditional information criteria. In addition to outperforming the benchmark ZOIB distribution, which has been favoured in practice, both proposed models also reduce the bias in the estimation of the mean. We also show how, by evaluating the sign of the empirical bias of the fitted models, a model that provides a more conservative estimation of the mean can be identified.

The proposed mixture models exhibit flexibility to more accurately capture bimodal data where substantial modes exist close to the boundary values. It is these areas that pose a specific problem for the beta distribution, which for $\alpha < 1$ and $\beta < 1$ has asymptotes at 0 and/or 1. This is not only a common occurrence in financial applications, but also in biological studies. For example, due to the analytical limitations of the laboratory instruments used to quantify the clinical markers of disease, many observations are defined as below the limit of quantification (LOQ). For data analysis, these values are often replaced by a constant, usually the LOQl, however, if the LOQ is unknown, values are replaced by zeros or a fraction of the minimum observed. The interested reader is referred to Van Reenen et al. [24] and the references therein for an in-depth discussion.

The beta distribution is a popular choice to model the dependent variable in regression analysis, and specifically for fractional data. The alternative mixture models proposed in this study may therefore be employed to more accurately characterise the target variable where this is U- or J-shaped. Possible avenues for future research would be to consider constrained optimisation techniques to estimate the parameters of the MSP model to circumvent possible identifiability problems, as well as investigating the potential of the proposed distributions to model doubly bounded fractional dependent variables.

Author Contributions: J.L., G.L.G., J.S.A. contributed equally. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author, upon agreement from the financial institution. The data are not publicly available due to a confidentiality agreement between the authors and the financial institution.

Acknowledgments: The work of J.S. Allison are based on research supported by the National Research Foundation (NRF). Any opinion, finding, and conclusion or recommendation expressed in this material is that of the authors and the NRF does not accept any liability in this regard.

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

References

1. Ramalho, E.; Ramalho, J.; Coelho, L. Exponential Regression of Fractional-Response Fixed-Effects Models with an Application to Firm Capital Structure. *J. Econom. Methods* **2016**, *7*, 20150019. [CrossRef]
2. Falls, L.W. The beta distribution—A statistical model for world cloud cover. *J. Geophys. Res.* **1974**, *79*, 1261–1264. [CrossRef]
3. Kousathanas, A.; Keightley, P.D. A comparison of models to infer the distribution of fitness effects of new mutations. *Genetics* **2013**, *193*, 1197–1208. [CrossRef] [PubMed]
4. Stewart, J. The Beta Distribution as a Model of Behavior in Consumer Goods Markets. *Manag. Sci.* **1979**, *25*, 813–821. [CrossRef]
5. Tasche, D. The single risk factor approach to capital charges in case of correlated loss given default rates. *arXiv* **2004**, arXiv:cond-mat/0402390.

6. Damme, G.V. A generic framework for stochastic Loss-Given-Default. *J. Comput. Appl. Math.* **2011**, *235*, 2523–2550. [CrossRef]
7. Farinelli, S.; Shkolnikov, M. Two models of stochastic loss given default. *J. Credit. Risk* **2012**, *8*, 3–20. [CrossRef]
8. Simone, R. On finite mixtures of Discretized Beta model for ordered responses. *TEST* **2022**, *31*, 828–855. [CrossRef]
9. Simone, R.; Tutz, G. Modelling uncertainty and response styles in ordinal data. *Stat. Neerl.* **2018**, *72*, 224–245. [CrossRef]
10. James, S.A.; West, C.L.; Davey, R.P.; Dicks, J.L.; Roberts, I.N. Prevalence and Dynamics of Ribosomal DNA Micro-heterogeneity Are Linked to Population History in Two Contrasting Yeast Species. *Sci. Rep.* **2016**, *6*, 28555. [CrossRef] [PubMed]
11. Memmel, C.; Sachs, A.; Stein, I. Contagion at the Interbank Market with Stochastic LGD. SSRN 2794059. 2011. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2794059 (accessed on 1 November 2022).
12. Basel Committee on Banking Supervision. *An Explanatory Note on the Basel II IRB Risk Weight Functions*; Technical Report; Bank for International Settlements: Basel, Switzerland, 2004. Available online: http://www.bis.org/bcbs/irbriskweight.htm (accessed on 28 September 2022).
13. Ospina, R.; Ferrari, S.L.P. Inflated beta distributions. *Stat. Pap.* **2008**, *51*, 111–126. [CrossRef]
14. Calabrese, R.; Zenga, M. Bank loan recovery rates: Measuring and nonparametric density estimation. *J. Bank. Financ.* **2010**, *34*, 903–911. [CrossRef]
15. Basel Committee on Banking Supervision. *Guidance on Paragraph 468 of the Framework Document*; Technical Report; Bank for International Settlements: Basel, Switzerland, 2005. Available online: https://www.bis.org/publ/bcbs115.htm (accessed on 28 September 2022).
16. Jones, M. Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. *Stat. Methodol.* **2009**, *6*, 70–81. [CrossRef]
17. Kumaraswamy, P. A generalized probability density function for double-bounded random processes. *J. Hydrol.* **1980**, *46*, 79–88. [CrossRef]
18. Cribari-Neto, F.; Santos, J. Inflated Kumaraswamy distributions. *An. Acad. Bras. CiêNcias* **2019**, *91*, e20180955. [CrossRef] [PubMed]
19. Leemis, L.M.; McQueston, J. Univariate Distribution Relationships. *Am. Stat.* **2008**, *62*, 45–53. [CrossRef]
20. Ribeiro, M. A Zero-One Inflated Mixture Model for Loss Given Default Data: The Beta Distribution Case. In Proceedings of the XIV Conference on Credit Scoring and Credit Control, Edinburgh, Scotland, 29 May 2015.
21. Ribeiro, M.; Louzada, F.; Henrique, G.; Pereira, A.; Moreira, F.; Calabrese, R. Inflated Mixture Models: Applications to Multimodality in Loss Given Default. SSRN 2634919. 2015. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2634919 (accessed on 1 November 2022). [CrossRef]
22. Schröder, C.; Rahmann, S. A hybrid parameter estimation algorithm for beta mixtures and applications to methylation state classification. *Algorithms Mol. Biol.* **2017**, *12*, 21. [CrossRef] [PubMed]
23. Bishop, C.M. *Pattern Recognition and Machine Learning*, 5th ed.; Information Science and Statistics; Springer: New York, NY, USA, 2007.
24. Van Reenen, M.; Westerhuis, J.A.; Reinecke, C.J.; Venter, J.H. Metabolomics variable selection and classification in the presence of observations below the detection limit using an extension of ERp. *BMC Bioinform.* **2017**, *18*, 83. [CrossRef] [PubMed]

Chapter 4

Article 3: Modelling the Time to Write-Off of Non-Performing Loans using a Promotion Time Cure Model with Parametric Frailty

The third article, *Modelling the Time to Write-Off of Non-Performing Loans using a Promotion Time Cure Model with Parametric Frailty*, has been published in *Mathematics 2023; 11, 2228. https://doi.org/10.3390/math11102228.*

A summary of the guidelines to authors from the journal is presented below.

Manuscript	No restrictions on the length of manuscripts, provided that the text is concise and comprehensive.
Title	The title should be concise, specific and relevant.
Abstract and keywords	An abstract (maximum length of 200 words) in a single paragraph and three to ten pertinent keywords must be included.
Figures, Schemes and Tables	All Figures, Schemes and Tables should be inserted into the main text close to their first citation and must be numbered following their number of appearance. All Figures, Schemes and Tables should have a short explanatory title and caption. All table columns should have an explanatory heading.
References	References must be numbered in order of appearance in the text (including table captions and figure legends) and listed individually at the end of the manuscript.The reference list should include the full title, as recommended by the ACS style guide. In the text, reference numbers should be placed in square brackets [ ], and placed before the punctuation.
General formatting	A LaTeXtemplate is provided for submission.
Additional information	https://www.mdpi.com/journal/mathematics /instructionspreparation

mathematics

Article

Modelling the Time to Write-Off of Non-Performing Loans Using a Promotion Time Cure Model with Parametric Frailty

Janette Larney [1,*,†], James Samuel Allison [2,†], Gerrit Lodewicus Grobler [2,†] and Marius Smuts [2,†]

[1] Centre for Business Mathematics and Informatics, North-West University, Potchefstroom 2531, South Africa
[2] School of Mathematical and Statistical Sciences, North-West University, Potchefstroom 2531, South Africa
* Correspondence: janette.larney@nwu.ac.za
† These authors contributed equally to this work.

Abstract: Modelling the outcome after loan default is receiving increasing attention, and survival analysis is particularly suitable for this purpose due to the likely presence of censoring in the data. In this study, we suggest that the time to loan write-off may be influenced by latent competing risks, as well as by common, unobservable drivers, such as the state of the economy. We therefore expand on the promotion time cure model and include a parametric frailty parameter to account for common, unobservable factors and for possible observable covariates not included in the model. We opt for a parametric model due to its interpretability and analytical tractability, which are desirable properties in bank risk management. Both a gamma and inverse Gaussian frailty parameter are considered for the univariate case, and we also consider a shared frailty model. A Monte Carlo study demonstrates that the parameter estimation of the models is reliable, after which they are fitted to a real-world dataset in respect of large corporate loans in the US. The results show that a more flexible hazard function is possible by including a frailty parameter. Furthermore, the shared frailty model shows potential to capture dependence in write-off times within industry groups.

Keywords: survival analysis; write-off time; loan recovery; promotion time cure model; parametric frailty; shared frailty

MSC: 62N01; 62N02; 62P05

Citation: Larney, J.; Allison, J.S.; Grobler, G.L.; Smuts, M. Modelling the Time to Write-Off of Non-Performing Loans Using a Promotion Time Cure Model with Parametric Frailty. *Mathematics* **2023**, *11*, 2228. https://doi.org/10.3390/math11102228

Academic Editors: David Zapletal and Manuel Alberto M. Ferreira

Received: 21 March 2023
Revised: 21 April 2023
Accepted: 5 May 2023
Published: 9 May 2023

1. Introduction

Risk management in the credit risk environment has advanced considerably over the last few decades, and during the same time, statistical models developed for this purpose have grown in variety and sophistication. Initially, statistical modelling focused on the assessment of a counterparty's creditworthiness (i.e., default risk), but the modelling of the recovery process and outcome after default are receiving increasing attention. Two main outcomes or resolutions are possible after default: either the outstanding amount is fully recovered or the loan is written off. Modelling the time to write-off of a defaulted loan is valuable to a bank, as the time on book of such loans not only affects its operational decisions but also its profitability [1]. Survival analysis has enjoyed increasing popularity in the credit risk environment for modelling of the time to default and, more recently, to model the time to resolution of a defaulted loan (see, e.g., [2–5]). Since write-offs are typically associated with longer resolution times than full recoveries, the number of write-offs is usually under-represented in samples of defaulted loans [6]. This makes survival analysis particularly suitable for modelling the time to write-off because the status at the end of the observation period is often recorded as "not resolved", i.e., censoring is present.

With the fast-growing body of literature on the use of survival analysis (and specifically the Cox proportional hazard (CPH) model) to model default risk, an exploration of the use of the methodology to describe the post-default outcome of a loan is a natural outflow.

Several authors have made use of Cox regression to demonstrate that loan and obligor characteristics, as well as macroeconomic factors, drive the resolution time of defaulted loans (see e.g., [3,4,7,8]).

Recently, both mixture cure and promotion time cure (PTC) models have found applications in the credit risk environment (see e.g., [9–12]). Recognising that there is a proportion of defaulted loans that is not "susceptible" to a full recovery, de Oliveira and Louzada [5] modelled the recovery process of defaulted loans using a PTC model. The model considers the full repayment of a defaulted loan as the event of interest and allows for latent factors that may impact the post-default resolution outcome.

The time to write-off or full recovery may, in addition to latent competing risks, also be influenced by common, unobservable drivers, such as the state of the economy [3]. In this study, we propose an expansion of the PTC model by including a frailty parameter to control for unobserved heterogeneity within our population of defaulted loans. Underlying the post-default recovery process is the macroeconomic environment, which has a systematic effect on all borrowers' ability to resolve their defaulted obligations. Although some of these are observable and can be included as covariates, we propose a frailty parameter that can account for common, unobservable factors, as well as for possible observable covariates not included in the model. We consider both the univariate and the shared frailty parameter. Global regulatory standards, specifically those set by the Basel Accords, require banks to have a comprehensive framework for model risk management, which includes ongoing monitoring and assessment of model performance. We demonstrate how inclusion of a frailty parameter in the PTC model results in improved model performance when estimating the time to write-off of a non-performing loan.

Our paper offers the following contributions. First, we formulate the PTC model with gamma frailty and derive its likelihood function. Second, we compare the performance of this model to that of the standard PTC model, as well as to a PTC model that includes an inverse Gaussian frailty parameter, as proposed by Barriga et al. [13]. For the comparison, we first perform a simulation study, then fit the models to both a simulated and a real-world dataset relating to the resolution outcomes of defaulted corporate loans. Third, we extend the PTC model to include a shared frailty parameter. We derive the likelihood function of this model and evaluate its performance on the same real-world dataset.

The probability of write-off (or recovery) is a key component within two-step loss given default (LGD) models, where the post-default outcome probability and loss severity are characterised separately (see, e.g., [14]). Improving on the accuracy of LGD models should therefore enable a bank to attain more accurate estimates of this risk parameter for the determination of credit loss provisions, which should, in turn, result in more stable earnings. It will also enable the bank to more accurately assess its regulatory and/or economic capital requirements, which are crucial for a bank's financial stability. Obtaining a more accurate estimate of the time to write-off will, furthermore, enable the bank to better manage its cash flow requirements and reduce the liquidity risk to which it is exposed.

The remainder of this article is structured as follows. In the next section, an overview of survival modelling is provided, including cure models and their application in the credit risk environment. The formulation of the proposed model, namely the PTC model with parametric frailty, follows in Section 3. In Section 4, the likelihood function of this model is derived, and the finite sample performance of maximum likelihood estimators (MLEs) of the frailty models is analysed through a Monte Carlo simulation study. We apply our model to both a simulated dataset and a real-world loan loss dataset in Section 5. In Section 6, we describe the extension of the PTC model with a shared frailty parameter, including an application to a real-world dataset. We provide some concluding remarks and avenues for future research in Section 7.

2. Background

Survival modelling is a statistical method used to analyse and predict the time until an event of interest occurs. The event of interest is often referred to as the "failure" or "death" event, and the time until this event is referred to as the "survival time". The event of interest in our context is the write-off of a defaulted loan. In general, the survival function is defined as $S(t) = P(T > t)$, where T is the random variable representing the time to the event, and t is a specific point in time.

Within the resolution process of defaulted loans, random right censoring is normally present due to the fixed duration of the observation period. Not all subjects experience the event of interest, i.e., write-off, within that time period. In right censoring, there are two latent random variables: T and C, where, T is the time of failure, and C is the censoring time. The random observed variable is (Y, δ), where

$$Y = \min(T, C)$$

and

$$\delta = \begin{cases} 1 & \text{if } T \leq C \\ 0 & \text{if } T > C \end{cases}.$$

The simplest way of modelling the time to write-off is to assign a distribution with positive support to T, for example, the gamma or Weibull distributions.

However, for credit risk modelling, specifying $S(t)$ unconditionally without taking covariates into account is too simplistic, and the CPH model has therefore dominated.

The CPH model, which was proposed in the seminal paper of Cox [15], assumes that the hazard function is a function of a set of covariates and that the effect of the covariates on the hazard function is proportional over time.

The hazard function $h(t) = f(t)/S(t)$ of a survival process represents the instantaneous risk of the event of interest at a given time, where f is the density of T. For the CPH model, the hazard function, conditional on a vector of covariates $\mathbf{x} = (x_1, x_2, \ldots, x_m)^T$, has the form

$$h(t \mid \mathbf{x}) = h_0(t)e^{\boldsymbol{\beta}^T\mathbf{x}}, \tag{1}$$

where $h_0(t)$ is the baseline hazard function, and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_m)^T$ is the vector of unknown regression parameters associated with $\mathbf{x}$. The conditional survival function is then given by

$$S(t \mid \mathbf{x}) = P(T > t \mid \mathbf{x}) = S_0(t)^{\exp(\boldsymbol{\beta}^T\mathbf{x})},$$

where $S_0(t) = S(t \mid \mathbf{x} = \mathbf{0})$ is the baseline survival function.

The earliest studies to make use of the CPH model to model default risk include those of Banasik et al. [16], Stepanova and Thomas [17] and Cao et al. [18]; subsequently, its application to credit risk problems has increased markedly. Several authors, including Betz et al. [3], Malwandla et al. [4] and Joubert et al. [8], have since made use of Cox regression to model the post-default outcome of non-performing loans.

Notwithstanding the popularity of the CPH model, cure models have recently enjoyed increasing attention within the credit risk domain. This is particularly true for the modelling of the probability of default, which, in many high-quality loan portfolios, represents a rare event (see, e.g., [9–11]). Cure models, as all survival models, had their origin in medicine, where Boag [19] first suggested that there exists a certain proportion of cancer patients who will never experience the event of interest, i.e., they can be considered cured. By not accounting for the cure proportion, the estimates of the survival of susceptible subjects may be biased [20].

In early work by Berkson and Gage [21] the following simple model was proposed:

$$S(t) = \pi_0 S^c(t) + (1 - \pi_0)S^s(t), \tag{2}$$

where S^c and S^s are the survival functions for non-susceptible and susceptible individuals, respectively, and π_0 is the proportion of individuals considered non-susceptible, also known as the cure proportion. Note that $S^c(t) = 1$ for all $t \geq 0$, as the individuals being modelled are not susceptible to the event of interest and therefore have a probability of survival equal to one.

The model in (2) was extended by Farewell [22] to allow for covariates. Suppose another set of covariates $\mathbf{y}$ that may be equal to $\mathbf{x}$ is observed. The conditional survival function for a mixture cure model is defined as:

$$S(t \mid \mathbf{x}, \mathbf{y}) = \pi_0(\mathbf{x}) + (1 - \pi_0(\mathbf{x}))S^s(t \mid \mathbf{y}), \tag{3}$$

where $\pi_0(\mathbf{x})$ is commonly modelled using a logistic regression model, and $S^s(t \mid \mathbf{y})$ is modelled by a CPH or accelerated failure time model. For a detailed overview of mixture cure models, the reader is referred to the review paper by Amico and van Keilegom [23].

The second type of cure model, known as the promotion time cure (PTC) model or, alternatively, the bounded cumulative hazard model, was first introduced by Yakovlev et al. [24]. The PTC model has an intuitive biological interpretation: an individual cancer patient has a number of carcinogenic cells (denoted by N) left active after initial treatment, which have the potential to activate over time and produce a significant relapse of cancer. Chen et al. [25] further proposed that an individual is at risk of relapse if he or she is exposed to at least one of the so-called latent factors, and relapse (or failure) is observed when one (or more) of these latent factors becomes activated. If the individual is not considered to be at risk, he or she is considered cured.

Formally, let N be the unobservable number of causes of the event of interest (write-off in our case) for a given subject in the population. For a subject for which $N \geq 1$, the event time is defined as

$$T = \min(Z_1, Z_2, \ldots, Z_N),$$

where $Z_1, Z_2, \ldots, Z_N$ represents the time required for the jth unobservable cause to be realised. A subject is cured if $N = 0$, which implies that $T = \infty$, i.e., write-off will not happen. The set of observed covariates is denoted by $\mathbf{x} = (1, x_1, x_2, \ldots, x_m)^T$. Rodrigues et al. [26] demonstrate that the long-term survival function satisfies the proportional hazards property if and only if the number of risks follows a Poisson distribution. Assume that N follows a Poisson distribution with a mean of $\theta(\mathbf{x}) > 0$, where $\theta(\mathbf{x}) = g(\boldsymbol{\beta}^T\mathbf{x})$ for a strictly increasing function g (see e.g., Zhang and Peng [27] and Portier et al. [28]). Furthermore, assume that $Z_1, Z_2, \ldots, Z_N$ are i.i.d. from a distribution $F_{\boldsymbol{\eta}}$ with parameter $\boldsymbol{\eta}$, which could be a vector and independent of N. We then arrive at the well-known PTC model,

$$S_{\boldsymbol{\eta}}(t \mid \mathbf{x}) = P(T > t \mid \mathbf{x}) = e^{-\theta(\mathbf{x})F_{\boldsymbol{\eta}}(t)}. \tag{4}$$

For a short derivation of (4), the interested reader is referred to the work of Chen et al. [25]. It is clear from (4) that the cure fraction (i.e., for those not susceptible to write-off) is given by

$$\lim_{t \to \infty} S_{\boldsymbol{\eta}}(t \mid \mathbf{x}) = e^{-\theta(\mathbf{x})} = \pi_0(\mathbf{x}).$$

To obtain a similar expression as in (3), the survival function in (4) can be decomposed to obtain

$$S_{\boldsymbol{\eta}}(t \mid \mathbf{x}) = \pi_0(\mathbf{x}) + (1 - \pi_0(\mathbf{x}))S^s_{\boldsymbol{\eta}}(t \mid \mathbf{x}),$$

where

$$S^s_{\boldsymbol{\eta}}(t \mid \mathbf{x}) = \frac{e^{-\theta(\mathbf{x})F_{\boldsymbol{\eta}}(t)} - \pi_0(\mathbf{x})}{1 - \pi_0(\mathbf{x})} \tag{5}$$

represents the survival function of those who are susceptible to write-off. Note that $S^s_{\boldsymbol{\eta}}(t \mid \mathbf{x})$ is a proper survival function in that $\lim_{t \to \infty} S^s_{\boldsymbol{\eta}}(t \mid \mathbf{x}) = 0$.

The distribution function $F_{\boldsymbol{\eta}}$ can be specified parametrically, e.g., by considering an exponential, Weibull or gamma distribution or, alternatively, left completely unspecified, resulting in a semiparametric PTC model. For an in-depth discussion relating to the estimation and inference of (4), the reader is referred to the review paper by Amico and van Keilegom [23].

We argue that the PTC model is also interpretable in the post-default loan resolution context. There exist several obligor- and loan-specific factors that may influence a loan's probability of being written off. Some of these factors are observable and can be included in a model as covariates, but some factors, for example, an individual's level of discretionary expenditure and undisclosed debt, are latent and therefore analogous to the carcinogenic cells modelled by Chen et al. [25]. There also exists a proportion of defaulted loans for which the outstanding amounts will be fully recovered and is therefore not exposed to write-off. In our context, these loans can be considered "cured".

de Oliveira and Louzada [5] made use of a PTC model to model the recovery process of defaulted loans within the context of latent competing risks. The considered event of interest is full loan recovery, and the cure proportion represents the proportion of obligors not susceptible to fully repay the outstanding loan amount.

The time to write-off or full recovery may, in addition to latent competing risks, also be influenced by common, unobservable drivers, such as the state of the economy. Within the credit risk environment and specifically in modelling the time to default, several authors have accounted for these unobservable factors by including a frailty parameter within the CPH framework (see e.g., [29–31]). An advantageous feature of the CPH model is the ease with which a frailty parameter can be included by multiplication with the baseline hazard function. A frailty parameter is a random variable that can account for unobserved heterogeneity in a population [32]. With the univariate frailty model, the frailty term is assumed to be a random variable that affects the failure times of individuals independently. In a shared frailty model, on the other hand, the frailty term is assumed to be shared among the individuals within a cluster or group [33]. For the CPH model, the frailty parameter is incorporated as follows:

$$h(t \mid \mathbf{x}, W) = W h_0(t) e^{\boldsymbol{\beta}^T \mathbf{x}}, \tag{6}$$

where the frailty term W is a random variable with non-negative support, such as the gamma distribution, inverse Gaussian (IG) distribution or log-normal distribution. The conditional survival function of the CPH model, given the frailty $W = w$, is

$$S(t \mid \mathbf{x}, w) = P(T > t \mid \mathbf{x}, W = w) = S_0(t)^{w \exp\left(\boldsymbol{\beta}^T \mathbf{x}\right)}.$$

In the context of the post-default loan resolution process, Betz et al. [3] accounted for unobservable systematic effects in the comovement of the resolution times of defaulted loans by including a random frailty parameter in the CPH model.

We propose including a frailty component in the PTC model of Chen et al. [25] to model the time to write-off. The frailty parameter, which is assumed to have a parametric form, can control for common unobservable factors, as well as for possible observable covariates not included in the model. The PTC model with an IG frailty was used by Barriga et al. [13] in a survival study of colorectal cancer patients, which is, to the best of our knowledge, the only study to include a frailty effect within the PTC framework.

3. Proposed Model: The Promotion Time Cure Model with Parametric Frailty

To account for the heterogeneity in the time to write-off of individual obligors, we propose an extension of the model in (4) by including a parametric frailty component. We opt for a parametric model due to its interpretability and analytical tractability. These characteristics are highly valued in the bank risk management environment, where most models are subjected to intense regulatory scrutiny. A further benefit of parametric models is the ease with which they can be subjected to stress and sensitivity testing.

We first consider the univariate case and include a gamma distributed frailty parameter as follows. Assume that N has a Poisson distribution with parameter $W\theta(\mathbf{x})$, where W is an unobserved gamma distributed random variable with parameter $\sigma^2 > 0$ and density function

$$f_{\sigma^2}(w) = \frac{\sigma^{-2/\sigma^2}}{\Gamma(1/\sigma^2)} w^{1/\sigma^2 - 1} e^{-w/\sigma^2},\ w \geq 0. \tag{7}$$

Li et al. [34] showed that the model is identifiable if the distribution function of W is specified. In this case, $\mathrm{E}[W] = 1$, which assures that the model is identifiable, and $\mathrm{Var}[W] = \sigma^2$.

From (4), we derive the conditional survival function, given $W = w$, as

$$\tilde{S}_{\boldsymbol{\eta}}(t \mid \mathbf{x},\ w) = P(T > t \mid \mathbf{x},\ W = w) = e^{-w\theta(\mathbf{x})F_{\boldsymbol{\eta}}(t)}. \tag{8}$$

The unconditional survival function is then

$$\begin{aligned} \tilde{S}_{\boldsymbol{\eta}}(t \mid \mathbf{x}) &= \int_0^\infty \tilde{S}_{\boldsymbol{\eta}}(t \mid \mathbf{x}, w) f_W(w) \mathrm{d}w \\ &= \int_0^\infty e^{-w\theta(\mathbf{x})F_{\boldsymbol{\eta}}(t)} f_W(w) \mathrm{d}w \\ &= \varphi_w\big(\theta(\mathbf{x})F_{\boldsymbol{\eta}}(t)\big) \\ &= \Big\{1 + \sigma^2\theta(\mathbf{x})F_{\boldsymbol{\eta}}(t)\Big\}^{-1/\sigma^2}, \end{aligned} \tag{9}$$

where $\varphi_w(s) = \mathrm{E}\big[e^{-sW}\big]$ is the Laplace transform of the frailty W.

It is clear from (9) that

$$\lim_{t\to\infty} \tilde{S}_{\boldsymbol{\eta}}(t \mid \mathbf{x}) = \Big\{1 + \sigma^2\theta(\mathbf{x})\Big\}^{-1/\sigma^2} = \pi_0(\mathbf{x}), \tag{10}$$

where π_0 is the cure fraction of the PTC model with gamma frailty, and σ^2 can be seen as a heterogeneity parameter.

From some algebraic manipulation, it follows that (9) can be expressed as

$$\tilde{S}_{\boldsymbol{\eta}}(t \mid \mathbf{x}) = \pi_0(\mathbf{x}) + (1 - \pi_0(\mathbf{x}))\tilde{S}^s_{\boldsymbol{\eta}}(t \mid \mathbf{x}), \tag{11}$$

where

$$\tilde{S}^s_{\boldsymbol{\eta}}(t \mid \mathbf{x}) = \frac{\{1 + \sigma^2\theta(\mathbf{x})F_{\boldsymbol{\eta}}(t)\}^{-1/\sigma^2} - \pi_0(\mathbf{x})}{1 - \pi_0(\mathbf{x})} \tag{12}$$

represents the survival function of those who are susceptible to write-off. Note that $\tilde{S}^s_{\boldsymbol{\eta}}(t \mid \mathbf{x})$ is a proper survival function in the sense that $\lim_{t\to\infty} \tilde{S}^s_{\boldsymbol{\eta}}(t \mid \mathbf{x}) = 0$ and $\lim_{t\to 0} \tilde{S}^s_{\boldsymbol{\eta}}(t \mid \mathbf{x}) = 1$. From (11), it is clear that the survival function of the non-susceptible individuals is 1, whereas the proper survival function of those who are susceptible is a function of the frailty parameter, σ^2. However, from (10), it follows that the cure fraction is dependent on the frailty parameter.

4. Estimation and Simulation

In this section, we discuss estimation of the parametric PTC model with a gamma frailty parameter using maximum likelihood estimation (MLE). We illustrate the small sample performance of the MLEs through a simulation study.

4.1. Maximum Likelihood Estimation of the PTC Model with Gamma Frailty

For the jth individual, $j = 1, 2, \ldots, n$, we observe the triplet $(Y_j, \delta_j, \mathbf{x}_j)$, where $Y_j = \min(T_j, C_j)$ is the jth observed time, $\delta_j = I\big(T_j \leq C_j\big)$ is the corresponding indicator, T_j is the time to write-off, C_j the censoring time and $\mathbf{x}_j = \big(1, x_{1j}, x_{2j}, \ldots, x_{mj}\big)^T$ is the vector of m covariates associated with the jth individual. Throughout, we assume that the model specified in (8) is a parametric model, i.e., we assume that F has a parametric form with an unknown parameter $\boldsymbol{\eta}$ and that θ is related to the covariates by the relationship $\theta_{\boldsymbol{\beta}}(\mathbf{x}_j) = e^{\boldsymbol{\beta}^T \mathbf{x}_j}$, where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_m)^T$ is a vector of unknown regression parameters.

The likelihood function is given by

$$\begin{aligned} L &= \prod_{j=1}^{n} \tilde{S}_{\boldsymbol{\eta}}(Y_j \mid \mathbf{x}_j) h(Y_j \mid \mathbf{x}_j)^{\delta_j} \\ &= \prod_{j=1}^{n} \tilde{S}_{\boldsymbol{\eta}}(Y_j \mid \mathbf{x}_j) \left(-\frac{\mathrm{d}}{\mathrm{d}t} \log \tilde{S}_{\boldsymbol{\eta}}(Y_j \mid \mathbf{x}_j \right)^{\delta_j}, \end{aligned}$$

where $\tilde{S}_{\boldsymbol{\eta}}(t \mid \mathbf{x}_j)$ is given in (9). The log-likelihood of the PTC model with gamma frailty is therefore

$$\ell\left(\boldsymbol{\beta}, \boldsymbol{\eta}, \sigma^2\right) = -\sum_{j=1}^{n} \frac{1}{\sigma^2} \log\left(1 + \sigma^2 \theta_{\boldsymbol{\beta}}(\mathbf{x}_j) F_{\boldsymbol{\eta}}(Y_j)\right) + \sum_{j=1}^{n} \delta_j \log\left(\frac{\theta_{\boldsymbol{\beta}}(\mathbf{x}_j) f_{\boldsymbol{\eta}}(Y_j)}{1 + \sigma^2 \theta_{\boldsymbol{\beta}}(\mathbf{x}_j) F_{\boldsymbol{\eta}}(Y_j)} \right). \quad (13)$$

For mixture cure models, a popular approach to estimate model parameters is to apply the EM algorithm. To this end, the mixture cure representation of the model in (11) is used. Recall that in a mixture cure model,

$$S^s(t \mid \mathbf{x}) = P(T > t \mid B = 1, \mathbf{x}),$$

where B is the cure status (or susceptibility) indicator, $B = 1$ indicates that an individual is susceptible to the event of interest (i.e., uncured) and $B = 0$ corresponds to individuals who are non-susceptible (i.e., cured). For the proposed model, (10) indicates that

$$P(B = 0 \mid \mathbf{x}) = \pi_0(\mathbf{x}) = \left\{1 + \sigma^2 \theta(\mathbf{x})\right\}^{-1/\sigma^2}.$$

Suppose that for the jth individual, $j = 1, 2, \ldots, n$, we observe not only the triplet $(Y_j, \delta_j, \mathbf{x}_j)$ but also B_j, the indicator showing whether an observation is susceptible or not. It is then possible to define the complete likelihood (see [35]) as

$$\begin{aligned} L = &\prod_{j=1}^{n} \left[(1 - \pi_0(\mathbf{x}_j)) \tilde{h}^s_{\boldsymbol{\eta}}(Y_j \mid \mathbf{x}) \tilde{S}^s_{\boldsymbol{\eta}}(Y_j \mid \mathbf{x}) \right]^{\delta_j B_j} \\ &\times \prod_{j=1}^{n} \left[\pi_0(\mathbf{x}_j)^{1 - B_j} \left\{ (1 - \pi_0(\mathbf{x}_j)) \tilde{S}^s_{\boldsymbol{\eta}}(Y_j \mid \mathbf{x}_j) \right\}^{B_j} \right]^{1 - \delta_j}, \end{aligned}$$

where $\pi_0(\mathbf{x}_j)$ and $\tilde{S}^s_{\boldsymbol{\eta}}$ are given in (10) and (12), respectively, while

$$\tilde{h}^s_{\boldsymbol{\eta}}(t \mid \mathbf{x}_j) = -\frac{\mathrm{d}}{\mathrm{d}t} \log \tilde{S}^s_{\boldsymbol{\eta}}(t \mid \mathbf{x}_j)$$

represents the hazard function of those who are susceptible to write-off.

This likelihood can be rewritten as

$$\begin{aligned} L = &\prod_{j=1}^{n} (1 - \pi_0(\mathbf{x}_j))^{\delta_j B_j} \pi_0(\mathbf{x}_j)^{(1-\delta_j)(1-B_j)} (1 - \pi_0(\mathbf{x}_j))^{(1-\delta_j) B_j} \\ &\times \prod_{j=1}^{n} \left[\tilde{h}^s_{\boldsymbol{\eta}}(Y_j \mid \mathbf{x}) \tilde{S}^s_{\boldsymbol{\eta}}(Y_j \mid \mathbf{x}) \right]^{\delta_j B_j} \left\{ \tilde{S}^s_{\boldsymbol{\eta}}(Y_j \mid \mathbf{x}_j) \right\}^{(1-\delta_j) B_j}. \end{aligned}$$

For the mixture cure model in (3), the survival function of the susceptible individuals $S^s(t \mid \mathbf{y})$ is not a direct function of the cure fraction π_0. Unfortunately, this is not the case for the survival function $\tilde{S}^s(t \mid \mathbf{x})$, when the PTC model is written in its mixture cure form, as can be seen from (12). Consequently, unlike for the "traditional" mixture cure model, the likelihood presented above cannot be split into two separate components, namely a latency part and an incidence part, which could have then be maximised separately (see, e.g., [9,36]). We therefore settle for the conventional method to obtain maximum likelihood estimators for the parameters of the PTC model with gamma frailty, i.e., we find the set of parameter values (β, $\boldsymbol{\nu}$ and σ^2) that maximises the log-likelihood function in (13). The required optimisation was performed by applying the Nelder–Mead algorithm, which is included in the optim function of the R statistical software package [37].

Similar to the study by Barriga et al. [13], we also consider a PTC model with IG frailty. In this case, the log-likelihood function is given by

$$\ell(\boldsymbol{\beta}, \lambda, k, \zeta) = \frac{n}{\zeta} + \sum_{j=1}^{n} \delta_j \log \theta_{\boldsymbol{\beta}}(\mathbf{x}_j) + \sum_{j=1}^{n} \delta_j \log F_{\boldsymbol{\eta}}(Y_j) - \frac{1}{2}\sum_{j=1}^{n} \delta_j \log\big(1 + 2\zeta\theta_{\boldsymbol{\beta}}(\mathbf{x}_j)F_{\boldsymbol{\eta}}(Y_j)\big) - \frac{1}{\zeta}\sum_{j=1}^{n} \sqrt{1 + 2\zeta\theta_{\boldsymbol{\beta}}(\mathbf{x}_j)F_{\boldsymbol{\eta}}(Y_j)}, \tag{14}$$

where ζ is the frailty parameter.

In the simulation study, we choose $F_{\boldsymbol{\eta}}$ as the Weibull and linear failure rate (LFR) distributions. In the case of the Weibull distribution, $\boldsymbol{\eta} = (\lambda, k)^T$, where $\lambda > 0$ is a scale parameter and $k > 0$ is a shape parameter with distribution function

$$F_{\lambda,k}(t) = 1 - e^{-\left(\frac{t}{\lambda}\right)^k},\ t \geq 0 \tag{15}$$

and density function

$$f_{\lambda,k}(t) = \frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1} e^{-\left(\frac{t}{\lambda}\right)^k},\ t \geq 0. \tag{16}$$

In the case of the LFR distribution, $\boldsymbol{\eta} = (a, b)^T$, where $a > 0$ and $b > 0$ are shape parameters, and the distribution function is given by

$$F_{a,b}(t) = 1 - e^{-at - \frac{b}{2}t^2},\ t \geq 0$$

with density function

$$f_{a,b}(t) = (a + bt)e^{-at - \frac{b}{2}t^2},\ t \geq 0.$$

4.2. Simulation Study

To evaluate the performance of the MLEs, based on the log-likelihood given in (13) (and also for comparison purposes, the log-likelihood given in (14)), we conducted a simulation study as follows. Two covariates are used; the first, $\mathbf{x}_{j1}, j = 1, \ldots, n$, is generated from a standard exponential distribution, whereas the second, $\mathbf{x}_{j2}, j = 1, \ldots, n$, is generated from a Bernoulli distribution with a probability of success of 0.6. These covariates remain fixed throughout the Monte Carlo simulation study. The cure proportion is given by

$$p_j = \left\{1 + \sigma^2\theta_{\boldsymbol{\beta}}(\mathbf{x}_j)\right\}^{-1/\sigma^2},$$

where

$$\theta_{\boldsymbol{\beta}}(\mathbf{x}_j) = e^{\boldsymbol{\beta}^T\mathbf{x}_j} = e^{\beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2}}.$$

As mentioned earlier, we take $F_{\boldsymbol{\eta}}$ to be either the Weibull or LFR distribution. In the former case, $\boldsymbol{\eta} = (\lambda, k)^T = (2, 1)^T$, and in the latter case, $\boldsymbol{\eta} = (a, b)^T = (1.5, 2)^T$. Censoring times $C_j, j = 1, \ldots, n$ are sampled from a standard uniform distribution, and the frailty parameter, σ^2, is set to 0.5. In order to control the cured and censoring proportions, we consider two different values for the regression coefficient vector, namely $\boldsymbol{\beta} = (2, -1, 2)^T$

and $\boldsymbol{\beta} = (0.5, -1, 2)^T$. The first calibration (in conjunction with the parameter choices for $F_{\boldsymbol{\eta}}$) results in an average cure proportion of 10% and a 20% censoring proportion, while the second results in an average cure proportion of 30%, with a 45% censoring proportion. The jth resolution time T_j from a PTC model with gamma frailty is generated as follows:

1. Generate B_j from a Bernoulli distribution:

$$B_j = \begin{cases} 1 & \text{with probability } 1 - p_j \\ 0 & \text{with probability } p_j \end{cases}$$

2. If $B_j = 0$, then set $T_j = \infty$;
3. If $B_j = 1$, generate $U_j \sim U[0, 1]$, and calculate

$$T_j = F_{\boldsymbol{\eta}}^{-1}\left\{ \frac{\left(U_j(1 - p_j) + p_j\right)^{-\sigma^2} - 1}{\sigma^2 \theta_{\boldsymbol{\beta}}(\mathbf{x}_j)} \right\},$$

   where $F_{\boldsymbol{\eta}}^{-1}$ is the inverse cumulative distribution function;
4. Generate censoring times $C_j \sim U[0, 1]$;
5. The simulated data are then $(Y_j, \delta_j, \mathbf{x}_{j1}, \mathbf{x}_{j2}), j = 1, \ldots, n$, where

$$Y_j = \min(T_j, C_j),$$

   and

$$\delta_j = I(T_j \leq C_j).$$

All calculated values are based on 1000 independent Monte Carlo replications for samples of size $n = 200,\ 300,\ 500$ and 1000.

Table 1 contains the approximate average values of the MLEs, as well as the corresponding standard errors given in parenthesis, for the PTC model with either the gamma or IG frailty and $F_{\boldsymbol{\eta}}$ as either a Weibull or LFR distribution. The average cure fraction is 10%, and the censoring proportion is 20%. Table 2 shows the results for a 30% cure fraction and a 45% censoring proportion, with all other aspects identical to the set-up of Table 1.

Tables 1 and 2 show the properties that we expect from MLEs, namely that they are unbiased and consistent. This is shown by the average estimates approaching the true values of the parameters, as well as the decreasing standard errors, as the sample size increases. Table 1 shows that for the IG frailty parameter ζ, the bias and the standard errors of the estimates are greater for smaller samples when $F_{\boldsymbol{\eta}}$ is the Weibull distribution compared to the LFR distribution. We also observe a slower rate of convergence in this case. The same holds for the rate of convergence in the case of a higher cure fraction and censoring percentage, as observed in Table 2. When $F_{\boldsymbol{\eta}}$ is the LFR distribution, larger standard errors are observed for smaller sample sizes ($n = 200$ and $n = 300$). Overall, the results demonstrate that the MLEs, based on the log-likelihood for the PTC model with gamma frailty, are accurate for small samples, especially when comparing their performance to those of the PTC model with an IG frailty. Although not shown in this paper, confidence intervals for the parameters can be estimated using the following bootstrap algorithm, which makes use of the "cases bootstrap" approach discussed in [38]:

1. Resample $(Y_1^*, \delta_1^*, \mathbf{x}_1^*), \ldots, (Y_n^*, \delta_n^*, \mathbf{x}_n^*)$ i.i.d from the empirical distribution function of $(Y_1, \delta_1, \mathbf{x}_1), \ldots, (Y_n, \delta_n, \mathbf{x}_n)$;
2. Fit the model in (9) on the resampled data to obtain estimates $\hat{\boldsymbol{\vartheta}}^* = \left(\hat{\boldsymbol{\beta}}^*, \hat{\boldsymbol{\eta}}^*, \hat{\sigma}^*\right)^T$. Denote the first of these by $\hat{\boldsymbol{\vartheta}}^*(1)$;
3. Repeat steps 1 and 2 B times to obtain $\hat{\boldsymbol{\vartheta}}^*(1), \hat{\boldsymbol{\vartheta}}^*(2), \ldots, \hat{\boldsymbol{\vartheta}}^*(B)$.

Now, suppose that we want to construct a confidence interval for the parameter β_1. From $\hat{\boldsymbol{\vartheta}}^*(1), \hat{\boldsymbol{\vartheta}}^*(2), \ldots, \hat{\boldsymbol{\vartheta}}^*(B)$, we have the bootstrap estimators $\hat{\beta}_1^*(1), \ldots, \hat{\beta}_1^*(B)$, from which a confidence interval can be obtained. Alternatively, one can make use of the large sample properties of MLEs to construct a confidence interval (see, e.g., [13]).

Table 1. Simulation results with 10% cure fraction and 20% censoring.

Parameter	$n = 200$	$n = 300$	$n = 500$	$n = 1000$
PTC with gamma frailty and F_η a Weibull distribution				
$\beta_0 = 2$	2.314 (0.988)	2.165 (0.706)	2.120 (0.512)	2.049 (0.330)
$\beta_1 = -1$	−1.035 (0.191)	−1.016 (0.154)	−1.016 (0.116)	−1.004 (0.085)
$\beta_2 = 2$	2.047 (0.347)	2.025 (0.276)	2.017 (0.216)	2.011 (0.152)
$\lambda = 2$	2.070 (1.217)	2.064 (0.988)	2.027 (0.782)	1.998 (0.560)
$\nu = 1$	1.018 (0.113)	1.012 (0.090)	1.009 (0.069)	1.002 (0.048)
$\sigma^2 = 0.5$	0.517 (0.302)	0.506 (0.247)	0.516 (0.179)	0.502 (0.127)
PTC with gamma frailty and F_η a LFR distribution				
$\beta_0 = 2$	2.318 (0.833)	2.199 (0.651)	2.170 (0.493)	2.110 (0.353)
$\beta_1 = -1$	−1.059 (0.195)	−1.028 (0.152)	−1.022 (0.108)	−1.009 (0.076)
$\beta_2 = 2$	2.103 (0.361)	2.050 (0.262)	2.029 (0.197)	2.023 (0.135)
$a = 1.5$	1.338 (0.804)	1.391 (0.682)	1.381 (0.551)	1.398 (0.415)
$b = 2$	3.337 (4.018)	2.820 (2.966)	2.531 (2.510)	2.257 (1.882)
$\sigma^2 = 0.5$	0.569 (0.308)	0.529 (0.238)	0.528 (0.161)	0.513 (0.107)
PTC with IG frailty and F_η a Weibull distribution				
$\beta_0 = 2$	2.556 (1.461)	2.385 (1.227)	2.312 (1.018)	2.107 (0.496)
$\beta_1 = -1$	−1.079 (0.234)	−1.057 (0.210)	−1.050 (0.167)	−1.018 (0.122)
$\beta_2 = 2$	2.134 (0.447)	2.110 (0.407)	2.089 (0.330)	2.040 (0.231)
$\lambda = 2$	2.228 (1.346)	2.181 (1.044)	2.096 (0.806)	2.046 (0.558)
$\nu = 1$	1.060 (0.168)	1.051 (0.159)	1.042 (0.130)	1.016 (0.090)
$\zeta = 0.5$	1.638 (3.506)	1.554 (3.553)	1.213 (2.284)	0.768 (1.261)
PTC with IG frailty and F_η a LFR distribution				
$\beta_0 = 2$	2.344 (0.973)	2.226 (0.703)	2.183 (0.497)	2.106 (0.332)
$\beta_1 = -1$	−1.058 (0.184)	−1.035 (0.150)	−1.023 (0.105)	−1.012 (0.077)
$\beta_2 = 2$	2.095 (0.337)	2.069 (0.278)	2.037 (0.196)	2.028 (0.139)
$a = 1.5$	1.362 (0.816)	1.382 (0.673)	1.374 (0.551)	1.404 (0.402)
$b = 2$	3.459 (4.147)	3.052 (3.301)	2.585 (2.561)	2.350 (2.054)
$\zeta = 0.5$	0.927 (1.549)	0.886 (2.742)	0.637 (0.503)	0.567 (0.293)

Table 2. Simulation results with 30% cure fraction and 45% censoring.

Parameter	$n = 200$	$n = 300$	$n = 500$	$n = 1000$
PTC with gamma frailty and F_η a Weibull distribution				
$\beta_0 = 0.5$	0.789 (0.887)	0.686 (0.755)	0.581 (0.409)	0.546 (0.254)
$\beta_1 = -1$	−1.067 (0.245)	−1.023 (0.189)	−1.022 (0.141)	−1.010 (0.102)
$\beta_2 = 2$	2.097 (0.426)	2.038 (0.337)	2.020 (0.247)	2.021 (0.179)
$\lambda = 2$	2.153 (1.188)	2.078 (0.982)	2.077 (0.751)	1.999 (0.533)
$\nu = 1$	1.039 (0.135)	1.020 (0.109)	1.013 (0.079)	1.005 (0.055)
$\sigma^2 = 0.5$	0.595 (0.480)	0.530 (0.375)	0.525 (0.273)	0.517 (0.182)
PTC with gamma frailty and F_η a LFR distribution				
$\beta_0 = 0.5$	0.812 (0.874)	0.695 (0.591)	0.618 (0.395)	0.579 (0.262)
$\beta_1 = -1$	−1.092 (0.264)	−1.039 (0.200)	−1.030 (0.142)	−1.014 (0.098)
$\beta_2 = 2$	2.158 (0.494)	2.073 (0.361)	2.038 (0.252)	2.029 (0.175)
$a = 1.5$	1.328 (0.830)	1.344 (0.669)	1.399 (0.518)	1.409 (0.370)
$b = 2$	3.045 (3.104)	2.705 (2.683)	2.476 (2.226)	2.145 (1.645)
$\sigma^2 = 0.5$	0.671 (0.578)	0.572 (0.401)	0.548 (0.280)	0.526 (0.174)
PTC with IG frailty and F_η a Weibull distribution				
$\beta_0 = 0.5$	0.958 (1.203)	0.802 (0.950)	0.746 (0.791)	0.623 (0.433)
$\beta_1 = -1$	−1.095 (0.271)	−1.062 (0.241)	−1.054 (0.205)	−1.033 (0.146)
$\beta_2 = 2$	2.164 (0.51)	2.135 (0.466)	2.108 (0.383)	2.069 (0.286)
$\lambda = 2$	2.202 (1.254)	2.206 (1.014)	2.131 (0.771)	2.018 (0.521)
$\nu = 1$	1.070 (0.179)	1.060 (0.166)	1.050 (0.137)	1.023 (0.100)
$\zeta = 0.5$	2.139 (5.047)	1.988 (4.589)	1.544 (3.243)	1.045 (2.389)
PTC with IG frailty and F_η a LFR distribution				
$\beta_0 = 0.5$	0.932 (0.970)	0.829 (0.970)	0.690 (0.488)	0.628 (0.341)
$\beta_1 = -1$	−1.098 (0.266)	−1.068 (0.235)	−1.038 (0.170)	−1.028 (0.120)
$\beta_2 = 2$	2.168 (0.483)	2.142 (0.426)	2.080 (0.318)	2.058 (0.238)
$a = 1.5$	1.256 (0.795)	1.298 (0.684)	1.350 (0.558)	1.374 (0.412)
$b = 2$	3.138 (3.305)	2.888 (2.833)	2.584 (2.381)	2.248 (1.888)
$\zeta = 0.5$	5.661 (70.567)	2.762 (20.101)	1.108 (2.504)	0.781 (1.191)

5. Application

In this section, we analyse whether the proposed models, as described in Section 3, are able to account for common unobservable factors by fitting them to two datasets. Before fitting the frailty models to a real-world dataset, we apply the models to a simulated dataset, where we simulate an "unobservable" variable not directly accounted for in our frailty models. We then proceed to apply the proposed models to a real-world loss dataset obtained from Global Credit Data (GCD).

5.1. Simulated Dataset

In order to further investigate the performance of the PTC model with gamma frailty, we simulate a single dataset $(Y_j, \delta_j, x_{j1}, x_{j2}), j = 1, \ldots, n$, of size $n = 1000$ from a PTC model without frailty, i.e., from the model in (4). The two covariates are generated as in Section 4.2, whereas the cure proportion is now given by $p_j = e^{-\theta_{\boldsymbol{\beta}}(\mathbf{x}_j)}$, where

$$\theta_{\boldsymbol{\beta}}(\mathbf{x}_j) = e^{\boldsymbol{\beta}^T \mathbf{x}_j} = e^{\beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2}}$$

with $\boldsymbol{\beta} = (2, -1, 2)^T$. $F_{\boldsymbol{\eta}}$ is taken to be the Weibull distribution, with the same parameters as in Section 4.2 and censoring times are again assumed to follow a standard uniform distribution. With this configuration, our simulated dataset has a non-susceptible percentage of 6.4%, and 16.2% of the resolution times are censored.

One of the covariates is now dropped from the simulated dataset so that we only "observe" the set $(Y_j, \delta_j, x_{j1}), j = 1, \ldots, n$. On this observed dataset, we fit three different models:

- The PTC model with no frailty in (4);
- The PTC model in (8) with gamma frailty and frailty parameter σ^2;
- The PTC model in (8) with an IG frailty and frailty parameter ζ.

For all three models, we have

$$\theta_{\boldsymbol{\beta}}(\mathbf{x}_j) = e^{\boldsymbol{\beta}^T \mathbf{x}_j} = e^{\beta_0 + \beta_1 x_{j1}}.$$

In Table 3, the estimates of all three models are shown, as well as their respective calculated log likelihoods, Akaike information criteria (AIC) and Bayesian information criteria (BIC).

Table 3. Parameter estimates and goodness-of-fit measures for all three models fitted to the simulated dataset.

Parameter	PTC	PTC with Gamma Frailty	PTC with IG Frailty
β_0	1.724	3.695	3.603
β_1	−0.629	−0.954	−0.978
λ	3.513	1.589	2.778
k	0.790	0.967	1.025
σ^2		0.825 *	
ζ			2.244 *
log-likelihood	1485.493	1493.001	1496.808
AIC	−2962.986	−2976.002	−2983.616
BIC	−2943.355	−2951.463	−2959.077

* These parameters are significant at a 5% significance level according to the adjusted likelihood ratio test (see Claeskens et al. [39]).

Based on both the AIC and BIC, the two frailty models outperform the PTC model, with the IG frailty yielding the best performance. The estimates of the β_0 and β_1 parameters for the two frailty models are very similar but differ notably from the estimate for the PTC model. This highlights the idiosyncratic attributes of the frailty models compared to the PTC model and, in conjunction with the lower AIC and BIC values, suggests that the frailty models are better at capturing the effect of the unobserved covariate than the PTC model.

5.2. Real-World Dataset

Our real-world loss dataset was provided by GCD, formally known as the Pan-European Credit Data Consortium (PECDC). GCD is a non-profit organisation that collects and analyses historical loss data from member banks [40].

We use a subsample of the loss database consisting of large corporations (LCs) from the United States. All loans that defaulted in the period 1 January 2005 to 31 December 2019 are included. For materiality purposes, we further filter the data to only include loans for which the exposure amount is larger than USD 500. The dataset includes information on when the loan went into default, whether (and when) the loan was fully settled, whether (and when) the loan was written off and whether the case has been resolved.

We include the industry of the obligor as a categorical covariate in the model based on the study by Schuermann [41], which demonstrates that the industry of an obligor has an impact on the recovery rate. In industries in which tangible assets represent a substantial portion of a company's assets, such as manufacturing or transportation, the bankruptcy process may be resolved more quickly, since the assets can be more easily sold to repay creditors.

Obligor industries are aggregated into five industry groups, as indicated in Table 4, with a sixth (denoted by -1) for loans for which the industry is unknown. Also shown in the rightmost column of Table 4 is the proportion of censored cases for each industry group. We consider defaulted loans that were fully repaid, as well as those for which the resolution process had not yet been completed at the end of the observation period, as censored.

Table 4. Summary of industry groupings, industry frequency and censored proportion of US large corporation dataset.

	Industry Group	Industry Description	Frequency	Censored
1	Primary Industries	Agriculture, Hunting and Forestry Fishing and Fishing Products Mining	14.5%	28.8%
2	Manufacturing, Utilities and Construction	Manufacturing Utilities Construction	33.3%	33.5%
3	Trade Industries	Wholesale and Retail Trade Transportation and Storage	24%	28.1%
4	Service Industries	Hotels and Restaurants Finance and Insurance Real Estate Rental and Leasing Professional, Scientific and Technical Services Private Sector Services (Household)	17%	27.9%
5	Other	Communications Public Administration and Defence Education Health and Social Services Other Community, Social and Personal Services	4.7%	35.6%
−1	Unknown	Unknown	6.5%	30.7%

In Figure 1, the Kaplan–Meier estimate of the survival probability of all data is shown on the left, and that of the data stratified by industry $(-1, 1, \ldots, 5)$ is shown on the right. From the left pane graph, it is evident that a cure proportion is present in the aggregated dataset. The narrow confidence interval can be attributed to the large sample size ($n = 1558$). From the right pane graph, notable differences in the cure proportions for the different industry groups can be observed, suggesting that the industry of a borrower has a definite impact on the probability of recovering the full loan amount after default.

Besides industry groups, which are included as dummy variables with corresponding coefficients $(\beta_2, \ldots, \beta_6)$, we include the annualised seasonally adjusted GDP growth in the month of default as a covariate, with the associated coefficient β_1. Our choice of GDP growth is motivated by the study of Khieu et al. [42], in which the authors report a positive relationship between annual GDP growth and recovery rates of defaulted corporate debt in the US. Furthermore, Betz et al. [43] found positive dependencies between default resolution times and final loan loss rates.

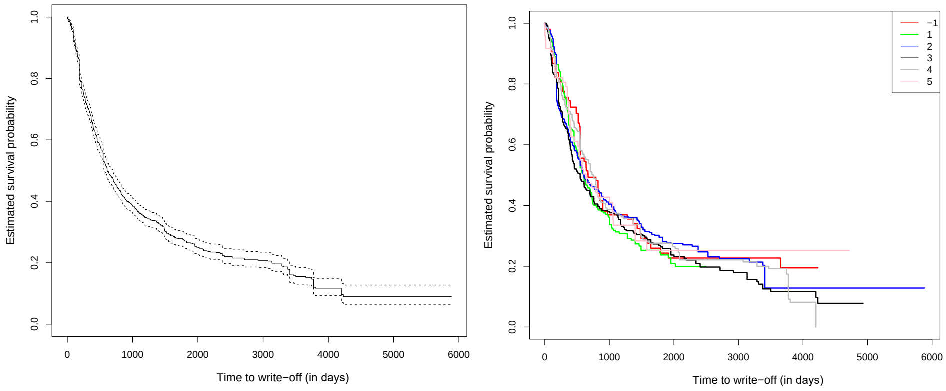

Figure 1. The Kaplan–Meier estimate of the survival probability of all data with 95% confidence intervals (**left** pane) and that of the data grouped by industry (**right** pane).

The MLEs of the model parameters are shown in the first three columns of Table 5. For both frailty models, the frailty parameters (σ^2 and ζ, respectively) are significant at the 5% level based on the adjusted log-likelihood test. In line with the findings of Khieu et al. [42] and Betz et al. [43], a negative relationship between GDP growth and time to write-off is observed. This means that higher GDP growth at the time of default is associated with a higher probability of full loan repayment. This makes intuitive sense, especially when considering the price at which collateral can be sold or the prospects for a firm to recover from financial distress and settle outstanding loan obligations. Also shown in Table 5 are the calculated log likelihoods, AIC and BIC. Based on all three goodness-of-fit measures, the PTC model with gamma frailty performs the best out of the three models.

Table 5. Estimation results of different models on the US large corporation dataset.

Parameter	PTC	PTC with Gamma Frailty	PTC with IG Frailty	PTC with Shared Frailty
β_0	3.920	6.968	2.141	0.843
β_1	−0.025	−0.021	−0.029	−0.023
β_2	0.187	0.094	0.134	
β_3	0.037	0.254	0.083	
β_4	0.175	0.501	0.297	
β_5	0.050	0.064	0.093	
β_6	0.076	0.168	0.117	
λ	4.462×10^{-4}	4.644×10^{-8}	2.523×10^{-5}	3.343×10^{-4}
k	0.852	1.579	1.327	1.073
σ^2		2.157 *		3.926
ζ			2.360 *	
log-likelihood	−8813.626	−8735.423	−8753.194	−8787.344
AIC	17,645.200	17,490.850	17,526.390	17,584.690
BIC	17,693.412	17,544.358	17,579.900	17,611.444

* These parameters are significant at the 5% level according to the adjusted likelihood ratio test.

In Figure 2, the survival curves (left pane) and hazard rates (right pane) for the PTC model, as well as the PTC with gamma and IG frailties, are plotted. From the left pane graph, it is evident that the fitted PTC model exhibits a much smaller cure fraction than the two frailty models, for which the frailty parameters were found to be significant. This further motivates the inclusion of a frailty parameter within the PTC model to capture residual heterogeneity. We observe that the hazard functions of the frailty models have very different shapes when compared to the standard PTC model. The hazard function has a significant impact on the survival function, especially on the cure rate, as can be seen from the graph on the left-hand side. The hazard functions of the two frailty models both exhibit a non-monotone shape. By including a frailty parameter in the model, the hazard function can become more flexible and may better capture non-standard patterns. Consequently, due to the flexibility of the frailty parameter, possibly less reliance may be placed on the choice of F_η.

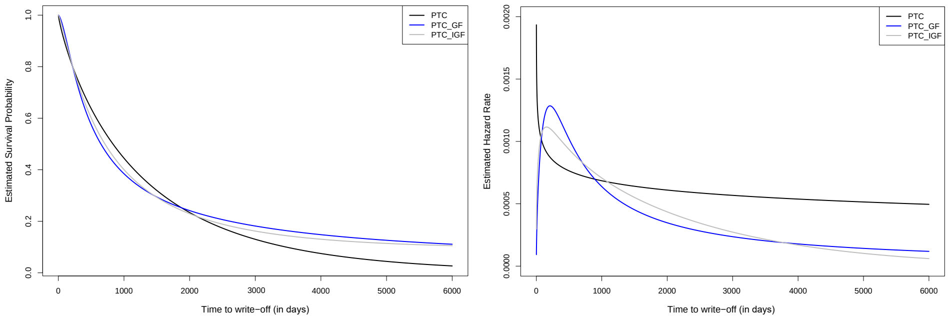

Figure 2. The estimated survival probability (**left** pane) and hazard rates (**right** pane) for PTC (black), PTC with gamma frailty (blue) and PTC with IG frailty (grey) models.

6. Extension to the Shared Gamma Frailty Model

In many survival studies, the failure times of individuals are not independent but dependent within groups or clusters. This can be due to shared environmental factors, such as exposure to air pollution in the case of cancer data or the state of the economy in the case of credit risk. By including a shared frailty term in a survival model, the dependence between the failure times of individual subjects within a cluster can be allowed for, leading to more accurate estimates of the hazard rates and the survival probabilities.

Suppose that we have n clusters of size n_i, $i = 1, 2, \ldots, n$, with each cluster sharing a common risk, assumed to be random, and modelled by an i.i.d. gamma distributed frailty W_i, $i = 1, \ldots, n$, each with a density of f_{σ^2} given in (7). The multivariate conditional survival function, given $W_i = w_i$, is then

$$\begin{aligned}\tilde{S}_{\boldsymbol{\eta}}(t_{i1}, \ldots, t_{in_i} \mid \mathbf{x}_i,\, w_i) &= P\big(T_{i1} > t_{i1}, \ldots, T_{in_i} > t_{in_i} \mid \mathbf{x}_i,\, W_i = w_i\big) \\ &= P\big(T_{i1} > t_{i1} \mid \mathbf{x}_i,\, W_i = w_i\big) \ldots P\big(T_{in_i} > t_{in_i} \mid \mathbf{x}_i,\, W_i = w_i\big) \\ &= e^{-w_i \sum_{j=1}^{n_i} \theta_{\boldsymbol{\beta}}(\mathbf{x}_{ij}) F_{\boldsymbol{\eta}}(t_{ij})},\end{aligned}$$

where $\mathbf{x}_i = \big(\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i}\big)$ is the covariate matrix of the individual in the ith cluster, and $F_{\boldsymbol{\eta}}$ is, as in the univariate case, some specified distribution with parameter vector $\boldsymbol{\eta}$.

For the jth individual in the ith cluster, $j = 1, 2, \ldots, n_i$, we observe the triplet $(Y_{ij}, \delta_{ij}, \mathbf{x}_{ij})$, where $Y_{ij} = \min(T_{ij}, C_{ij})$, $\delta_{ij} = I\big(T_{ij} \le C_{ij}\big)$, and $\mathbf{x}_{ij} = (1, x_{1ij}, x_{2ij}, \ldots, x_{mij})^T$ is the vector of m covariates associated with the jth individual in the ith cluster. Similar to the univariate case, $\theta_{\boldsymbol{\beta}}(\mathbf{x}_{ij}) = e^{\boldsymbol{\beta}^T \mathbf{x}_{ij}}$, where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_m)^T$ is a vector of unknown regression parameters. Furthermore, we denote the number of observed events in the ith cluster by $d_i = \sum_{j=1}^{n_i} \delta_{ij}$. The likelihood is then given by

$$L = \prod_{i=1}^{n} \int_0^{\infty} \left(\prod_{j=1}^{n_i} \big(w_i \theta_{\boldsymbol{\beta}}(\mathbf{x}_{ij}) f_{\boldsymbol{\eta}}(Y_{ij})\big)^{\delta_{ij}} e^{-w_i \theta_{\boldsymbol{\beta}}(\mathbf{x}_{ij}) F_{\boldsymbol{\eta}}(Y_{ij})} \right) f_{\sigma^2}(w_i) \mathrm{d}w_i,$$

where $f_{\boldsymbol{\eta}}$ is the density associated with distribution $F_{\boldsymbol{\eta}}$.

By some algebraic manipulation, and by taking the natural logarithm, the log-likelihood function in closed form is given by

$$\begin{aligned}\ell(\boldsymbol{\beta}, \sigma, \boldsymbol{\eta}) = &\sum_{i=1}^{n} \left(\sum_{j=1}^{n_i} \delta_{ij} \log\{\theta_{\boldsymbol{\beta}}(\mathbf{x}_{ij}) f_{\boldsymbol{\eta}}(Y_{ij})\} \right) + \sum_{i=1}^{n} \log \Gamma\left(\frac{1}{\sigma^2} + d_i\right) \\ &- \sum_{i=1}^{n} \left(\frac{1}{\sigma^2} + d_i\right) \log\left(\frac{1}{\sigma^2} + \sum_{j=1}^{n_i} \theta_{\boldsymbol{\beta}}(\mathbf{x}_{ij}) F_{\boldsymbol{\eta}}(Y_{ij})\right) \\ &- \frac{2n}{\sigma^2} \log \sigma - n \log \Gamma\left(\frac{1}{\sigma^2}\right).\end{aligned}$$

For the application of the shared frailty model to our real-world dataset described in Section 5.2, we cluster our data with reference to industry group based on the intuitive reasoning that obligors within the same industry group are subject to similar forces in the economy, which is a source of dependency within these groups. Instead of treating the industry group as a covariate, a shared frailty parameter is fitted as a means to account for the dependence of write-off times within industry groups. Again, we assume $F_{\boldsymbol{\eta}}$ to be the Weibull distribution function with $\boldsymbol{\eta} = (\lambda, k)$ as defined in (15).

The final column of Table 5 lists the MLEs of the PTC model with shared frailty. It is important to note that the significant frailty variance parameter σ^2 of the univariate gamma frailty model confirms that the model succeeds in capturing unobserved heterogeneity. However, for the shared frailty model, a significant frailty variance parameter σ^2 confirms that a relationship exists within each industry cluster and that its value can be interpreted to measure the dependence between loan write-off times within each industry cluster (see discussion in [32] (p. 144)).

According to the values of the calculated log likelihood, AIC and BIC, the shared frailty model performs favourably compared to the standard PTC model, despite the shared frailty model being a more parsimonious model. This also suggests that the shared frailty component does succeed in capturing some of the dependence between loan write-off times attributable to the industry of the obligor.

7. Concluding Remarks

In this paper, a PTC model with gamma frailty was developed and applied to a real-world loss dataset. The applicability of a cure proportion in the context of the post default loan resolution process was demonstrated, and we also showed how the inclusion of a frailty parameter in the PTC model can allow for the modelling of a more flexible shape of the hazard function. This may improve the accuracy of the survival model and, equally importantly, provide a better understanding of the underlying processes that drive the occurrence of write-off over time. Furthermore, the flexibility offered by the inclusion of a frailty parameter affords for possibly less reliance to be placed on the selection of a "baseline" distribution $F_{\boldsymbol{\eta}}$ in the PTC model.

The proportional hazard assumption of the popular CPH model implies that the relative hazard rate remains constant over time with different covariate levels. By including a parametric frailty parameter in the PTC model with fewer covariates, a more flexible relationship between potential (latent) predictors and the time to write-off can be accommodated. Through a Monte Carlo study, we demonstrated that the parameter estimates of the PTC model with gamma frailty are reliable.

We also showed that, by including a shared gamma frailty term in a PTC model, the dependence between the failure times of individual subjects within a cluster can be accounted for, which, in turn, may lead to more accurate estimates of the hazard rate and the write-off times. Our initial exploration of the use of a shared frailty model warrants future investigation, with specific focus on the small sample performance of shared frailty models. Other PTC models with shared frailties can also be considered, such as a log-normal shared frailty model or a compound Poisson shared frailty model that includes two important special cases, namely gamma and IG distributions.

A shared frailty model takes the dependency of the event times within clusters into account, but other forms of dependency may also exist that were not explored in this study. The models we considered assume independence between the censoring time and write-off time of individual loans. A model that allows for possible dependence between these times may be considered in a future study. This is referred to as dependent censoring and is rapidly gaining traction in contemporary research (see, e.g., [44–46]).

In a highly regulated industry such as banking, parametric models are often preferred to non-parametric models due to their analytical tractability and interpretability. Parametric models also easily allow for sensitivity analysis and stress testing by varying parameter values. As demonstrated by this study, parametric frailty models offer the potential to characterise unexplained heterogeneity in the write-off times of individual loans. Besides the time to write-off, which can be predicted more accurately, these models also offer the prospect of modelling the probability of write-off and full repayment in the context of censored data. As such, the model offers a viable alternative for modelling the resolution outcome in two-step loss, given default modelling, while making use of the most recent (and therefore most relevant) loan resolution data. The reliance of these models on accurate and trustworthy data requires banks to make substantial investments in systems, expertise and validation efforts. Modellers should take care to avoid spurious accuracy and to understand the economic and operational drivers of credit risk before and after loan default. Inclusion of a frailty parameter to capture residual heterogeneity should therefore not be regarded as a substitute for including known risk drivers as covariates within a prediction model.

Author Contributions: Conceptualization, J.L., J.S.A., G.L.G. and M.S.; Methodology, J.L., J.S.A., G.L.G. and M.S.; Software, M.S.; Validation, J.L., J.S.A. and G.L.G.; Formal analysis, J.S.A. and G.L.G.; Investigation, J.L.; Resources, J.L.; Writing—original draft, J.L.; Writing—review & editing, J.L., J.S.A., G.L.G. and M.S.; Visualization, J.S.A. and G.L.G.; Supervision, J.S.A., G.L.G. and M.S.; Project administration, J.L. The authors all contributed equally. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset presented in this study is available on application from Global Credit Data. https://globalcreditdata.org/.

Acknowledgments: The work of J.S. Allison is based on research supported by the National Research Foundation (NRF). Any opinion, finding, conclusion or recommendation expressed in this material is that of the authors, and the NRF does not accept any liability in this regard.

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

References

1. Botha, A.; Beyers, C.; De Villiers, P. Simulation-based optimisation of the timing of loan recovery across different portfolios. *Expert Syst. Appl.* **2021**, *177*, 114878. [CrossRef]
2. Fenech, J.P.; Yap, Y.K.; Shafik, S. Modelling the recovery outcomes for defaulted loans: A survival analysis approach. *Econ. Lett.* **2016**, *145*, 79–82. [CrossRef]
3. Betz, J.; Krüger, S.; Kellner, R.; Rösch, D. Macroeconomic effects and frailties in the resolution of non-performing loans. *J. Bank. Financ.* **2017**, *112*, 105212. [CrossRef]
4. Malwandla, M.C.; Marimo, M.; Breed, D.G. A cross-sectional survival analysis regression model with applications to consumer credit risk. *S. Afr. Stat. J.* **2017**, *51*, 217–234.
5. de Oliveira, M., Jr.; Louzada, F. Recovery Risk: Application of the Latent Competing Risks Model to Non-Performing Loans. *arXiv* **2014**, arXiv:1408.4380.
6. Hibbeln, M.; Gürtler, M. Pitfalls in Modeling Loss Given Default of Bank Loans. *J. Bank. Financ.* **2013**, *37*, 2354–2366.
7. Allen, L.N.; Rose, L.C. Financial survival analysis of defaulted debtors. *J. Oper. Res. Soc.* **2006**, *57*, 630–636. [CrossRef]
8. Joubert, M.; Verster, T.; Raubenheimer, H.G. Making use of survival analysis to indirectly model loss given default. *ORiON* **2019**, *34*, 107–132. [CrossRef]
9. Tong, E.N.C.; Mues, C.; Thomas, L.C. Mixture cure models in credit scoring: If and when borrowers default. *Eur. J. Oper. Res.* **2012**, *218*, 132–139. [CrossRef]
10. Smuts, M.; Allison, J.S. An overview of survival analysis with an application in the credit risk environment. *ORiON* **2021**, *36*, 89–110. [CrossRef]
11. Dirick, L.; Bellotti, T.; Claeskens, G.; Baesens, B. Macro-Economic Factors in Credit Risk Calculations: Including Time-Varying Covariates in Mixture Cure Models. *J. Bus. Econ. Stat.* **2019**, *37*, 40–53. [CrossRef]
12. de Oliveira, M., Jr.; Moreira, F.; Louzada, F. The zero-inflated promotion cure rate model applied to financial data on time-to-default. *Cogent Econ. Financ.* **2017**, *5*, 1395950. [CrossRef]
13. Barriga, G.D.C.; Cancho, V.G.; Garibay, D.V.; Cordeiro, G.M.; Ortega, E.M. A new survival model with surviving fraction: An application to colorectal cancer data. *Stat. Methods Med Res.* **2018**, *28*, 2665–2680. [CrossRef] [PubMed]
14. Leow, M.; Mues, C. Predicting loss given default (LGD) for residential mortgage loans: A two-stage model and empirical evidence for UK bank data. *Int. J. Forecast.* **2012**, *28*, 183–195. [CrossRef]

15. Cox, D.R. Regression models and life-tables. *J. R. Stat. Soc. Ser. B (Methodol.)* **1972**, *34*, 187–202. [CrossRef]
16. Banasik, J.; Crook, J.N.; Thomas, L.C. Not if but when will borrowers default. *J. Oper. Res. Soc.* **1999**, *50*, 1185–1190. [CrossRef]
17. Stepanova, M.; Thomas, L. Survival analysis methods for personal loan data. *Oper. Res.* **2002**, *50*, 277–289. [CrossRef]
18. Cao, R.; Vilar, J.M.; Devia, A. Modelling consumer credit risk via survival analysis. *Sort-Stat. Oper. Res. Trans.* **2009**, *33*, 3–30.
19. Boag, J.W. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J. R. Stat. Society. Ser. B (Methodol.)* **1949**, *11*, 15–53. [CrossRef]
20. Peng, Y.; Yu, B. *Cure Models: Methods, Applications, and Implementation*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2021.
21. Berkson, J.; Gage, R.P. Survival curve for cancer patients following treatment. *J. Am. Stat. Assoc.* **1952**, *47*, 501–515. [CrossRef]
22. Farewell, V.T. The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* **1982**, *38*, 1041–1046. [CrossRef] [PubMed]
23. Amico, M.; van Keilegom, I. Cure models in survival analysis. *Annu. Rev. Stat. Appl.* **2018**, *5*, 311–342. [CrossRef]
24. Yakovlev, A.Y.; Asselain, B.; Bardou, V.; Fourquet, A.; Hoang, T.; Rochefediere, A.; Tsodikov, A. A simple stochastic model of tumor recurrence and its application to data on premenopausal breast cancer. *Biom. Anal. Donnees Spatio-Temporelles* **1993**, *12*, 66–82.
25. Chen, M.H.; Ibrahim, J.G.; Sinha, D. A New Bayesian Model For Survival Data with a Surviving Fraction. *J. Am. Stat. Assoc.* **1999**, *94*, 909–919. [CrossRef]
26. Rodrigues, J.; Cancho, V.G.; de Castro, M.; Louzada-Neto, F. On the unification of long-term survival models. *Stat. Probab. Lett.* **2009**, *79*, 753–759. [CrossRef]
27. Zhang, J.; Peng, Y. A new estimation method for the semiparametric accelerated failure time mixture cure model. *Stat. Med.* **2007**, *26*, 3157–3171. [CrossRef]
28. Portier, F.; El Ghouch, A.; van Keilegom, I. Efficiency and bootstrap in the promotion time cure model. *Bernoulli* **2017**, 23, 3437–3468. [CrossRef]
29. Delloye, M.; Fermanian, J.D.; Sbai, M. Dynamic frailties and credit portfolio modelling. *Risk* **2006**, *19*, 100.
30. Chih-Wei, L.; Chang, M.J. A credit risk model with dynamic frailties for default intensity estimation. *Asia Pac. Manag. Rev.* **2008**, *13*, 557–566.
31. Chamboko, R.; Bravo, J.M. Frailty correlated default on retail consumer loans in Zimbabwe. *Int. J. Appl. Decis. Sci.* **2019**, *12*, 257–270. [CrossRef]
32. Wienke, A. *Frailty Models in Survival Analysis*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2010; pp. 57–63.
33. Balan, T.A.; Putter, H. A tutorial on frailty models. *Stat. Methods Med Res.* **2020**, *29*, 3424–3454. [CrossRef] [PubMed]
34. Li, C.S.; Taylor, J.M.; Sy, J.P. Identifiability of cure models. *Stat. Probab. Lett.* **2001**, *54*, 389–395. [CrossRef]
35. Legrand, C. *Advanced Survival Models*; CRC Press: Boca Raton, FL, USA, 2021.
36. Dirick, L.; Claeskens, G.; Baesens, B. An Akaike information criterion for multiple event mixture cure models. *Eur. J. Oper. Res.* **2015**, *241*, 449–457. [CrossRef]
37. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022.
38. Efron, B.; Tibshirani, R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.* **1986**, *1*, 54–75. [CrossRef]
39. Claeskens, G.; Nguti, R.; Janssen, P. One-sided tests in shared frailty models. *Test* **2008**, *17*, 69–82. [CrossRef]
40. Brumma, N.; Winckle, P. LGD Report 2018-Large Corporate Borrowers. *SSRN* **2018**. [CrossRef]
41. Schuermann, T. What do we know about Loss Given Default? Wharton Financial Institutions Center Working Paper. *SSRN* **2004**. [CrossRef]
42. Khieu, H.D.; Mullineaux, D.J.; Yi, H.C. The determinants of bank loan recovery rates. *J. Bank. Financ.* **2012**, *36*, 923–933. [CrossRef]
43. Betz, J.; Kellner, R.; Rösch, D. Time matters: How default resolution times impact final loss rates. *J. R. Stat. Soc. Ser. C* **2021**, *70*, 619–644. [CrossRef]
44. Emura, T.; Chen, Y.H. *Analysis of Survival Data with Dependent Censoring: Copula-Based Approaches*; Springer: Berlin/Heidelberg, Germany, 2018.
45. Deresa, N.W.; van Keilegom, I. A multivariate normal regression model for survival data subject to different types of dependent censoring. *Comput. Stat. Data Anal.* **2020**, *144*, 106879. [CrossRef]
46. Deresa, N.W.; van Keilegom, I.; Antonio, K. Copula-based inference for bivariate survival data with left truncation and dependent censoring. *Insur. Math. Econ.* **2022**, *107*, 1–21. [CrossRef]

Chapter 5

Conclusion

In this Chapter the main findings, including avenues for future research, of the three thesis articles are reported and some concluding remarks are provided.

5.1 Overview of results and future research

The general objective of this thesis was to contribute to the literature on LGD modelling, specifically in the context of EC models. Three articles have been presented, each making a unique contribution to distinct aspects of LGD modelling. The results and conclusions of these three articles, as well as avenues for future research, are now presented.

In the first article, ***A Cost of Capital Approach to determining the LGD Discount Rate***, we have demonstrated how an LGD discount rate can be inferred from a market-consistent (or fair) price for defaulted exposures. Here, determination of the market consistent price has been inspired by the Solvency II regulatory regime's cost of capital approach. The main drivers of the LGD discount rate were found to be the mean and variance of recovery cash flows, which in turn drive the economic capital requirement for post default credit risk. Further drivers of the discount rate include the average risk free rate applicable over the reference period, the average term of the cash flows, and the cost of capital rate. We also demonstrated that the discount rate is fairly sensitive to the cost of capital rate, more so when recovery rates are low. The proposed methodology satisfies the regulatory requirement for the LGD discount rate to account for the time value of money and to include a risk premium that reflects the undiversifiable risk of recovery cash flows. It therefore represents a viable option to determine an LGD discount rate for both regulatory and economic capital purposes.

Due to data limitations, a comparative study between the proposed cost of capital discount rate and the average contract rate (of the two retail bank portfolios analysed) was excluded from the scope of our study. To provide a more comprehensive analysis of the proposed method, future research could consider such a comparative study to fully explore the difference between the cost of capital risk premium and that embedded in the contractual loan rate.

In the second article, ***Introducing Two Parsimonious Standard Power Mixture Models for Bimodal Proportional Data with Application to Loss Given Default***, two alternative models to the commonly used beta distribution have been proposed to better capture the bimodality of LGD data. Besides parsimony, the objective was to formulate a model that reduces the difference between the empirical and distributional means. The two proposed models, both mixtures of standard power distributions, were fitted to LGD data and their performance evaluated using traditional information criteria. In addition to outperforming the benchmark beta distribution, both proposed models also reduced the bias in the estimation of the mean. It was further demonstrated how, by evaluating the sign of the empirical bias of the fitted models, a model that provides a more conservative estimation of the mean LGD could be identified.

The beta distribution is a popular choice to model the dependent variable in regression analysis, and specifically for fractional data. The alternative mixture models proposed in this study may therefore be employed to more accurately characterise the target variable where this is U- or J-shaped, and future research could be directed to such an investigation.

In the third article, ***Modelling the Time to Write-Off of Non-Performing Loans using a Promotion Time Cure Model with Parametric Frailty***, an improvement to existing survival modelling approaches was proposed, by including a frailty component in the promotion time cure model to account for common unobservable factors, and for possible observable covariates not in the model. Both gamma and inverse Gaussian distributed frailty parameters were considered, as well as a shared frailty model. We demonstrated the reliability of the parameter estimates with a Monte Carlo simulation study, after which the models were applied to a simulated data set as well as a real world loan loss data set, provided by Global Credit Data. The gamma frailty model produced the best fit to our data and we illustrated how the flexibility of the frailty parameter can allow for the modelling of a non-monotone shape of the hazard function. Finally, it was demonstrated how, by including a shared frailty term in a survival model, the dependence between the failure times of individual subjects within a cluster can be allowed for, which in turn may lead to more accurate estimates of the hazard rates and the write-off times. The proposed model offers an attractive alternative to modelling the probability of write-off in two-step LGD models.

Our initial exploration into the use of a shared frailty model has shown promising results, but further investigation is needed to fully understand its small sample performance. Additionally, future studies may consider other promotion time cure models with shared frailties and could also explore models that allow for possible dependence between write-off and censoring times.

5.2 Concluding remarks

The increasing attention LGD modelling is receiving, from academics and practitioners alike, reflects its importance with regards to the assessment of credit risk. The three articles of this thesis should make a meaningful contribution to the extensive body of literature in this domain.

Given the rapid inroads that machine learning models are making, and the overall increased sophistication in risk modelling, LGD predictions may soon be even more accurate. However, considering the intense regulatory scrutiny that capital models are typically subjected to, parametric modelling approaches should retain their rightful place in the arsenal of LGD modelling techniques, due to their analytical tractability and interpretability. Therefore, despite the extensive research in the field, there remains ample opportunities for further contributions and advancements.