MODELING LOSS GIVEN DEFAULT ON UNSECURED CONSUMER LOANS – CASE STUDY OF A FINNISH FINANCIAL INSTITUTION

Lappeenranta–Lahti University of Technology LUT

Master's Programme in Strategic Finance and Analytics, Master's thesis

2023

Joonas Rantanen

Supervisor/Examiner: Post-doctoral Researcher, Jyrki Savolainen

Examiner: Professor, Mikael Collan

ABSTRACT

Lappeenranta–Lahti University of Technology LUT

LUT School of Business and Management

Business Administration

Joonas Rantanen

Modeling loss given default on unsecured consumer loans – case study of a Finnish financial institution

Master's thesis

2023

49 pages, 4 figures and 3 tables

Supervisor/Examiner: Post-doctoral Researcher Jyrki Savolainen

Examiner: Professor Mikael Collan

Keywords: unsecured consumer loan, credit risk, loss given default, modeling

Given the impact of the Basel II Accord and IRB approach, financial institutions must better understand Loss Given Default (LGD) models. LGD models enable banks to more accurately estimate the potential losses that may result from a default, allowing them to price their loans appropriately and manage their risk exposure more effectively. This thesis explores the state-of-the-art methods for modeling loss given default on unsecured consumer loans and investigates how these methods are currently being applied in practice in the case of company x operating in Finland.

A comprehensive literature review found that the most used LGD modeling techniques are regression-based models. However, the choice of the most suitable method depends on the specific characteristics of the loan portfolio. The semi-structured interview with company x found that the company's LGD modeling approach was consistent with the literature. The company's regression-based model used customer demographics, credit behavior, and credit account application variables to predict LGD. However, the study also identified areas for improvement, including exploring more sophisticated methods for LGD estimation and addressing data representativeness more explicitly. The recommendations of this thesis can be generalized to recommendations for companies looking to improve their LGD modeling practices.

TIIVISTELMÄ

Lappeenrannan–Lahden teknillinen yliopisto LUT

LUT-kauppakorkeakoulu

Kauppatieteet

Joonas Rantanen

Vakuudettomien kulutusluottojen tappio-osuuden mallintaminen – tapaustutkimus suomalaisesta rahoituslaitoksesta

Kauppatieteiden pro gradu -tutkielma

2023

49 sivua, 4 kuvaa ja 3 taulukkoa

Tarkastajat: Tohtoritutkija Jyrki Savolainen ja Professori Mikael Collan

Avainsanat: vakuudeton kulutusluotto, luottoriski, tappio-osuus, mallinnus

Ottaen huomioon Basel II -sopimuksen ja IRB-lähestymistavan vaikutukset, rahoituslaitosten on ymmärrettävä paremmin LGD-malleja. LGD-mallit auttavat pankkeja arvioimaan velan takaisinmaksun laiminlyönnistä johtuvia tappioita. Tällöin pankit voivat hinnoitella lainansa asianmukaisesti ja hallita riskiä tehokkaammin. Tässä pro gradu tutkielmassa tarkastellaan nykyaikaisia menetelmiä vakuudettomien kulutusluottojen LGD mallintamiseen ja miten näitä menetelmiä sovelletaan tällä hetkellä käytännössä Suomessa toimivan yrityksen x tapauksessa.

Kattavassa kirjallisuuskatsauksessa havaittiin, että regressiopohjaiset mallit ovat eniten käytettyjä LGD-mallinnustekniikoita. Sopivimman menetelmän valinta riippuu kuitenkin lainasalkun erityispiirteistä. Puolistrukturoidussa haastattelussa yrityksen x kanssa havaittiin, että yrityksen LGD-mallinnustapa oli yhdenmukainen kirjallisuuden kanssa. Yhtiön regressiopohjainen malli käytti muun muassa asiakkaiden demografisia, luottokäyttäytymisen ja luottohakemuksen muuttujia ennustamaan LGD:tä. Tutkimuksessa tunnistettiin kuitenkin myös parannuskohteita, mukaan lukien kehittyneempien LGD-estimointimenetelmien tutkiminen ja datan edustavuuden parantaminen. Tämän pro gradu tutkielman suositukset voidaan yleistää suosituksiksi yrityksille, jotka haluavat parantaa LGD-mallinnuskäytäntöjään.

ACKNOWLEDGEMENTS

Better late than never. This trip has been longer than it originally should have been. Here I am writing acknowledgements, although it is still hard to believe. This whole writing process could have been finished before work, buying a house, getting married, and having the first child.

First, I would like to thank my supervisor Jyrki Savolainen, who helped me finish this project. I was able to do this with your guidance and patience. In addition, I would like to thank LUT's competent staff, who I had to bother with constant questions, especially in the latter half of the project.

However, the biggest thanks go to my wife and son. Thank you for continuing to believe that this day will come. Thank you for the evenings during which I could focus on writing and when I needed something else to think about. Without you, this would not have been possible. Thanks to my son, whose birth gave me new motivation to finish this process. Thanks also to my little brother for all the years together. Now you can rest.

In Kirkkonummi, 5th March 2023

Joonas Rantanen

Table of contents

1 Introduction

Loss given default (LGD) measures the economic loss, expressed as a percentage of the loan exposure, in case of default and plays a role in determining the expected loss on a loan portfolio and companies' long-term strategy. (Loterman et al., 2012, 161) This thesis studies different ways of modeling loss given default on unsecured consumer loans and how to use this information to select between available methods in LGD modeling. A financial company operating in Finland is used as a case study. In the literature, there aren't many studies revolving around different methods of modeling or using LGD models in Finnish financial institutions or case companies making this topic worth studying. Results can be used in real-life case scenarios in case company x or other financial institutions operating in the Finnish market.

1.1 Background and motivation

According to Loterman et al. (2012, 161), the Basel II Accord has had a significant impact on financial institutions by enabling them to construct credit risk models for three critical risk parameters: the probability of default (PD), LGD, and exposure at default (EAD). They continue that LGD has particularly focused in credit risk analysis from these parameters. They conclude that the credit risk research focus is changing from PD to estimation and validation of the LGD modeling. (Loterman et al., 2012)

As discussed in EBA (16/2017, 3), to guide financial institutions to reduce the unjustified variability of risk parameters and own funds requirements, the European Banking Authority (EBA) has initiated various efforts, including "*guidelines on PD estimation, LGD estimation, and the treatment of defaulted exposures*." These initiatives are part of a broader review of the Internal Ratings-Based (IRB) approach, as outlined in "*the Report on the Review of the IRB Approach,"* published in February 2016. (EBA, 16/2017, 3) The guidelines primarily emphasize the definitions and modeling techniques in estimating risk parameters for non-defaulted and defaulted exposures. At the same time, the guidelines continue how other regulatory products developed in the review process will clarify other aspects associated with the IRB approach. The guidelines have been applicable since 1 January 2021. (EBA, 16/2017)

Given the impact of the Basel II Accord and IRB approach, financial institutions must better understand LGD models. By understanding LGD models, banks can more accurately estimate the potential losses that may result from a default, enabling them to price their loans appropriately and manage their risk exposure more effectively.

1.2 Research questions

This thesis aims to comprehensively investigate the existing literature on LGD modeling to identify best practices for modeling unsecured consumer loans. The research will utilize a literature review and then a semi-structured interview with case company x to gain insights into how the company could improve its LGD modeling. The existing literature has extensively covered LGD modeling, but it is crucial to provide clear indications of best practices due to regulatory uncertainties. From the literature review findings, best practices for LGD modeling will be established. The thesis will focus on three key areas: LGD, unsecured consumer loans, and modeling, as displayed in Figure 1.

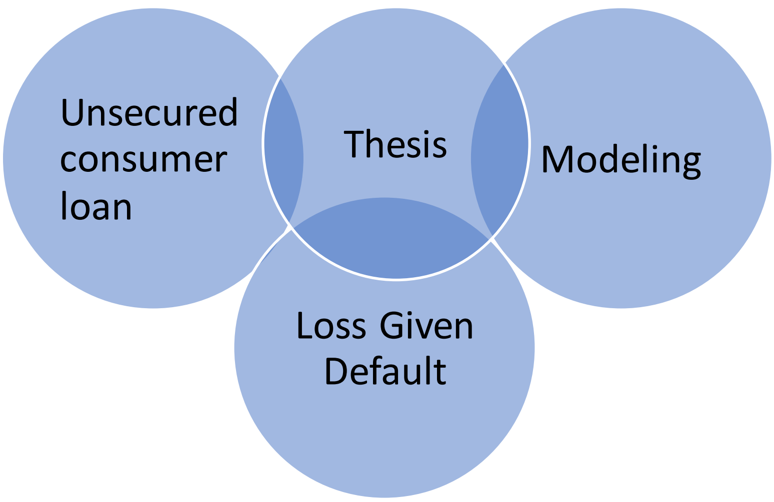

Figure 1. Focus of the thesis

The study will thoroughly review the literature to identify best practices for LGD modeling in an unsecured consumer loan environment. This information and insights from case company x will be used to identify critical areas for further development.

The objective of this thesis is to answer the following research questions:

1. Based on the existing literature, what are the state-of-the-art methods to model loss given default on unsecured consumer loans, and how are they currently applied in practice?

2. How is LGD modeling currently implemented in case company x and what would be the key points to develop it further?

Question 2 will be answered based on the findings from RQ1, as well as insights from the semi-structured interview with case company x. The study intends to contribute to the existing literature on LGD modeling in unsecured consumer loan environments and provide practical recommendations for companies to improve their LGD modeling practices.

1.3 Data and Methodology

This thesis is qualitative research by nature. The research approach is qualitative because it seeks to understand the field-level knowledge and real-world problems associated with LGD modeling in unsecured consumer loans rather than focusing solely on mathematical formulas. Two methods were utilized to answer the research questions: a literature review and a semi-structured interview.

The literature review was conducted to answer the first research question, which aimed to identify the state-of-the-art methods for modeling LGD in unsecured consumer loans and how they are currently applied in practice. The review process involved a systematic search of academic journals and industry reports. Relevant articles were identified and analyzed to comprehensively overview the current best practices in LGD modeling.

The second research question focused on how LGD modeling is currently done in case company x and what key points could be developed further. A semi-structured interview was conducted with key employees at case company x. The interview allowed for a deeper exploration of the company's LGD modeling practices and an opportunity to identify areas for improvement.

1.4 Structure of the research

This thesis is structured around six different chapters. A walkthrough of the content is introduced in Figure 2.

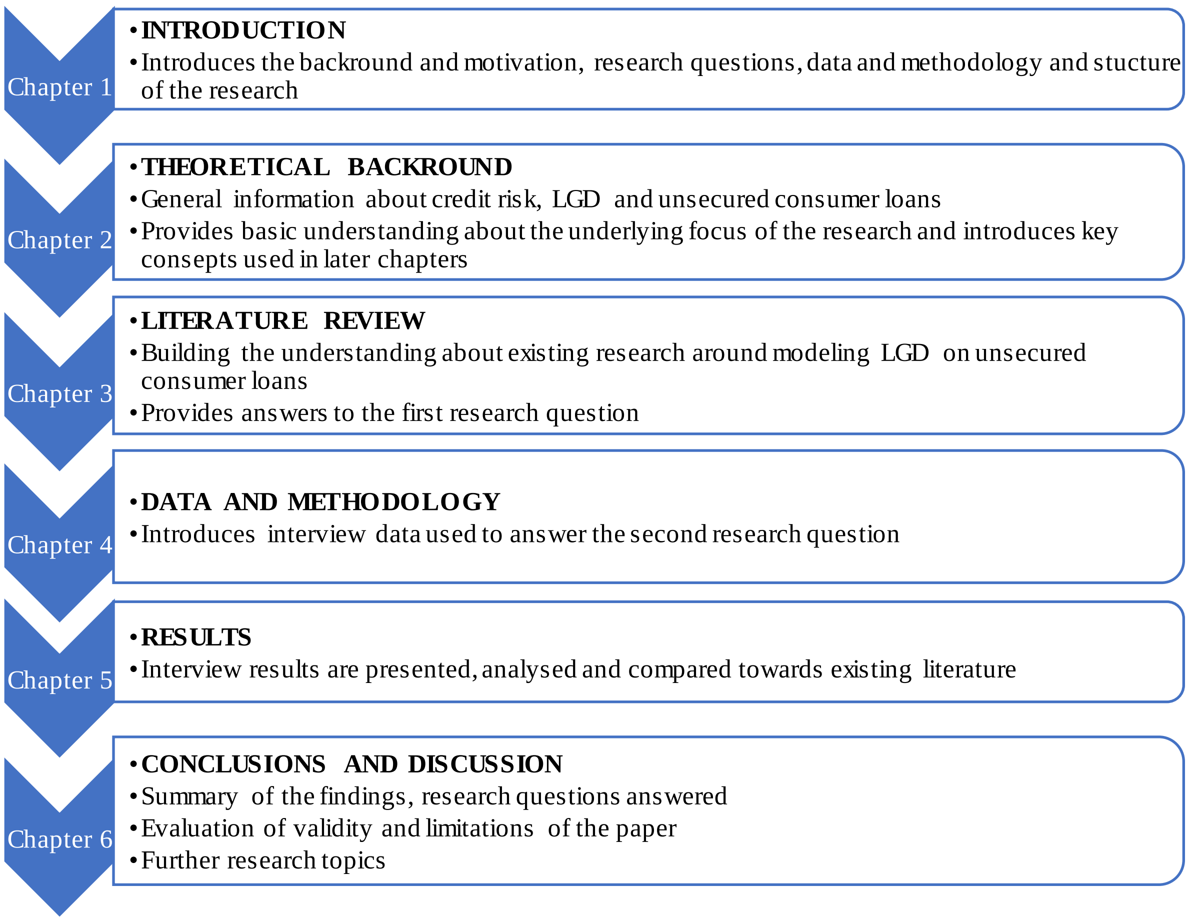

Figure 2. Structure of the Research

2 Theoretical background

A critical parameter that financial institutions consider when assessing credit risk is LGD, according to Loterman et al. (2012, 161), while continuing that LGD represents a borrower's expected loss in default. Kaposty F. et al. (2020, 248) see accurate LGD modeling as crucial for effective credit risk management and pricing of loans. They discuss how Basel II Accord has significantly impacted credit risk management, leading to a growing importance in the modeling and forecasting of LGD. They conclude how financial institutions that employ their own LGD estimations can adopt simpler or more complex approaches where simpler methods are easier to implement and comprehend, whereas complex techniques offer the potential for greater accuracy in predictions.

The following section presents the theoretical background for modeling LGD on unsecured consumer loans. The first two chapters focus on unsecured consumer loans, credit risk parameters, and default. The last eight chapters focus on LGD, modeling LGD, and areas needed for LGD modeling.

2.1 Unsecured consumer loans

When discussing consumer loans or credits, European Central Bank (ECB 2016, 1) defines consumer credit as *"loans granted mainly for personal consumption of goods and services"* in their bank lending survey. They continue listing typical examples of loans in this category, notably financing vehicles, domestic appliances, and other consumer durables. The survey concludes that loans included in the consumer category are with or without collateral by various forms of security or guarantee. When discussing consumer loans in this thesis, loans are without security or guarantee and are considered unsecured.

One of the main challenges in unsecured consumer loan lending is assessing credit risk. Amadi (2012, 22) discusses how, for unsecured consumer loans, consumers pay an extra premium on loan rates to offset the higher risk of delinquency and loan write-offs for financial institutions. Pickert (2017, 45-46) points out that credit risk is one of the most critical risks financial institutions granting unsecured consumer loans must face and manage.

2.2 Credit risk parameters and default

Credit risk refers to the inability or unwillingness of a customer or counterparty to fulfil their obligations in financial transactions, such as lending, trading, settlement, and hedging pointed out by Spuchľáková et al. (2015). Since Basel II, Hurlin et al. (2018) discuss how banks have been able to use internal rating models to calculate their credit risk capital charge through the IRB approach. They continue that this approach can be considered an external risk model based on four crucial risk parameters: EAD, PD, LGD, and maturity (M).

Default is seen by Frye (2004) as a situation *"when an obligor fails to meet a financial obligation,"* while McNab and Wynn (2000) list reasons for default, such as disputes with the lender, financial naivety, fraud, marital breakdown, or job loss. Thomas et al. (2016, 476) define default: *"as borrowers being 90 days overdue or there is evidence to the lender that the borrowers will not repay"*. They continue to discuss that when a debtor defaults on a loan, the collections process is triggered as the lender attempts to retrieve the outstanding debt, and the effectiveness of the collections process is typically measured by the Recovery Rate (RR) attained. RR is calculated as 1 - LGD. Matuszyk A. et al. (2010, 393-394) highlight that the Basel Accord mandates lenders to estimate LGD for loans both in default and not in default.

2.3 Loss given default

Schuermann (2004, 3) discusses how LGD is defined as "*the ratio of losses to exposure at default*" and continues how after default, LGD losses can be either the loss of principal, costs from carrying the non-performing loans, or collection expenses. Leymarie et al. (2018, 350) see LGD also as the ratio of the loss that never is recovered by a financial institution in case of customer default, while Leow, M. et al. (2014, 363) see LGD models forecasting "*losses as a proportion of the outstanding loan, if a debtor were to default"*. The definition of the

LGD is agreed upon in the literature, but it could be argued that selecting suitable models and measurements can be more complex.

2.4 LGD in financial organizations

As stated in EBA (16/2017, 5), a particular methodology for LGD estimation is not given while continuing that various LGD methodologies may be valid depending on specific circumstances, portfolios, and processes. That said, for organizing LGD modeling in financial institutions, some guidelines are given as EBA (16/2017, 11) has provided an internal cycle of risk parameter estimates represented in Figure 3. The steps are equally important for banks approved under the IRB approach as regulatory supervisors evaluate institutions. The IRB method is a comprehensive approach to risk assessment and is required under Regulation (EU) No 575/2013, Article 143.

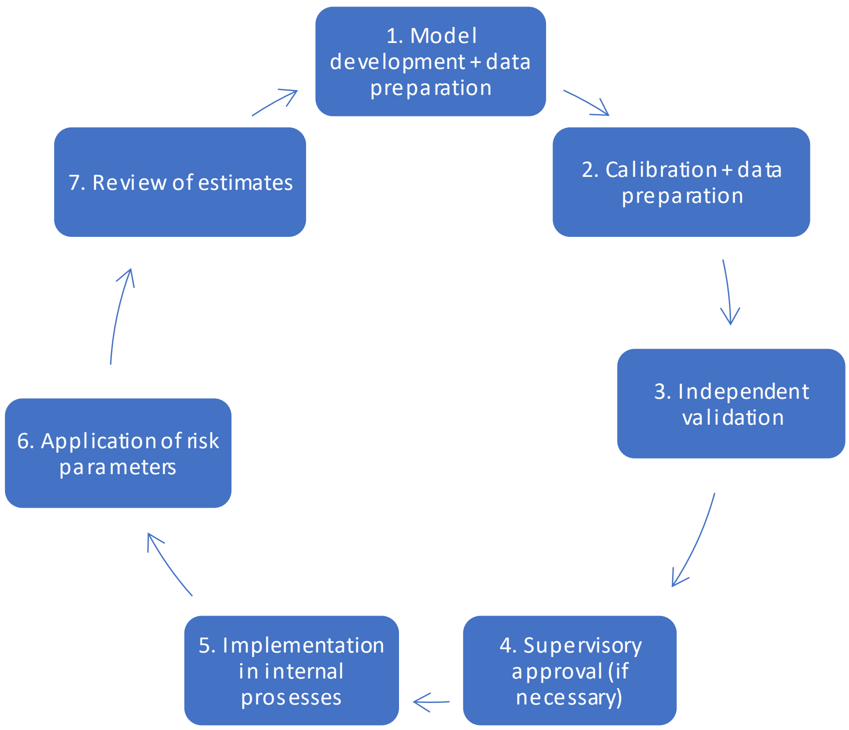

Figure 3. The internal cycle of risk parameter estimates. (EBA 16/2017, 11)

It is discussed in EBA (16/2017, 8) that model development is understood as part of estimating risk parameters that lead to appropriate risk differentiation. Financial institutions must carefully choose a dataset for developing their models to avoid any negative impact on the model's performance when applied to the actual portfolio. To ensure that the data is representative, institutions must analyze it during the model development stage, considering various factors such as the scope of the model's application, how the default is defined, risk characteristics distribution, lending standards, and recovery policies. By doing so, institutions can increase the accuracy and reliability of their models, resulting in better risk management practices. (EBA 16/2017, 53)

2.5 Modeling methods and techniques used in LGD modeling

Numerous studies on modeling LGD have been conducted, utilizing various methods, including regression techniques, decision trees, Markov chain models, and many more. (Matuszyk, A. et al., 2010; Loterman et al., 2012; Thomas et al., 2016). These methods are thoroughly analyzed in the literature review.

As stated by Miller & Töws (2018), the estimation of LGD can be split into two categories, linear and non-linear methods. They continue how both methods contain various techniques, but there are inconsistent results regarding the comprehensibility of these models while concluding that linear regression is the most frequently applied model for LGD estimation. (Miller & Töws, 2018, 190). The general linear regression model in Figure 4. represents Zhang & Thomas (2012) take on the model.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m + \varepsilon,$$

where, in this case,

y is the recovery rate or recovery amount;

$\beta_0, \beta_1, \ldots, \beta_m$ are unknown parameters;

$x_1, x_2, \ldots, x_m$ are independent variables which describe various characteristics of the loan and the borrower; and

ε is a random error term.

Figure 4. The general linear regression model, according to Zhang & Thomas (2012, 205-206)

Zhang & Thomas (2012, 207-208) see the recovery rate as [recovery amount] / [default amount], where for the single distribution model and other models predicted recovery rate can be obtained by dividing the predicted recovery amount by the default amount. They mention that distribution in realized LGD is more likely to be bimodal due to realized LGD values mostly being near zero and one. They conclude how bimodal distribution suggests that sophisticated methods would be more accurate for LGD estimation. Loterman et al. (2012) also studied LGD estimation methods and concluded that non-linear models are superior to linear ones.

As discussed in EBA (2017), most LGD models for performing exposures are work-out LGD models, and Leymarie et al. (2018, 350) state the meaning of the work-out LGD as the historical data used by institutions from all default exposures and identifying relevant risk drivers. Bijak K. et al. (2015, 343) mention that LGD models for unsecured retail loans can be classified as one-stage or multi-stage approaches, and for the one-stage approach, they mention several regression models that are suggested in the literature: Ordinary Least Squares (OLS) regression, Least Absolute Value (LAV) regression, robust and ridge regression, beta regression, and fractional regression. They continue with other one-stage models mentioned in the literature, including Tobit and two-tailed Tobit and survival analysis, classification, and regression trees (CART), neural networks (NN), Multivariate Adaptive Regression Splines (MARS), and Least Squares Support Vector Machines (LSSVM).

Regarding the multi-stage approach, Bijak K. et al. (2015, 343) mention that there are stages in which separate models are estimated where the first model discriminates positives from zeroes and negatives, and in the two-stage approach, the second model allows for the estimation of the positive values. They conclude how logistic regression and decision trees are discrimination models in the first two stages.

Matuszyk, A. et al. (2010, 396-397) discussed how the LGD could be estimated using linear regression with the weight of evidence (WOE) approach. They continue how the WOE approach classified the target variable as whether the LGD value was above or below the mean. They conclude that with univariate analysis, five variables emerged as the most robust predictors of the LGD for cases where LGD was greater than zero, listing those as the number of months in arrears throughout the entire life of the loan, the number of months in arrears during the last 12 months, the application score, the loan amount, and the duration of the loan until default. (Matuszyk, A. et al., 2010)

2.6 Evaluating the goodness of LGD estimates

Leymarie et al. (2018, 349) point out how complex evaluating the goodness of LGD estimates can be. They also conclude that usually, with banks and in academic papers, LGD model comparison consists of a three-step process. First, they mention that the sample of defaulted credits is split into test and training data sets. Second, they mention that all models to be compared are estimated on the training set. Thirdly they see that the models are tested with the test set, and the comparison is made on traditional methods such as mean squared error (MSE) or mean absolute error (MAE). In their approach, Leymarie et al. (2018, 349) evaluate the goodness of LGD estimates regarding regulatory capital and the bank's capacity to face unexpected losses on its credit portfolio.

In EBA (16/2017, 102), Back-testing is mentioned, which involves comparing the actual loss experience on defaulted consumer loans to the estimated loss rate and assessing the model's accuracy based on the differences between the actual and estimated loss rates.

2.7 Significant factors to consider in LGD modeling

When considering factors to be included in LGD modeling, Han & Jang (2013, 21) point out these significant factors: the loan size, collateral, debt seniority, product type, firm size, creditworthiness, financial ratios, age of the firm, industry classification, macroeconomic condition, and collateral. They also highlight how different studies suggest different factors, and the only factor that studies agree on is collateral. Han & Jang (2013, 21)

Bijak K. et al. (2015) point out that an LGD model should be characterized by good performance with low errors and high correlation coefficients. They also mentioned that stability and intuitive covariates are essential and classified into five groups. Matuszyk A. et al. (2010) also identified important characteristics in LGD modeling. Both these findings are presented in Table 1.

Table 1. Factors to consider in LGD modeling (Bijak K. et al., 2015; Matuszyk, A. et al., 2010)

Bijak K. et al. (2015) findings	Matuszyk, A. et al. (2010) findings
socio-demographic variables, such as customer age	amount of the loan at opening
customer's financial situation, such as income	number of months with arrears within the whole life of the loan
account details, such as loan amount	number of months with arrears in the last 12 months
payment history, such as outstanding balance	time at the current address
macroeconomic variables.	joint applicant.

2.8 Data in LGD modeling

EBA (16/2017, 15) highlights that good data quality is fundamental to developing a robust rating system. It is continued that data representativeness may influence the estimates' accuracy, and to ensure the excellent performance of the models and their good predictive power, EBA (16/2017, 15) points out how institutions should have adequate policies, processes, and methods for assessing the representativeness of data used for the estimation of risk parameters. It is recommended that banks should be cautious when using multiple data sources and ensure that consistently applied data standards are in place for all data used during the estimation process. (EBA, 16/2017, 15)

According to the EBA (16/2017, 28), LGD estimates should be based on the institution's experience, which requires proper recording and storage of all relevant data. This extensive data includes the date of default, all cash flows and events after default, and information on the obligors and transactions that could be used as risk drivers in model development. Therefore, the reference data set (RDS) should include all necessary information for LGD model development and calibration. (EBA, 16/2017, 28)

2.9 Challenges in LGD modeling

Zhang & Thomas (2012, 204) argue that modeling LGD is more complex than modeling PD for two main reasons. Firstly, they note that a considerable portion of the data may be censored, as debts are still being repaid over an extended period. This poses an issue for linear regression, which does not handle censored data well. Secondly, Zhang & Thomas (2012, 204) suggest that different reasons may lead to debtor defaults, resulting in different repayment patterns. They explain that while some individuals intentionally avoid repayment, others may be unable to do so due to permanent changes in their circumstances. Consequently, for some, the reason for non-repayment may only be temporary. (Zhang & Thomas, 2012)

Bijak K. et al. (2015, 342) also raise the issue of how LGD for unsecured retail loans is difficult to model. They discussed how LGD takes values from the interval zero to one, and some models cannot use values outside this interval. They point out that the LGD distribution peaks at zero since many customers default, but continue that even at default, the customers pay their defaulted debt fully. They also mention another case where the customers peak in LGD one, meaning customers pay nothing. They conclude that these peaks may lead to challenges in modeling LGD with the notion that retail LGD has at least a five-year-old observation period under the IRB approach. (Bijak K. et al., 2015)

2.10 Improving LGD modeling

The EBA (16/2017) highlights the importance of regular reviews of LGD modeling and rating systems as part of the validation function. EBA (16/2017, 44-45) mentions that institutions should have internal policies for changing models and risk parameters based on regular reviews, independent validation, legal environment changes, internal audit review, and competent authority review. It is also mentioned that identified weaknesses should be analyzed appropriately, which may result in model changes. These measures aim to improve the accuracy and effectiveness of banks' LGD modeling and rating systems. (EBA 16/2017, 100)

3 Literature review

This section will review prior research on modeling loss given default in unsecured consumer loans. The primary objective of this literature review is to gain a thorough understanding of the various methodologies used in modeling LGD for unsecured consumer loans. By summarising the knowledge and information gathered from the literature, the review aims to provide insights and answers to the first research question, which pertains to the current state-of-the-art modeling loss given default in unsecured consumer loans.

The literature review represents a critical component of this thesis, serving as the foundation upon which subsequent analyses and discussions are built. After completing the literature review, the findings will be used as the foundation of the interview results obtained, thereby contributing to a more thorough understanding of the subject matter.

3.1 Literature selection process

In this chapter, the overall literature selection process is described. The literature search was done with the LUT Primo service, which provides a wide range of scientific articles from multiple databases, including Elsevier SD Complete Freedom Collection [SCCMFC], SpringerLink Palgrave Journals, and EBSCO eBook Academic Collection.

Search terms used in the first phase were:” Modeling loss given default on unsecured consumer loans” to get an overall view of the subject and articles. The search was limited to articles written in English and was peer-reviewed, complete, and available online. The search resulted in 165 results. Apostrophes were added to loss given default to give more accurate outcomes resulting in 38 results. These results were analyzed closely to identify articles on different modeling techniques for unsecured consumer loans.

3.2 Various modeling techniques

Qi and Zhao (2011) compared six modeling methods for LGD on a dataset of 3,751 defaulted securities in the US from 1985 to 2008. Their six methods included four parametric methods: ordinary least squares regression, fractional response regression, inverse Gaussian regression, and inverse Gaussian regression with beta transformation, and two non-parametric methods: regression tree and neural network.

Qi and Zhao (2011, 2855) found that non-parametric methods, such as regression trees and neural networks, performed better than parametric methods in both in-sample and out-of-sample predictions, provided over-fitting was appropriately controlled. They continue how fractional response regression performed slightly better than ordinary least squares regression among the parametric methods. They also noticed that the transformation methods' performance was sensitive to the random error term (ε), a slight adjustment to LGDs of 0 or 1 before the transformation. Therefore, Qi and Zhao (2011, 2855) suggested that models that produce strong bi-modal patterns may need better model fit and accurate LGD predictions. They conclude that even with an optimal ε, the performance of the transformation methods could only match that of ordinary least squares regression. (Qi and Zhao, 2011)

3.3 Various regression techniques

Loterman et al. (2012) organized a major LGD benchmarking study that evaluated various regression techniques for modeling and predicting LGD. Their research with 24 techniques analyzed six loss datasets from international banks. Notably, their research identified several modeling techniques, including one-stage models such as ordinary least squares regression, beta regression, robust regression, ridge regression, regression splines, neural networks, support vector machines, regression trees, and two-stage models that combine multiple techniques. (Loterman et al., 2012, p. 161)

Loterman et al. (2012, 169) concluded that there is a clear trend that non-linear techniques, particularly support vector machines and neural networks, perform significantly better than more traditional linear techniques. Additionally, it is noted that a significant proportion of the variability in LGD remains unaccounted for, as evidenced by the modest range of R-squared values, spanning from 4% to 43%, for the models' average predictive performance. Finally, they conclude that two-stage models built by a combination of linear and non-linear techniques are shown to have similarly good predictive power, with the added advantage of having a comprehensible linear model component. (Loterman et al., 2012, 169)

3.4 Linear regression and survival analysis

Zhang & Thomas (2012) compared linear regression and survival analysis models to model recovery rates and amounts. Their goal was to predict the loss given default for unsecured consumer loans and credit cards, but as the thesis scope suggests, we focus more on the unsecured consumer loans side of the research. They also provide their perspective on the benefits and drawbacks of utilizing single and mixture distribution models to estimate these quantities. (Zhang & Thomas, 2012, 204)

For linear regression models, Zhang & Thomas (2012, 208) employed one model with the recovery rate as the target variable and another with the recovery amount as the target variable. They point out that the linear regression model is the most obvious choice for predicting the recovery rate and presents only the recovery amount via the recovery rate model. According to Zhang & Thomas (2012, 209), Survival analysis represents a practical methodology for modeling recovery rate and LGD, given that standard linear regression models cannot accommodate debts that are still being paid off due to the non-normal distribution of recovery rates and the resulting violation of linear regression assumptions. They further conclude that survival analysis models can incorporate repayments as censored and integrate them into the model (Zhang & Thomas, 2012., 213).

3.5 Decision tree

Matuszyk A. et al. (2010) researched modeling LGD for unsecured personal loans with a decision tree approach as they modeled the collection process. They used individual-level data of defaulted personal loans granted by a UK financial organization. For collection models on the macro level, Matuszyk et al. (2010, p. 394) view LGD as a result of a combination of factors, including uncertainty regarding the borrower's ability and willingness to repay, as well as the lender's decisions on the appropriate collection strategy to be employed. They divided the collection process into three distinct phases, each of which had its own specific LGD limits:

1. In-house collection (where no penalties are imposed) - LGD values ranging from 0 to 1

2. Collection process outsourced to an agent (with a commission rate of 40%) - LGD values ranging from 0.4 to 1.

3. Selling off the debt (at 5% of the face value) - LGD value set at 0.95.

Matuszyk et al. (2010, p. 394) describe how a decision tree can be employed to model a problem in which a combination of decisions and random factors influences the outcome. In building an LGD repayment model for identifying the class of repayer, they used debtor and loan characteristics in a two-stage process. They utilized logistic regression in the first stage to differentiate between two distinct groups: those loans with LGD equal to zero and those with LGD greater than zero. Subsequently, in the second stage, they developed regression-based models for each group to calculate the LGD estimate for individual loans. (Matuszyk A. et al., 2010)

Matuszyk A. et al. (2010, 397) analyzed multiple regression techniques to identify the method that produced the most optimal fit. They mention several methods, such as standard linear regression using beta distribution transformation, before applying regression, a log-normal transformation, the Box–Cox method, and the WOE approach with linear regression.

They observed that R-squared values were relatively low, indicating that LGD values are challenging to predict with a high accuracy. (Matuszyk A. et al., 2010)

Matuszyk et al. (2010, 397-398) observe a significant overlap between collection scoring models and LGD modeling. They argue that modeling both lender decisions and debtor repayment risks is critical for a decision tree approach and represent an ideal methodology. They suggest that the resulting distribution is a mixture distribution, indicating that a two-stage process should be used to obtain estimates. They explain that they first employ logistic regression to determine the debtor's classification, followed by a regression approach to estimate LGD values for debtors in that class. Lastly, they conclude that the decision tree methodology allows for model adjustments in unfavorable economic conditions due to changes in lender collection policies and borrower repayment capabilities. (Matuszyk et al., 2010).

3.6 Markov chain models

Thomas et al. (2016) conducted a case study that modeled repayment patterns in the collections process for unsecured consumer debt. Specifically, they employed Markov chain models of payment patterns to estimate recovery rates. The research states that models were tested using an extensive portfolio of UK retail loans over ten years. They continue by utilizing two models with the Markov chain approach and a hazard rate approach, all of which were developed to model the payment patterns of debtors during the collections process for unsecured consumer debt. They observed that defaulters often balance between sequences of repayment and non-repayment. In the former, borrowers repay their debt in every payment period, while borrowers do not make any payments in the latter. Additionally, they calculated the LGD for portfolios by employing various write-off strategies and compared these results to actual LGD outcomes (Thomas et al., 2016, 478)

In their conclusions, Thomas et al. (2016, 486) discuss a way of modeling LGD and Recovery Rates (RR) since the LGD can be calculated as one minus the recovery rate (LGD = 1 – RR) for unsecured consumer loans. They continue about the RR model and its ability to illustrate how debtors repay their debts after default. They make a point that while these models can accurately predict LGD, it is important to note that the specific write-off policies of the collectors influence LGD values. They discuss how collectors can identify the most suitable write-off strategies by estimating the additional proportion of debt recovered and the extra effort required when a write-off policy is relaxed. They highlight that the Markov chain approach employs the sequences of consecutive or missed payments as the fundamental building blocks, alongside the average recovery rate for each payment sequence. They also noticed that this approach requires fewer data and leads to an analytic solution. (Thomas et al., 2016)

Thomas et al. (2016, 486) explained how the discrete hazard rate methodology could be utilized to estimate the probability of a defaulter making or missing payment within a specific month. They see that this approach involves more parameters and must be solved iteratively. They also conclude that these models are progressing and indicate what is possible with this repayment pattern approach. They further elaborate on the benefits of the discrete hazard rate approach, stating that it is not dependent on the LGD distribution form and can account for the collector's operating decisions, including their write-off policies. Additionally, they mention that this method can incorporate economic effects, making it a more comprehensive approach to LGD modeling. They conclude that these three issues, LGD distribution form, write-off policies, and economic effects, have historically been challenging to address in LGD modeling, and the discrete hazard rate approach provides a valuable solution to these difficulties (Thomas et al., 2016)

3.7 Ordinary least squares model

Leow M. et al. (2014) researched the role of macroeconomic variables in loan-level retail LGD models by incorporating macroeconomic variables in retail LGD models. Their research used an OLS model for an unsecured personal loan data set. Leow M. et al. (2014, 364) used the same data source as Matuszyk et al. (2010) and Loterman et al. (2012), but a different OLS method and as discussed in their research, a linear regression LGD model developed by Loterman et al. (2012) with defaulted loans characteristics. Leow et al. (2014, 365) discusses the variables used in their LGD model, including customer-related data and loan commitment variables. They mention that customer-related data consisted of the application score, indicator for a joint applicant, marital status, length of time at address, employment status, and residential status of the account holder at the start of the loan. They further continue that loan commitment variables included whether the borrower had a mortgage, current savings or personal account, loan amount, loan term, loan purpose, length of time the loan had been with the bank, and whether the loan was ever in arrears, as well as the extent of any arrears. (Leow M. et al., 2014)

Leow M. et al. (2014, 373) found that with unsecured personal loan LGD, net lending growth at default was the only statistically significant macroeconomic variable. They continue that despite no improvement in R-squared value, all macroeconomic variables are statistically insignificant. After accounting for loan-level characteristics, they conclude that the economy less affects personal loan LGD. (Leow M. et al., 2014)

3.8 Bayesian methods

Bijak. & Thomas (2015, 342) compared Bayesian methods with the frequentist two-step approach to model LGD for unsecured consumer loans. They highlighted the challenges of fitting a model to the data due to the complex nature of the LGD distribution. To address this issue, they proposed multi-stage models, such as the two-step approach presented by Matuszyk et al. (2010). (Bijak & Thomas, 2015)

Bijak and Thomas (2015, 342-343) observed that in the frequentist approach, two separate models are independently estimated, which they believed could create difficulties when combining them to forecast LGD. They continue to explain how the first logistic regression model differentiates between positive values and zeroes, whereas the linear regression model can be utilized to estimate the positive values. They explain that the Bayesian framework is a more cohesive way to model LGD, employing a single hierarchical model instead of two separate models. Bijak and Thomas (2015, 348) continue how this results in a unique predictive distribution of LGD for each loan rather than just a single number. They note that with a distribution, one can use its features, like quantiles. (Bijak & Thomas, 2015)

Bijak & Thomas (2015, 349-350) summarize their findings by stating that while the frequentist and Bayesian models have similar performance, the Bayesian model has several advantages over the frequentist approach. They argue that the Bayesian model is more coherent and provides a better uncertainty description. In their opinion, the most significant advantage of the Bayesian model is that it generates an individual predictive distribution of LGD for each loan, whereas the frequentist model only provides a point estimate of LGD. They explain that predictive distributions can be advantageous in LGD stress testing and approximating the downturn LGD, as they provide valuable information and benchmarks for LGD estimates. (Bijak & Thomas, 2015)

3.9 Results from the literature review

The literature review showed that various authors suggest different methods for modeling LGD in the context of unsecured consumer loans. Matuszyk et al. (2010) used a decision tree approach, while Qi and Zhao (2011) used a regression tree and neural network. Loterman et al. (2012) suggested a two-stage model that combines linear and non-linear techniques. Zhang & Thomas (2012) found linear regression better than survival analysis in single distribution models. Leow et al. (2014) used an OLS model, but only net lending growth was found to be statistically significant. Bijak and Thomas (2015) used Bayesian methods with a frequentist two-step approach, which generated an individual predictive distribution of LGD for each loan. Finally, Thomas et al. (2016) suggested a Markov chain approach and another hazard rate approach to deal with collector's operating decisions and include economic effects. A summary of the results is provided in Table 2.

Table 2. The key finding from the literature review

Authors	**Suggested methods modeling LGD**	**Key findings**
Matuszyk, A. et al. (2010)	Decision tree approach	Use logistic regression to estimate which class a debtor is in and then a

		regression-type approach for each class to estimate the LGD values of debtors in that class
Qi, Zhao (2011)	Regression tree and neural network	Non-parametric methods (regression tree and neural network) perform better than parametric methods
Loterman et al. (2012)	Two-stage model: linear OLS regression with non-linear support vector machines or neural networks	Two-stage models built by a combination of linear and non-linear techniques are shown to have a good predictive power
Zhang & Thomas (2012)	Linear regression is better than survival analysis in single distribution models with higher R-squared value and Spearman rank coefficient	Single distribution models with linear regression, as results indicate that these methods worked the best in the research
Leow M. et al. (2014)	OLS model for an unsecured personal loans data set	Only net lending growth is statistically significant, but including this variable did not improve R-squared
Bijak, K. & Thomas, L. C. (2015)	Bayesian methods with the frequentist two-step approach to model LGD for unsecured retail loans. The first model (logistic regression) separates positive values from zeroes, whereas the second model (linear regression) allows	The most important advantage of the Bayesian model is that it generates an individual predictive distribution of LGD for each loan, whereas the frequentist approach only produces a point estimate

	for the estimation of the positive values	
Thomas et al. (2016)	Markov chain approach and another hazard rate approach	It has the advantage that it does not depend on the form of the LGD distribution, can deal with collector's operating decisions, such as their write-off policy, and could include economic effects

4 Data and Methodology

A semi-structured interview was used to collect data around LGD modeling for unsecured consumer loans in case company x. Two individuals from company x were interviewed, an analyst and manager working with around LGD. The credit risk manager has worked in retail banks' credit business and credit risks for eight years. The credit risk analyst has worked in retail banks' credit business and credit risks for over three years.

A semi-structured interview was chosen due to its flexibility, data richness, and standardization. Kvale, S. (1994) sees how semi-structured interviews allow for flexibility in the interview process, as they allow the researcher to follow up on interesting responses or explore new topics that emerge during the interview. Silverman, D. (2001) points out how semi-structured interviews can provide rich and detailed information about participants' experiences, thoughts, and perspectives, allowing the researcher to understand the topic under study. Seidman, I. E. (2006) also highlights how semi-structured interviews can be somewhat standardized by using a set of predetermined questions, allowing for better participant comparison and more reliable data.

Based on the literature review findings, the interview questions were designed to produce specific information about LGD modeling for unsecured consumer loans. The interview data were collected during three separate meetings, first separately with the manager and analyst, and then a final meeting with both participants to consolidate the results from the interview. The participants were selected based on their expertise and experience in credit risk and financial modeling. The participants were also familiar with LGD modeling for unsecured consumer loans, which ensured that the data collected was relevant and accurate.

The interview questions were developed based on the literature review. The questions aimed to prompt insightful and informative responses from case company x, thereby contributing to a deeper understanding of the thesis topic. The questions for the interview were as follows:

- Q1: How is LGD modeling for unsecured consumer loans organized in your organization?
- Q2: What kind of methods do you use for unsecured consumer loan LGD modeling?
- Q3: What kind of techniques do you use for unsecured consumer loan LGD modeling?
- Q4: How do you evaluate the goodness of unsecured consumer loan LGD estimates?
- Q5: What factors are considered significant for unsecured consumer loan LGD modeling in your organization?
- Q6: What kind of data do you use to model unsecured consumer loan LGD?
- Q7: What kind of challenges have you faced while modeling unsecured consumer loan LGD?
- Q8: What actions are you currently taking to improve unsecured consumer loan LGD modeling?

5 Results

The results of the interview are discussed below. Interviews from a manager and an analyst are combined into one coherent answer.

Q1: How is LGD modeling for unsecured consumer loans organized in your organization?

The modeling of customers' unsecured consumer loans has been done with internal resources based on bank credit agreement data. The monitoring and annual validation of the models are also handled internally in the regulatory functions. The company applies the principle of three lines of defense. Business ownership of credit risk models belongs to the Credit Risk Management team, which is part of the first line of defense. As the owner of the models, Credit Risk Management is responsible for the operation, regulatory compliance, and monitoring of the models. In the development projects of new models, the team is mainly responsible. However, due to the significant workload of the modeling projects, several stakeholders participate in the modeling, mainly from the company's business, analytics, and risk control functions.

In modeling exercises, Credit Risk Management is primarily responsible for the regulatory compliance of the model and that the models are as suitable as possible for the company's business needs. Business representatives are responsible for applying models in different business processes and systems. The primary role of analytics is to ensure that the data used for modeling is high quality and can be maintained even after modeling. Risk control does not actively participate in decisions during modeling but evaluates the functioning and regulatory compliance of the models with independent assessments.

The primary task of Risk Control, which represents the second line of defense, is the independent assessment of the functionality of the models and development needs and reporting to the company's management and the board. The third line of defense (Internal Audit) task is to perform independent validation and evaluation directly for the bank's board.

Q2: What kind of methods do you use for unsecured consumer loan LGD modeling?

LGD models aim to predict the amount of conditional financial loss realized after default from the debt balance at the moment of default during the next 36 months (i.e., account-level target). The current models are based on statistical modeling, and regression models (fractional response regression) are used as forecasting models due to their ease of implementation and comprehensibility.

Modeling exercises often start with a preliminary investigation, explaining what models are needed and why. In addition, the company will find out what kind of data is available for the purpose in question. The model's purpose and the data used determine the technology used in modeling. Therefore, the company has not separately defined primary modeling techniques or methods; instead, they are evaluated on a model-by-model basis. The choice of modeling methods aims to avoid complexity. Often, the modeling material has significant shortcomings compared to the regulatory requirements, which is why using complex modeling methods does not necessarily improve the result. The methods are chosen to make the models' documentation, use, and maintenance as efficient as possible.

Q3: What kind of techniques do you use for unsecured consumer loan LGD modeling?

The models are produced using statistical methods, and various mathematical applications are often tested during the modeling phase. In the modeling phase, internationally known partners specializing in modeling who use advanced modeling software are often used as help. The benefit of the software is that it can be used to effectively test different modeling techniques and choose from the final results either the option with the best predictive ability, the simplest option in terms of modeling technique, or the desired compromise between the first two.

Regression models are a commonly used technique because they often achieve a good balance between the model's predictive ability and simplicity.

The stages of modeling can be divided into six different categories, which are

1. collecting data from the data warehouse fields
2. data analysis and time window selection
3. selection and formation of possible explanatory variables
4. choice of modeling method and variables
5. validation and testing of the final model
6. implementation and monitoring of the selected model

The explanatory variables that are candidates for the model are intended to be formed as widely as possible from different account and counterparty level data, utilizing the experts' knowledge of the best market practices and the bank's available information. The selection technique of models and variables combines statistical and business criteria. In the modeling, we aim for historical data that is as clean as possible, within which no significant changes have occurred in operational production processes, and the data is as consistent as possible over time. In addition to standard statistical methods, the model selection criteria also take into account business needs, and the overall criteria can be simplified into three categories:

1. statistical model selection criteria

2. validation and test results of different methods

3. business needs

Q4: How do you evaluate the goodness of unsecured consumer loan LGD estimates?

The risk control validation process assesses the reliability of the LGD estimates. The sparseness of data causes challenges to the process. The company does not calculate loans realized LGD values, which could be evaluated against the estimates produced by LGD models. The aim is to estimate realized losses as accurately as possible for validation needs.

The model's prediction ability was evaluated using statistical methods such as Pearson correlation and R-square in the modeling phase. Due to the lack of monitoring, there is no certainty about the continuity of the forecasting ability. Remodeling the LGD models is the primary way to improve the plan models' predictive ability and correct the monitoring challenges. In connection with the development and implementation of the model, statistical methods are used to measure the model's functionality. The partly incomplete monitoring environment of the current production process makes it challenging to manage the continuity of the model, and the validation and monitoring results can only be considered indicative.

Q5: What factors are considered significant for unsecured consumer loan LGD modeling in your organization?

The model's target variable is the realized credit loss after collection actions. In other words, the models aim to predict the company's final loss after all available collection means, considering the time value of money and the collection costs. In the current models, the moment of insolvency is identified as when collection actions start, but the company's future models will be based on the concept of insolvency defined in the solvency regulation.

The aim is to produce the most extensive possible set of potential explanatory variables from the data available for modeling, from which the variables that best predict the target variable are selected. The variables are primarily selected according to statistical evidence, but, the quality and validity of the variables, compatibility with the company's strategy and goals, and the acceptability of the data are noted. To guarantee the principle of equality and personal data protection, the use of certain information in the models is limited, such as age and gender. In general, the explanatory variables of LGD models measure customer demographics, customer or credit behavior, and credit account application data.

Q6: What kind of data do you use to model unsecured consumer loan LGD?

LGD models are based on customer and account information available to the company. The aim is to develop the models using the company's internal data, but it may be supplemented occasionally with information from partners or publicly available information. From the point of view of LGD models, the data must describe credit events as accurately as possible after insolvency up to a possible credit loss or account closure. In addition, it should be possible to form variables that predict the amount of the final credit loss as well as possible from the data.

Demographic data of the counterparty, behavioral data of the counterparty and/or the account, and credit account application data are generally used to form explanatory variables. As supplementary information, market databases or generally available information that can be classified into them can be used. Target is formed from account/contract level information, but information from the counterparty level is also utilized to form explanatory variables.

Q7: What kind of challenges have you faced while modeling unsecured consumer loan LGD?

The primary challenges are related to the available data. Modeling can only be based on the data produced by the company's processes. In the planning and development of the processes, the quality of the data produced by the process has not necessarily been regarded as a primary goal, and therefore they primarily need to meet the requirements for modeling data given in the regulation. The challenges related to data quality are common to the company, but according to industry consultants, the challenges are expected. Data-related challenges are also significant because the material needed for modeling should represent a long-time span, usually at least five years. Therefore, it is usually impossible to fix deficiencies quickly, but accumulating sufficient historical data can take considerable time.

Other challenges are the low use of models in business processes, the challenges related to calculating realized LGD values, and the consideration of macroeconomic situations in the models, the so-called downturn LGD calculation. External challenges are faced because a general feature of the Finnish credit market, low default cases, makes modeling difficult. Additional segmentation of the models may not be able to be produced after the initial segmentation, which reduces the models' performance. In addition to this, regulatory requirements, for example, the length of modeling data, create additional challenges.

Internal challenges are faced when operational production processes have not been designed with the models' needs in mind, so the data quality does not always meet the required modeling standard. The development of operational production processes is slow and expensive, and the needs of a few models are not enough to increase the development priority of the processes enough.

Q8: What actions are you currently taking to improve unsecured consumer loan LGD modeling?

The LGD models currently in use in the company are used to generate the parameters of the ECL calculation. Due to the shortcomings of the models, the model use has yet to be extended to the company's other processes. The models are at the end of their life cycle, and the primary plan to improve the quality of the models is to remodel the LGD models. Business processes are being developed in the company, especially in credit monitoring and collection. In addition to other business needs, the development of the processes aims to promote the production of high-quality material for the company's future LGD modeling. In addition, the company promotes the development of data warehouses, which can significantly facilitate the management and utilization of the material needed for modeling in the future.

5.1 Results Analysis

The results from the interview and theory are summarized in Table 3. After the summarized results, the analysis continues with a more thorough analysis of interview results versus theory.

Table 3. Interview and theory

Question	**Interview**	**Theory/regulation**
Q1: How is LGD modeling for unsecured consumer loans organized in your organization?	-Banks credit agreement data -Validation of the models are handled internally -The principle of three lines of defense -Business ownership of credit risk models belongs to the Credit Risk Management team	-Internal cycle of risk parameter estimates. (EBA 16/2017, 11)
Q2: What kind of methods do you use for unsecured consumer loan LGD modeling?	-LGD models aim to predict the amount of conditional financial loss realized after default from the debt balance, at the moment of default during the next 36 months -Statistical modeling, and regression models, but evaluated on a model-by-model basis. -Efficiency aspect: model maintenance and documentation versus the complex modeling	-Numerous studies utilizing various methods, including regression techniques, decision trees and Markov chain models. (Thomas et al., 2016) -Mostly work-out LGD models (EBA 2017) -One-stage or multi-stage approaches. (Bijak K. et al., 2015)

Q3: What kind of techniques do you use for unsecured consumer loan LGD modeling?	-In the modeling phase, internationally known partners specializing in modeling who use advanced modeling software different modeling techniques and choose from a) the option with the best predictive ability b) the simplest option in terms of modeling technique c) the desired compromise between a and b.	-Linear regression is the most frequently applied for LGD estimation. (Miller & Töws, 2018, 190).
Q4: How do you evaluate the goodness of unsecured consumer loan LGD estimates?	-Pearson correlation and R-square -Remodeling the LGD models statistical methods are used to measure the model's functionality.	-A three-step process. The sample of defaulted credits is split into test and training data sets. -Models to be compared are estimated on the training set. -Models are tested with the test set, and the comparison is made on traditional methods such as mean squared error (MSE) or mean absolute error (MAE) (Leymarie et al., 2018, 349)
Q5: What factors are considered significant for unsecured consumer loan LGD modeling in your organization?	-The explanatory variables of LGD models measure customer demographics, customer or credit behavior, and credit account application data.	-Socio-demographic variables, such as customer age. -Customer's financial situation, such as income -Account details, such as loan amount -Payment history, such as outstanding balance -Macroeconomic variables. (Bijak K. et al. 2015, 344)
Q6: What kind of data do you use to model unsecured	-Demographic data -Behavioral data -Credit account application data	-Good data quality -Relevant data must be appropriately recorded and stored. -Representativeness of data

consumer loan LGD?		(EBA 16/2017, 15, 28)
Q7: What kind of challenges have you faced while modeling unsecured consumer loan LGD?	-Data-related challenges -Low use of models in business processes	-The data may be censored (Zhang & Thomas, 2012). -Debtors have different reasons for defaulting (Zhang & Thomas, 2012). -LGD has at least a five-year-old observation period under the IRB approach. (Bijak K. et al., 2015)
Q8: What actions are you currently taking to improve unsecured consumer loan LGD modeling?	-The primary plan is to improve the quality of the models is to remodel the LGD models. -Business processes are being developed in the company, especially in credit monitoring and collection.	-Regular reviews of LGD modeling (EBA, 16/2017) -Institutions should have internal policies for changing models and risk parameters based on regular reviews, independent validation, changes in the legal environment, internal audit review, and competent authority review. (EBA 16/2017, 44-45)

In Q1, regarding LGD modeling for consumer loans in organizations, case company x's answers align with the EBA (16/2017). According to EBA (16/2017), financial institutions must carefully choose a dataset for developing their models and analyze it during the model development stage to ensure that the data is representative. Case company x mentions that the data used for modeling is of high quality and that the analytics team ensures that the data can be maintained even after modeling. EBA (16/2017) also highlights the importance of an internal risk parameter estimate cycle. Case company x mentions that the monitoring and annual validation of the models is handled internally in the regulatory functions. Case company x also applies the three lines of defense principle, which aligns with EBA (16/2017) guidance.

Questions Q2 and Q3, which asked about methods and techniques for consumer loan LGD modeling, are analyzed as a whole, as methods and techniques go hand in hand. In Q2 and Q3, it is mentioned that in case company x, regression models are a commonly used technique because they often achieve a good balance between the model's predictive ability and simplicity. Case company x also mentioned that the primary modeling techniques are evaluated on a model-by-model basis. The literature also mentioned that Linear regression is the most frequently applied model for LGD estimation. (Miller & Töws, 2018, 190). It is worth mentioning that in the literature, Zhang & Thomas (2012) noted that distribution in realized LGD is more likely to be bimodal, which is due to realized LGD values mostly being near zero and one. This bimodal distribution suggests that sophisticated methods would be more accurate for LGD estimation.

In Q4, concerning evaluating the goodness of consumer loan LGD estimates, the case company's approach and the theory presented by Leymarie et al. (2018, 349) share similarities in using statistical methods to evaluate the predictive ability of LGD models. However, the case company's approach does not mention using test and training datasets or traditional comparison methods like MSE or MAE. In Q5, relating to significant factors for consumer loan LGD modeling, there are some similarities between the case company's answer and the factors identified in the literature (Han & Jang 2013, 21; Bijak K. et al., 2015; Matuszyk, A. et al., 2010), but there are also some differences. The case company's approach to selecting variables based on statistical evidence and compatibility with the company's strategy and goals is consistent with the literature's emphasis on intuitive covariates and stability in LGD modeling. The case company does not go into specific variables when describing significant factors, whereas the literature details these variables.

While in Q6, about data to model consumer loan LGD, and the EBA (16/2017) stress the importance of using relevant and high-quality data to model consumer loan LGD, there are some key differences between the two sources. The case company x's answer focuses more on the specific types of data used to form explanatory variables for LGD models, such as demographic and behavioral data of the counterparty and/or the account. Case company x answer also mentions that publicly available information or market databases may be used as supplementary information. In contrast, the EBA (16/2017) guidelines offer detailed guidance on the necessary data scope for accurate LGD estimation and emphasize the significance of properly recording and storing all relevant data for LGD modeling purposes. Another key difference is that the case company x's answer does not explicitly address the issue of data representativeness. At the same time, the EBA (16/2017) emphasizes the need for institutions to have adequate policies and processes for assessing data representativeness to ensure the accuracy of LGD estimates.

In Q7, touching on challenges faced by modeling consumer loan LGD, the case company's answer focuses on the challenges related to available data and data quality, low use of models in business processes, calculating realized LGD values, macroeconomic factors, and the low number of default cases in the Finnish credit market. On the other hand, the theory by Zhang & Thomas (2012) and Bijak K. et al. (2015) highlights the difficulties in modeling LGD due to censored data and the different reasons for defaulting among debtors, as well as the challenges in modeling LGD for unsecured retail loans that take values from the interval zero to one.

In Q8, about improving consumer loan LGD modeling, both case company x answer and the EBA (16/2017, 44-45, 100) emphasize the importance of regularly reviewing and improving consumer loan LGD modeling. Case company x focuses more on specific strategies such as model remodeling, process development, and data warehouse promotion, while the EBA (16/2017, 44-45, 100) provides more comprehensive guidance on the validation function and the need for internal policies and analysis of identified weaknesses.

6 Conclusions and Discussion

This thesis aimed to explore the state-of-the-art methods for modeling loss given default on unsecured consumer loans and to investigate how these methods are currently being applied in practice in the case company x. A comprehensive literature review found that the most used LGD modeling techniques are regression-based models. However, the choice of the most suitable method depends on the specific characteristics of the loan portfolio. In the case study of company x, it was found that the company's LGD modeling approach was consistent with the literature. The company's model was a regression-based model that used customer demographics, customer or credit behavior, and credit account application variables to predict LGD.

6.1 Answering the research questions

The literature review addressed prior research on modeling loss-given default in unsecured consumer loans. The purpose was to answer the first research question:

1. *Based on the existing literature, what are the state-of-the-art methods to model loss given default on unsecured consumer loans, and how are they currently applied in practice?*

Literature review results confirm that there is no straightforward way to model LGD on unsecured consumer loans. Loterman et al. (2012) benchmarked regression algorithms for LGD modeling and found that machine-learning algorithms performed better than classical regression methods. Matuszyk et al. (2010) proposed a decision tree approach for modeling LGD in unsecured personal loans. Zhang & Thomas (2012) compared linear regression, and survival analysis using single and mixture distribution approaches in modeling LGD and found that survival analysis was more effective. Bijak and Thomas (2015) proposed a Bayesian method for modeling LGD in unsecured retail loans with two-stage approach. Thomas et al. (2016) developed a case study on modeling repayment patterns in the collections process for unsecured consumer debt. Leow et al. (2014) studied the impact of the economy on LGD and found that unemployment rates and housing prices were significant factors. Qi and Zhao (2011) compared modeling methods for LGD and found that ensemble methods performed better than individual methods.

The conclusion on how to go about modeling unsecured consumer loans could be to use a combination of machine learning algorithms and survival analysis while considering the specific economic factors, repayment patterns, and lending policies relevant to the specific type of loan being modeled. Bayesian and decision tree approaches may also be helpful depending on the situation.

The interview addressed a practical viewpoint and answered the second research question:

2. *How is LGD modeling currently implemented in case company x, and what would be the key points to develop it further?*

Based on the provided material, it can be concluded that case company x follows the EBA (16/2017) guidance on LGD modeling for consumer loans, especially in terms of data selection and analysis, risk parameter estimates, and the three lines of defence principle. The company primarily uses regression models for LGD estimation, but there is a suggestion from the literature that more sophisticated methods would be more accurate, given the bimodal distribution of realized LGD values. The case company's approach to selecting variables based on statistical evidence and compatibility with its strategy and goals aligns with the literature's emphasis on intuitive covariates and stability in LGD modeling. However, unlike the literature, the case company does not provide specific variables describing significant factors. Although, this may be due to confidentiality.

The case company's approach to data collection and usage focuses on specific data types used to form explanatory variables for LGD models, such as demographic and behavioral data of the counterparty and the account. The EBA (16/2017) provides more comprehensive guidance on the necessary data scope for accurate LGD estimation and emphasizes the importance of assessing data representativeness. The challenges in LGD modeling mentioned by the case company include available data and data quality, low use of models in business processes, calculating realized LGD values, macroeconomic factors, and the low number of default cases in the Finnish credit market. The literature also highlights difficulties in modeling LGD due to censored data, different reasons for defaulting among debtors, and challenges in modeling LGD for unsecured retail loans that take values from the interval zero to one. The case company and the EBA (16/2017) stress the importance of regularly reviewing and improving consumer loan LGD modeling. However, the EBA (16/2017) provides more comprehensive guidance on the validation function and the need for internal policies and analysis of identified weaknesses. The case company focuses on strategies, such as model remodeling, process development, and data warehouse promotion.

To develop LGD modeling further in case company x, some key areas for improvement could be exploring more sophisticated methods for LGD estimation, addressing data representativeness more explicitly, and conducting regular and comprehensive validation of models. Additionally, the company could consider incorporating more traditional comparison methods like MSE or MAE and including test and training datasets if these are not in use already.

6.2 Verification and Validation

The primary sources of information for the literature review were academic journals, books, and other relevant publications related to LGD modeling for unsecured consumer loans. These sources were verified to be credible and reliable, as they were published in reputable academic journals and written by experts in credit risk, financial modeling, and their respective research fields.

The empirical part considered only one company which restricts the generalizability of the results. The limitations of a semi-structured interview should be recognized: Flick, U. (2009) points out how semi-structured interviews can be time-consuming in conducting, transcribing, and analyzing the data. Silverman, D. (2001) explains how semi-structured interviews are subject to researcher bias, particularly in selecting questions and interpreting responses. Kvale, S. (1994) also highlight that semi-structured interview depends on participant's willingness to be interviewed and their verbal skills.

6.3 Future research

Future research could investigate LGD modeling on secured loans, as the thesis did not study LGD modeling on loans with collaterals. The thesis could have also included a case study of another company to compare and determine LGD modeling practices. Additionally, larger focus groups could have complemented the semi-structured interview to obtain a broader perspective on the company's LGD modeling practices. Future studies could also explore modeling LGD on different loan types and add more case companies.

References

Amadi, C.W. (2012) An examination of the adverse effects of consumer loans. International Journal of Business and Management, 7 (3).

Bijak, K. & Thomas, L. C. (2015) Modelling LGD for unsecured retail loans using Bayesian methods. The Journal of the Operational Research Society. 66 (2), 342–352.

European Banking Authority. (2017) EBA Report on IRB modelling practices. [Online] Available at: EBA Report on IRB modelling practices.pdf (europa.eu) [Accessed 5 February, 2023].

European Banking Authority. (16/2017) Guidelines on PD estimation, LGD estimation and the treatment of defaulted exposures. [Online] Available at: https://www.eba.europa.eu/documents/10180/2033363/Guidelines+on+PD+and+LGD+estimation+%28EBA-GL-2017-16%29.pdf/6b062012-45d6-4655-af04-801d26493ed0 [Accessed 5 February, 2023].

European Central Bank. (2016) Bank lending survey for the euro area. [Online] Available at: https://www.ecb.europa.eu/stats/pdf/ecbblsglossary.en.pdf [Accessed 26 October 2022].

Flick, U. (2009). An Introduction to Qualitative Research (4th ed.). London: SAGE Publications.

Han, C. & Jang, Y. (2013) Effects of debt collection practices on loss given default. Journal of Banking & Finance. 37 (1), 21–31.

Hurlin, C., Leymarie, J. and Patin, A. (2018) Loss functions for loss given default model comparison. European Journal of Operational Research, 268 (1), 348–360.

Kaposty, F. et al. (2020) Predicting loss given default in leasing: A closer look at models and variable selection. International journal of forecasting. 36 (2), 248–266.

Kvale, S. (1994) Interviews: An introduction to qualitative research interviewing. Sage Publications, Inc.

Leow, M. et al. (2014) The economy and loss given default: evidence from two UK retail lending data sets. The Journal of the Operational Research Society. 65 (3), 363–375.

Leymarie, J., Hurling, C. and Patin, A., (2018) Loss Functions for LGD Models Comparison. European Journal of Operational Research, 268(1), 348-360.

Loterman, G. et al. (2012) Benchmarking regression algorithms for loss given default modeling. International journal of forecasting. 28 (1), 161–170.

Loterman, G. et al. (2014) A proposed framework for backtesting loss given default models. Journal of risk model validation. 8 (1), 69–90.

Matuszyk, A. et al. (2010) Modelling LGD for unsecured personal loans: decision tree approach. The Journal of the Operational Research Society. 61 (3), 393–398.

Pickert, J. (2017) Risk assessment of unsecured loans – example of entering a new market. The Central European Review of Economics and Management. 1 (3).

Qi, M. & Zhao, X. (2011) Comparison of modeling methods for Loss Given Default. Journal of banking & finance. 35 (11), 2842–2855.

Regulation (EU) (2013) No 575/2013 of the European Parliament and of the Council of 26 June 2013 on prudential requirements for credit institutions and investment firms and amending Regulation (EU) No 648/2012

Schuermann, T. (2004). What do we know about loss given default? Federal Reserve Bank of New York.

Seidman, I. E. (2006) Interviewing as Qualitative Research: A Guide for Researchers in Education and the Social Sciences. Teachers College Press

Silverman, D. (2001) Interpreting Qualitative Data: Methods for Analysing Talk, Text and Interaction. Sage

Spuchľáková, E., Valašková, K. and Adamko, P. (2015). The credit risk and its measurement, hedging and monitoring. Procedia Economics and Finance, 24, 675–681.

Thomas, L. C. et al. (2016) Modelling repayment patterns in the collections process for unsecured consumer debt: A case study. European journal of operational research. 249 (2), 476–486.

Zhang, J. & Thomas, L. C. (2012) Comparisons of linear regression and survival analysis using single and mixture distributions approaches in modelling LGD. International journal of forecasting. 28 (1), 204–215.