Modeling recovery rates for small- and medium-sized entities in the US

Aleksey Min, Matthias Scherer, Amelie Schischke, Rudi Zagst

Technical University of Munich, Munich, Germany.

Abstract. A sound statistical model for recovery rates is required for various applications in quantitative risk management. We compare different models for predicting the recovery rate on borrower level including linear and quantile regressions, decision trees, neural networks and mixture regression models. We fit and apply these models on the worldwide largest loss and recovery dataset for commercial loans provided by Global Credit Data, where we focus on small- and medium-sized entities in the US. Additionally, we include macroeconomic information via a predictive Crisis Indicator. The horserace is won by the mixture regression model with regressed weight probabilities.

Keywords: Decision Tree, Loss Given Default, Mixture Model, Neural Network, predictive Crisis Indicator

1. Introduction

Additional capital requirements and an increased awareness of the importance of credit risk modelling are one consequence of the financial crisis of 2007. Capital requirements, like the internal ratings-based approach of Basel II, allow financial institutions to estimate their credit risk by own models. The main determinants of credit risk are the probability of default (PD), the exposure at default (EAD) and the loss given default (LGD), the latter is linked to the recovery rate (RR) via $RR = 1 - LGD$. We focus on the modeling of the recovery rate and compare different methods to estimate a firm-specific one.

According to §297 of the Basel Comittee on Banking Supervision (2004), LGD has to be measured as "loss given default as a percentage of the EAD." However, there exist several methods to calculate the LGD (resp. RR), namely the market LGD, the implied market LGD and the

Adress for correspondence: Amelie Schischke, Department of Mathematics, Technical University of Munich, Boltzmannstr. 3, 85747 Garching, Germany, E-Mail: amelie.schischke@tum.de

workout LGD. For loan data, the appropriate definition is the workout RR, which are the revenues (R) financial institutions can collect reduced by all administration costs (A) during the resolution period in case of a default divided by the outstanding amount at default (EAD). Ye and Bellotti (2019) define the RR for a defaulted loan with exposure at default EAD as:

$$RR = \frac{R - A}{EAD} = \frac{\sum \text{Collections} - \sum \text{Admin Fee}}{\text{Outstanding Balance at Default}}.$$

With this definition, it is actually possible that the RR can take values greater than one or smaller than zero. Keijsers et al. (2015) mention as an example for a RR smaller than zero principal advances. On the contrary, in cases of penalty fees, additional interest and recovered principal advances, the RR can attain a value greater than one. Both is frequently observed in our data.

Our objective is to compare different methods to model the recovery rate, namely linear regression, quantile regression, decision trees with linear/quantile regression, neural networks and mixture regression models. Thereby, we investigate how information on an economic crisis affects these models.

We base this study on the LGD&EAD platform provided by Global Credit Data (GCD) which contains information about defaulted loans; see Section 4.1 for more details. Moreover, we aggregate the information on loan level to borrower level and focus on small- and medium sized entities (SME) in the US. Inspired by the findings of Brumma and Winckle (2017), who observe that the macroeconomic behavior during the resolution time has an influence on the recovery rate, we include a predictive Crisis Indicator which predicts whether a crisis might occur during the time of resolution in our models.

This paper is structured as follows: In the second section, we survey the literature for modelling the RR. Subsequently, we provide a theoretical background of the techniques used in this study in Section 3. Thereafter, Section 4 presents the structure of the database. The results of the models are shown in Section 5. Section 6 summarizes the results and we discuss possibilities for further research on the RR.

2. Literature Review

In literature, several models are suggested to estimate the RR. We give a short overview with focus on regression models, decision trees, neural networks and mixture models.

According to Yao et al. (2015) and Bastos (2010b), classical linear regression models are the most popular and most straightforward techniques to estimate the RR. However, Dermine and

Carvalho (2006) and Qi and Zhao (2011) mention as drawback that in reality, RRs are bounded and not normally distributed. Nevertheless, the linear regression model outperforms the Tobit model and the decision tree model for UK credit card accounts in the study of Bellotti and Crook (2009).

Many authors have adapted regression models to the situation of RRs. The inverse Gaussian (IG) regression transforms the RR by an inverse Gaussian distribution function from the interval $(0, 1)$ to the real line. Qi and Zhao (2011) compare this to the inverse Gaussian regression with beta transformation, which is also used by Gupton and Stein (2005), Loterman et al. (2012) and Yao et al. (2015), where the assumption of beta distributed LGDs is postulated and subsequently, the inverse Gaussian distribution is applied. Linear regression aims at predicting the mean, whereas a quantile regression can analyze the influence of covariates on the entire distribution. Krüger and Rösch (2017) emphasize that quantile regression might hence be better suited for downturn scenarios.

In order to model the concentration of RRs at the boundaries $\{0, 1\}$, Bellotti and Crook (2009) propose a decision tree model which is also used by Yao et al. (2015). A logistic regression model decides whether the RR takes the values 0 or 1. Subsequently, an ordinary least squares method is used inside $(0, 1)$. Similarly, Loterman et al. (2012) use a logistic regression to determine whether the RR takes the boundary values and different parametric as well as non-parametric models to explain the RR inside $(0, 1)$, but the single application of the non-parametric models, esp. neural networks and the least squares support vector machines, outperforms the combinations.

Besides Loterman et al. (2012), non-parametric models are also studied by several researchers: In Bastos (2010b) and Qi and Zhao (2011), neural networks outperform the fractional response regression resp. linear regression, Inverse Gaussian regression, Inverse Gaussian regression with Beta transformation and the fractional response regression. However, Qi and Zhao (2011) mention as drawback that neural networks are a "black-box," because there is no straightforward method to interpret the relationship between the independent and dependent variables.

Another type of models, which are considered in different ways to predict the RR, are finite mixture models. Krüger and Rösch (2017) use a normal mixture distribution with two components for LGD and find that it performs best with their quantile regression on the GCD subset of US SMEs. However, Ye and Bellotti (2019) propose a two-stage model to apply a beta mixture model for the RRs in $(0, 1)$. This two-stage model outperforms the OLS, OLS with lasso as well as the beta regression. In addition, Tomarchio and Punzo (2019) present zero-and-one inflated mixture models. A three-level multinomial model first decides, whether the LGD takes the value

0, 1 or lies in $(0, 1)$. Subsequently, finite mixture distributions are applied to $(0, 1)$, in which they test different component distributions.

In the study of Altman and Kalotay (2014), the transformed RR by the inverse normal distribution is approximated as a mixture of Gaussian distributions, where only the probability belonging to a certain component depends on covariates. Wang et al. (2018) extend this mixture model on the Moody's Ultimate Recovery Database by introducing a Markov switching model with two states, representing crisis and non-crisis periods to capture cyclical aspects. For each state, there is a mixture model with four components for the transformed RR which enables the determination of the influence of covariates.

Similar to the decision trees, Calabrese (2012) presents a mixed continuous-discrete model. In her further work, Calabrese (2014) extends her model by introducing a mixture model. The LGD is modeled as a mixture of the expansion and recession distribution where each distribution is represented by the mixed model in Calabrese (2012). The mixtures represent the credit cycle whether there are bad or good times.

3. Modeling Methods

This section provides a theoretical background of the techniques used in this study. We focus on decision trees, neural networks and mixture regression models. We refer to Fahrmeir et al. (2013), Hosmer and Lemeshow (2013) and Krüger and Rösch (2017) for more information on regression methods, in particular for quantile regression models as well as for model selection techniques.

3.1. Decision Tree

Since RRs are not normally distributed, a linear regression might not be adequate. As an alternative, the RR can first be transformed and then, on the transformed data, a linear regression can be applied. In the literature, e.g. in Gupton and Stein (2005), a beta transformation is used. The transformed RR is:

$$\text{Transformed RR} = \Phi^{-1}(F_{Beta}(RR, \alpha, \beta)), \tag{1}$$

where Φ^{-1} is the quantile function of the standard normal distribution function and $F_{Beta}(x, \alpha, \beta)$ is the distribution function of the beta distribution with shape parameters α and β, which have to be estimated. However, this transformation can only be applied to $RR \in (0, 1)$. As our dataset

also contains observations with a RR smaller than zero or greater than one, we use a decision tree approach as displayed in Figure 1.

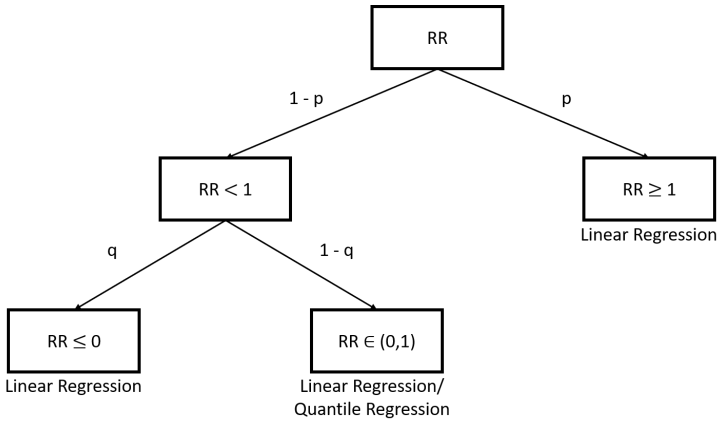

Fig. 1: Structure of the decision tree model.

Firstly, a logistic regression (or a neural network) determines the probability p that the RR is greater than or equal to 1. Then, a second classification model, i.e. a logistic regression (resp. neural network), estimates the probability q that the RR takes a value less than or equal to 0, given that it is smaller than 1. We use a linear regression to predict the rates $RR_{\geq 1}$ and $RR_{\leq 0}$. Inside $(0,1)$, we apply the beta transformation (1) to the RR. Subsequently, a linear regression (or a quantile regression) estimates the rate $RR_{(0,1)}$. If the linear regression had been applied to the raw RR, there would be predicted values outside $(0,1)$. Therefore, we first apply the beta transformation and, on the transformed RR, we can use the linear regression. In contrast to the linear regression, the estimates of the quantile regression would not exceed the open unit interval. Therefore, we apply this regression type on the raw RR and compare the results. We mention that, according to our results, it is better to apply the quantile regression on the raw $RR \in (0,1)$. Hence, in the following, the corresponding results are presented. The expected RR is expressed as a weighted average, where the weights are p, $(1-p)\cdot q$ and $(1-p)\cdot(1-q)$. Hence, the expected RR is:

$$E[RR] = p \cdot RR_{\geq 1} + (1-p)\cdot q \cdot RR_{\leq 0} + (1-p)\cdot(1-q)\cdot RR_{(0,1)}.$$

3.2. Neural Networks

In this section, we present the structure of feedforward neural networks following Hastie et al. (2001) and Günther and Fritsch (2010).

In a neural network, whose structure is presented in Figure 2, neurons are structured in layers. The neurons are connected by synapses, which are graphs between them and the neurons of the subsequent layers. In order to keep the model simple, we consider feedforward neural networks with one hidden layer. The input layer contains all covariates, the so-called input variables $X_1, ..., X_p$, which represent the separate neurons. Each numerical attribute has its own neuron. In case of categorical variables, dummy coding as in a linear regression is applied. The output layer has K neurons $O_1, ..., O_K$. For regression problems with one response variable as well as for classification problems with two categories, we have $K = 1$. For classification problems with C classes, there are $K = C$ output neurons, each representing one category. The hidden layer with neurons $H_1, ..., H_M$ lies in between and cannot be observed directly. A bias can be added to the input and hidden layers as an extra neuron B_I resp. B_H.

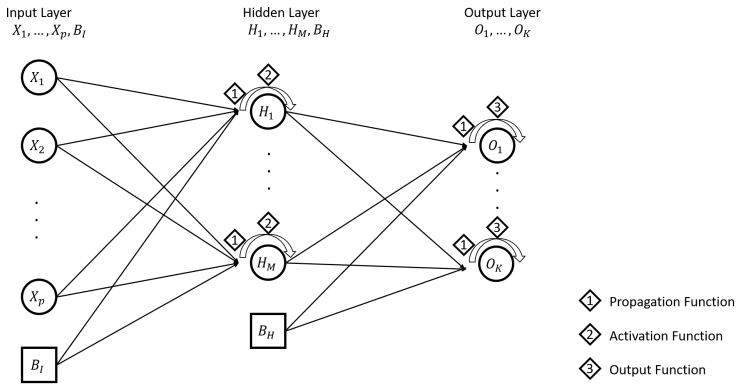

Fig. 2: Structure of neural networks.

The propagation function connects the output values of the previous layer $O_{j,\text{previous layer}}$, $j \in \textit{previous layer}$ such that the result can be used as input $I_{i,\text{current layer}}$ for a neuron $i \in \textit{current layer}$ in the current layer. We use the weighted sum:

$$I_{i,\text{current layer}} = \sum_{j \in \text{previous layer}} w_{i,j} O_{j,\text{previous layer}}.$$

The activation function σ transforms this value $I_{i,\text{current layer}}$ to the output value of the neuron $O_{i,\text{current layer}} = \sigma(I_{i,\text{current layer}})$. For this, we use the sigmoid function:

$$\sigma(x) = \text{sigmoid}(x) = \frac{1}{1 + \exp(-x)}.$$

The propagation function is applied again to receive the input for the output layer. Then, for the output neuron O_k, $k = 1, ..., K$, in the output layer, we apply a final transformation by the

output function g_k instead of the activation function. In case of regression problems, we use as g_k the identity function, whereas we apply the softmax function $g_k(I_k) = \frac{\exp(I_k)}{\sum_{l=1}^{K} \exp(I_l)}$ in case of classification problems.

The weights have to be estimated in the training process. Therefore, we use the backpropagation algorithm in case of neural networks for classification problems as *"backpropagation is the most widely used algorithm for supervised learning with multilayered feed-forward networks"* according to Riedmiller and Braun (1993). In case of regression problems, we use the extension RPROP+ algorithm of Riedmiller and Braun (1993) and refer for more information to their original paper.

3.3. Mixture Models

In a linear regression model, we assume that the dependent variable relates to the covariates by a fixed parameter β over all observations. This assumption is often too restrictive, calling the need for models in which the regression coefficient can change over different clusters among the observations. One family of models are finite mixture models, which will be presented following Frühwirth-Schnatter (2006), Grün et al. (2008), Leisch (2004) and Murphy (2012).

In general, a finite mixture regression model with K components has the form:

$$h(y|\boldsymbol{x}, \boldsymbol{\psi}) = \sum_{k=1}^{K} \pi_k f(y|\boldsymbol{x}, \boldsymbol{\theta}_k), \tag{2}$$

where π_k, $k = 1, ..., K$ are the weights with $\pi_k \geq 0$, $\sum_{k=1}^{K} \pi_k = 1$ and $\boldsymbol{\psi} = (\pi_1, ..., \pi_K, \boldsymbol{\theta}_1', ..., \boldsymbol{\theta}_K')$ is the vector of all unknown parameters. $\boldsymbol{\theta}_k$ denotes the component specific parameter vector for the density function f. If f is a univariate normal density with component specific mean $\boldsymbol{\beta}_k'\boldsymbol{x}$ and variance σ_k^2, we get a mixture of standard linear regression models with $\boldsymbol{\theta}_k = (\boldsymbol{\beta}_k', \sigma_k^2)'$.

The weights π_k, $k = 1, ..., K$ in Equation (2) are usually independent of the covariates. One extension is the concomitant variable model by Grün et al. (2008), which assumes that the weights depend on some variables, the so-called concomitant variables denoted by $\boldsymbol{c}$. Then, the mixture model can be written as:

$$h(y|\boldsymbol{x}, \boldsymbol{\psi}) = \sum_{k=1}^{K} \pi_k(\boldsymbol{c}, \boldsymbol{\alpha}) f(y|\boldsymbol{x}, \boldsymbol{\theta}_k), \tag{3}$$

where $\boldsymbol{\alpha}$ denotes the parameter vector of the concomitant variables and $\boldsymbol{\psi}$ contains all parameters including $\boldsymbol{\alpha}$. The remaining arguments are defined as in Equation (2) and the weights have to satisfy the conditions $\pi_k(\boldsymbol{c}, \boldsymbol{\alpha}) > 0$ and $\sum_{k=1}^{K} \pi_k(\boldsymbol{c}, \boldsymbol{\alpha}) = 1$, $k = 1, ..., K$. Similar to Grün et al.

(2008), we assume a multinomial logit model for the weights π_k, which can be written as:

$$\pi_k(\boldsymbol{c}, \boldsymbol{\alpha}) = \frac{\exp(\boldsymbol{c}'\boldsymbol{\alpha}_k)}{\sum_{u=1}^{K} \exp(\boldsymbol{c}'\boldsymbol{\alpha}_u)}, \tag{4}$$

for all $k = 1, ..., K$ and with $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_k')'_{k=1,...,K}$ and $\boldsymbol{\alpha}_1 \equiv 0$.

For parameter estimation, we write the log-likelihood function of a sample of n observations $(\boldsymbol{x}_1, y_1), ..., (\boldsymbol{x}_n, y_n)$ as:

$$\log L = \sum_{i=1}^{n} \log h(y_i|\boldsymbol{x}_i, \boldsymbol{\psi}) = \sum_{i=1}^{n} \log \left(\sum_{k=1}^{K} \pi_k(\boldsymbol{c}, \boldsymbol{\alpha}) f(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}_k) \right).$$

Since the membership to the components is unknown, this likelihood function cannot be computed directly. We use for maximum likelihood estimation the iterative Expectation-Maximation (EM) algorithm introduced in Dempster et al. (1977). It is implemented for the concomitant mixture models in the R-package `flexmix`.

4. Data

4.1. The Global Credit Data (GCD) database

As Krüger and Rösch (2017), we use a dataset of US-based small- and medium-sized entities (SME) from Global Credit Data (GCD) for our empirical analysis. GCD is a dutch-based, not-for-profit registered association whose owners are more than 50 Member-banks across the world. The objective of GCD is to be a credit risk data pooling initiative to support the Member-banks by their internal credit risk models inter alia for the advanced internal ratings-based approach of Basel II. We use the LGD&EAD platform, which is the worldwide largest loss and recovery dataset for commercial loans, and contains data relating to credit defaults since 1998 until the end of 2016. This time period encompasses more than one full economic cycle as required by §472 in Basel Comittee on Banking Supervision (2004). Table 9 in the Appendix gives an overview over all variables used.

We adjust the data following Höcht and Zagst (2008). First, the exposure at default has to be strictly greater than zero as the focus of this study lies on real losses. Second, we only consider loans where EAD + Principal Advance + Financial Claim $\geq$ €5,000, such that very small exposures are excluded. Third, the default date lies in the interval [January 2002, December 2015]. We exclude cases before the year 2002 due to modified banking regulations. As the cases after 2015 might still be unresolved, we exclude them as well. Fourth, to exclude all facilities that are not fully resolved or exhibit unreasonable cash flows, the following rule is applied according

to Höcht and Zagst (2008): If the total sum of all reported cash flows (including charge-offs and waivers) divided by the outstanding amount at default is smaller than 90% or greater than 105%, the facility is not considered. Fifth, only cases with resolved default status are of interest. Finally, the RR lies in the interval $[-0.5, 1.5]$. All observations with smaller or greater RR are excluded to avoid outliers.

Furthermore, we split the data into 3 groups: training, validation and test set. The training set, in regression problems the so-called in-sample set, contains 80% of the data according to Murphy (2012) and is used to estimate the models. In order to get an impression how well a model can create new predictions, the trained models are applied on the test set, which is also called out-of-sample set. This data is not used in the estimation of the model and therefore, these results are reliable and can be compared. Some models need hyperparameters, for example the number of hidden neurons in a neural network. Since the training data is already used and the test data should remain independent of the modeling process, we use a third dataset, the validation set, to fit the hyperparameters. The test set as well as the validation set both contain 10% of the data.

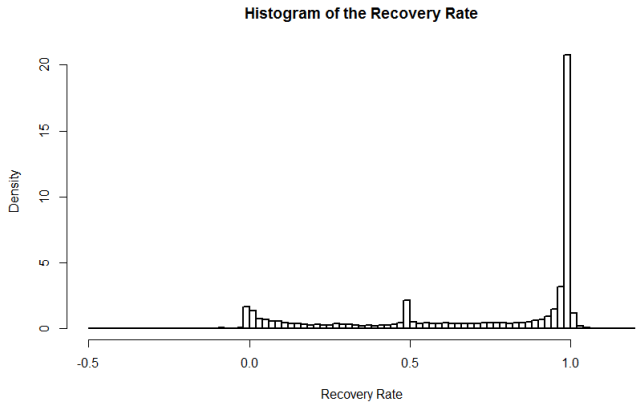

Fig. 3: Histogram of the recovery rate in our data (SME, US-based).

The histogram of the RR is presented in Figure 3 and shows a high concentration at full recovery. Furthermore, there are two additional peaks near 0 and 0.5. In literature, the RR has frequently been modeled using a bimodal structure, for example in the studies of Altman and Kalotay (2014), Bastos (2010a), Bastos (2010b) and Qi and Zhao (2011). Similar to our data, Ye and Bellotti (2019) use a trimodal distribution.

4.2. Predictive Crisis Indicator

Some studies, for example Calabrese (2012) and Höcht and Zagst (2008), find that the recovery rate tends to be lower during economic downturns. Brumma and Winckle (2017) observe that the macroeconomic behavior during the resolution time has an influence on the recovery. Therefore, we use a predictive crisis indicator, which indicates whether a crisis might occur in the next 18 months (the average resolution time).

To model the predictive Crisis Indicator, we first calculate a daily Crisis Indicator using a modified version of the algorithm of Ernst et al. (2009), where we use two-year highs instead of half-year highs. The algorithm of Ernst et al. (2009) can be applied to any stock index, but for the focus on recovery rates of SMEs in the US, we chose S&P500. With the Ernst et al. (2009) algorithm, a daily Crisis Indicator is determined. To get a monthly aggregated Crisis Indicator, we apply the following decision rule: If at least 2 days within a month are indicated as crisis, the month in total is considered as crisis.

In the next step, a predictive Crisis Indicator needs to be built. For every month m, we consider the period of the next 18 months $[m+1, ..., m+1+18]$. If there is at least one month in crisis, the predictive Crisis Indicator for m is set to 1 (indicating a crisis).

Up to this point, the calculations are made on historical data and the predictive Crisis Indicator can only be obtained, once the data for the next 18 months is available. Since the goal of this study is to predict the RR at the date of default, the required information is not yet available. Therefore, the predictive Crisis Indicator has to be modeled. For this, we set up a logistic regression model with macroeconomic data and Table 1 shows the included attributes and their impact.

Variable	Positive (+) or negative (-) impact	Description
(Intercept)	+	
Implied Volatility	+	Implied Volatility
$(\text{TEDRATE})^2$	+	TED Spread^2
TCU	-	Capacity Utilization Rate for Total Industry
FEDFUNDS	+	Effective Federal Funds Rate
FEDFUNDS : $(\text{TEDRATE})^2$	-	Interaction between FEDFUNDS & $(\text{TEDRATE})^2$

Variable	Positive (+) or negative (-) impact	Description
T10Y3MM	+	Spread between 10-Year treasury Constant Maturity and 3-Month Treasury Constant Maturity
$(\text{OECD\_6NME})^2$	+	Composite Leading Indicator given by the OECD
log(UNRATE)	-	Civilian Unemployment Rate
OECD_6NME	-	Composite Leading Indicator given by the OECD

Table 1: Estimated logistic regression model for the Crisis Indicator.

In this paper, the Crisis Indicator is used in different ways:

(C1) We do not include the crisis information at all.

(C2) The predicted Crisis Probability calculated from the logistic regression model is included as a covariate.

(C3) The Crisis Indicator is included as a covariate.

(C4) We split the data into crisis and non-crisis dataset and train the models on each subset.

5. Empirical Results

The focus of this study lies on mixture regression models and thus, we only briefly present the best results of the regression models, decision trees and neural networks and subsequently, we concentrate on the mixture models. We also give an overall comparison of all models.

In order to decide for the best model, we use the mean squared error (MSE) measure of fit, defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2,$$

where $y_i, i = 1, ..., n$ are the observed RRs and $\hat{y}_i$ are the estimated RRs. This measure is also used, e.g., in Calabrese (2012), Gupton and Stein (2005) and Ye and Bellotti (2019).

5.1. Regression Models

First, we consider the results of the regression models. We apply stepwise selection for model selection in the different regression problems based on the BIC, as it penalizes model complexity

on a larger scale compared to the AIC. We will use this model selection criterion in every regression problem in the following.

We concentrate on the linear regression model including the Crisis Indicator as well as the models trained on the crisis and non-crisis subsets, since the results of these models performed best (see Table 6). Regarding the included covariates within the linear models, which are presented in Table 2, we recognize that in case of crisis, the RR is only determined by the information whether a guarantee or collateral is given and the size of EAD. Moreover, in the crisis case, the Collateral Indicator has an impact on the RR, whereas the Primary Industry Code as well as the Utilization Rate only have an impact in the non-crisis case. However, the linear regression model with the Crisis Indicator includes all these attributes and additionally the variable Nature of Default.

Variable	LR on crisis subset	LR on non- crisis subset	LR incl. Crisis Indicator	QR incl. Crisis Indicator
Crisis Indicator			x	x
Country of Business				x
Leveraged Finance Indicator				x
Operating Company Indicator				x
Primary Industry Code		x	x	x
Nature of Default			x	x
Collateral Rank of Security				x
Guarantor Rating Moodys				x
Guarantee Indicator	x	x	x	x
Collateral Indicator	x		x	x
log(EAD)	x	x	x	x
Utilization Rate		x	x	x

Table 2: Included attributes in the linear regression models (LR) on the crisis/non-crisis subset resp. the entire subset including the Crisis Indicator and in the quantile regression model (QR) including the Crisis Indicator.

In case of the quantile regression model, we regress the median in order to compare the results of the quantile regression to the results of the linear regression. The quantile regression including

the Crisis Indicator outperforms the remaining models considering MSE (see Table 6). Therefore, we have a look at the included variables of this model and recognize that the model selection results in more attributes for the quantile regression as in case of the linear regression models. In particular, the variables Country of Business, Leveraged Finance Indicator, Operating Company Indicator, Collateral Rank of Security as well as Guarantor Rating Moodys are only included in the quantile regression model, whereas the remaining covariates are part of both models.

5.2. Decision Tree

We use a decision tree approach in order to apply the beta transformation on the RR. Similar to the regression models in Section 5.1, we use stepwise selection with BIC for model selection.

Besides the logistic regression, a neural network can be applied for the categorization problems. Therefore, we use the function `nnet` of the R-package `nnet`. As input, all available attributes are used. The network is trained by the backpropagation algorithm and only has one hidden layer for simplicity. Furthermore, the optimal number of hidden neurons is tested. Therefore, the network is trained on the training data with different numbers of hidden neurons from 1 to 10. The prediction error of the validation set is the selection criteria and results for both classification problems, whether the RR is greater than or equal to 1 and whether the RR is smaller than or equal to 0, in one hidden neuron. †

We first remark that we report the trees with the quantile regression for the median applied on the raw RR in the open unit interval, because they outperform those on the beta transformed RR. One reason for this might be that the RR in our dataset is trimodal and the beta distribution might not fit well.

Table 3 gives an overview which covariates are included in the different regression problems. We focus in the unit interval on the linear regression model without crisis information and the quantile regression including the Crisis Probability, as the results of the decision trees including these models outperform the other decision trees. Similar to the regression problems above, the quantile regression model including the Crisis Probability contains more variables than the linear regression model. Furthermore, the Crisis Indicator affects only the logistic regression to decide whether the RR attains a value smaller than or equal to zero.

†In general, a neural network with one hidden neuron and the sigmoid function as activation function equals a logistic regression model. Since the estimation method is different (backpropagation algorithm in case of a neural network and maximum likelihood estimation in case of a logistic regression), the parameters of the two models can be different.

Variable	LR without crisis info for $RR \in (0,1)$	QR incl. Crisis Probability for $RR \in (0,1)$	LR for $RR \geq 1$	LR for $RR \leq 0$	Log. Reg. $RR \geq 1$	Log. Reg. $RR \leq 0$
Crisis Indicator						x
Crisis Probability		x				
Country of Business	x	x			x	
Public Private Indicator			x	x	x	
Leveraged Finance Indicator	x	x				
Operating Company Indicator	x	x				
Primary Industry Code		x				
Nature of Default		x		x	x	x
Lender Issued Amount				x		
Collateral Rank of Security	x	x				
Guarantor Rating Moodys		x				
Guarantee Indicator	x	x				x
Collateral Indicator	x	x	x			
Entity Sales			x			
log(Number of Loans)					x	
log(EAD)	x	x				x
Utilization Rate	x	x	x		x	

Table 3: Included attributes in the regression models of the decision trees.

5.3. Neural Network

Another possibility to model the RR are neural networks whose results are shown in the following. We begin with the description of the predetermined model parameters and present the results in

Table 6.

In order to train neural networks for regression problems, we use the function `neuralnet` of the R-package `neuralnet`, which applies the RPROP+ algorithm. Therefore, we set the multiplication factors for the upper and lower learning rate to $\eta^- = 0.5$ and $\eta^+ = 1.2$, the parameter threshold to 0.01 and the maximum number of iterations to $1e7$. We use the sigmoid activation function, the identity as output function and the sum of squared errors as error function.

For reasons of simplicity, all neural networks contain one hidden layer and the optimal number of hidden neurons is determined by minimizing the MSE on the validation set.

We use all available attributes as input variables and do not apply any model selection in advance, because the neural network should identify the important variables by its own. For categorical covariates, dummy variables are created just like for the regression problems. According to Lantz (2015), we scale the metric data to the unit interval $[0, 1]$ as it is not normally distributed.

5.4. Mixture Regression Models

In this subsection, we present the results of the mixture regression models from Section 3.3. At first, we apply the model selection to identify relevant covariates. Subsequently, the results of the mixture models with and without concomitant variables are shown.

5.4.1. Model Parameters

Our motivation to investigate mixture regression models stems from the observation of multiple modes. Since the RR is trimodal in our dataset, our mixture regression models have three components. We use the package `flexmix` in R to fit our models. Unfortunately, there is no model selection implemented. In addition, the EM-algorithm to fit the model does not converge for every possible combination of input variables. For a pre-selection and in order to reduce the overall number of covariates, we use the input variables of Krüger and Rösch (2017), who base their study on loan data of SMEs in the US provided by GCD. We use our crisis information instead of macroeconomic data and focus our analysis on entity level, hence we can not use all attributes of Krüger and Rösch (2017). In conclusion, the resulting variables are log-transformed EAD, Guarantee Indicator, Collateral Indicator, Primary Industry Code as well as the Crisis Indicator resp. Probability. Subsequently to this pre-selection, all possible combinations of

variables are formed and it is tested whether the EM-algorithm converges. We compare the mixture regression models with the BIC and the best model contains the covariates Collateral Indicator, log-transformed EAD and Crisis Probability. For simplicity reasons, this combination of variables is also used as concomitant variables in the following. We notice that it is better to use the Crisis Probability than the Crisis Indicator in this method.

The package `flexmix` provides information about the standard error as well as z- and p-value for every coefficient in every component. In case of a negative entry in the diagonal of the variance-covariance matrix, the standard error can not be computed. This is partially the case for the coefficient of the log-transformed attribute EAD.

Therefore, we exclude this attribute for the regression problems, but it is still part of the multinomial model. In order to distinguish the different models, we refer to the model including the EAD as covariate by the name "M1" and denote the model without the EAD as "M2." If, additionally to the components, the probabilities to belong to the components are regressed, we denote the models "M1C" and "M2C," since the attributes included in the multinomial models are called concomitant variables.

5.4.2. Model Description

Comp. 1

	M1	M2	M1C	M2C
(Intercept)	0.9967	0.9920	0.9796	0.8051
Collateral Indicator: No	-0.0045	-0.0044	-0.0111	0.0151
Collateral Indicator: Yes	0.0030	0.0014	-0.0057	-0.0285
log(EAD)	-0.0003	-	-0.0004	-
Crisis Probability	-0.0008	-0.0012	-0.0055	0.0006
Sigma	0.0052	0.0077	0.0287	0.1818

Comp. 2

	M1	M2	M1C	M2C
(Intercept)	0.5158	-0.0143	0.9978	0.9921
Collateral Indicator: No	-0.0087	0.0280	-0.0057	-0.0045
Collateral Indicator: Yes	0.4687	0.0362	0.0015	0.0014
log(EAD)	-0.0007	-	-0.0003	-

Crisis Probability	-0.0089	0.0206	-0.0007	-0.0008
Sigma	0.0195	0.0366	0.0053	0.0074

Comp. 3

	M1	M2	M1C	M2C
(Intercept)	-0.3386	0.6348	-0.3593	0.3214
Collateral Indicator: No	-0.2457	-0.1046	-0.1549	-0.0410
Collateral Indicator: Yes	-0.2637	0.0465	-0.2147	-0.2223
log(EAD)	0.0814	-	0.0763	-
Crisis Probability	-0.0070	-0.0505	0.0172	0.0720
Sigma	0.3077	0.2702	0.2858	0.1744

Table 4: Summary of the estimated mixture regression models.

Table 4 displays the results of the estimated models. Firstly, we consider model M1. The first component of M1 is mainly determined by the intercept near one, since the remaining coefficients are vanishing. Furthermore, the intercept at 0.5 as well as the attribute Collateral Indicator have the most impact on the second cluster. In contrast to the other clusters, the characteristic Yes of the Collateral Indicator has a negative impact on the third cluster. Moreover, the influence of the characteristic No of the Collateral Indicator as well as the Crisis Probability is negative. In addition, the EAD has a slightly positive impact. The third component has the highest fluctuations (represented in a Sigma of 0.308).

Having a look at M2, we notice that the first component is mainly influenced by the intercept near one and the attributes have little impact. In comparison to the control group Unknown, the categories Yes and No of the Collateral Indicator have an positive impact on the second component. As the parameter of Yes of the Collateral Indicator is higher than the coefficient for No, we would expect higher values for entities having a collateral. It is counterintuitive that this component increases its value if the Crisis Probability increases. The highest values of the third component are expected in case of a collateral, whereas the lowest values will be attained when there is no collateral. In addition, the value of this cluster will be higher if the Crisis Probability is small. The attributes have the highest impact on the third cluster due to their higher absolute values. Sigma with a value of 0.270 underlines this finding, as it is higher than the sigma of the first and second cluster.

The first as well as the second component of the mixture regression model M1C are mainly determined by the intercept near one, as the coefficients of the covariates are near zero and influence them little. Sigma of the third component is the highest, indicating higher fluctuations. The characteristics of the Collateral Indicator influence the third component negative due to their negative parameters. Moreover, the log-transformed EAD has a positive coefficient, which indicates that the value of the third cluster correlates to the EAD. In addition, a higher Crisis Probability results in a higher value for this component which is counterintuitive.

In model M1C, we regress the probabilities that an observation belongs to a certain component. In this study, a multinomial logit model is assumed for the weights π_k, $k = 1, ..., K$ as depicted in Equation (4). One assumption of this model is $\alpha_1 \equiv 0$. Therefore, only the parameters of the second and third component are given in Table 5.

Comp. 2

	M1C	M2C
(Intercept)	1.5822	3.6216
Collateral Indicator: No	1.1220	0.8513
Collateral Indicator: Yes	-0.1528	-0.0661
log(EAD)	-0.0345	-0.2217
Crisis Probability	-0.4787	-0.5281

Comp. 3

	M1C	M2C
(Intercept)	2.2059	7.0839
Collateral Indicator: No	1.4026	2.0530
Collateral Indicator: Yes	-0.1795	0.2222
log(EAD)	-0.1054	-0.6220
Crisis Probability	0.2783	0.3027

Table 5: Summary of the estimated concomitant regression models.

The probability of belonging to the second component increases if no collateral is available and decreases if a collateral is given. Moreover, a borrower with lower EAD (resp. a lower Crisis Probability) is expected to have a higher probability of belonging to the second cluster. Furthermore, the probability that an entity belongs to the third cluster is expected to be lower

in case of a collateral and higher in case of no collateral. The log-transformed EAD has again a negative influence, whereas the Crisis Probability has a positive one.

In model M2C, we would expect that the value of the first component is higher in case of no collateral than in case of a collateral. In addition, a higher Crisis Probability will lead to higher values of the first component. This cluster has the highest variation which is displayed in the high value of Sigma of 0.182. The covariates have little impact on the second component, which is mainly determined by its intercept near one. Moreover, the value of the third cluster is lower for an observation with a collateral than for one without any collateral. Furthermore, the third component is expected to attain a higher value for a higher Crisis Probability.

The probability that an observation belongs to the second cluster increases if it has no collateral. However, having a collateral decreases the probability. In addition, the higher the log-transformed EAD or the higher the Crisis Probability, the lower the probability that an observation belongs to the second cluster. Furthermore, the probability that an observation belongs to the third cluster reaches a maximum if there is no collateral. Moreover, we would expect a lower probability that an entity belongs to the third cluster if the EAD is high or the Crisis Probability is low.

5.5. Comparison of all models based on MSE

Table 6 shows the in-sample as well as the out-of-sample results for all models including the linear regression model with different assumptions (C1),...,(C4) on the crisis information. For the linear regression, MSE prefers in-sample the model including the Crisis Indicator and out-of-sample to separate the data into crisis and non-crisis subsets. Since the models on the split data give more insights in the determinants of the RR in crisis and non-crisis case, this approach might be preferred, as the goodness-of-fit of the models is similar.

	In-sample				Out-of-sample			
	(C1)	(C2)	(C3)	(C4)	(C1)	(C2)	(C3)	(C4)
LR	0.1095	**0.1085**	0.1086	0.1087	0.1095	0.1089	0.1089	**0.1082**
QR	0.1382	**0.1310**	0.1312	0.1377	0.1418	**0.1352**	0.1353	0.1404
DT LR LogReg	0.1432	0.1435	0.1434	0.1457	0.1458	0.1465	0.1464	0.1475
DT LR NN	**0.1395**	0.1407	0.1407	0.1431	**0.1417**	0.1432	0.1432	0.1441
DT QR LogReg	0.1546	0.1504	0.1504	0.1590	0.1575	0.1542	0.1540	0.1624
DT QR NN	0.1513	**0.1476**	**0.1476**	0.1569	0.1541	0.1513	**0.1511**	0.1596

	In-sample				Out-of-sample			
	(C1)	(C2)	(C3)	(C4)	(C1)	(C2)	(C3)	(C4)
NN	0.0471	0.0467	**0.0454**	0.0995	0.1032	0.1011	**0.0981**	0.1010
M1			0.0375				0.0379	
M2			**0.0257**				**0.0268**	
M1C			0.0323				0.0326	
M2C			**0.0101**				**0.0107**	

Table 6: In-sample and out-of-sample results for the estimated linear regression (LR), quantile regression (QR) models, decision trees (DT) with linear or quantile regression in the unit interval (LR/QR) and logistic regression or neural network for the classification problems (LogReg or NN), neural networks (NN) and mixture regression models.

Having a look at the quantile regression, the model including the Crisis Indicator outperforms the remaining models considering MSE. Moreover, the linear regression models outperform the quantile regressions. This fact might be explained by the different optimization problems. The estimation method of the linear regression minimizes the least squares error, whereas the quantile regression for the median minimizes the mean absolute error. However, the quantile regression models give more insights into the structure of the distribution, since different quantiles can be modeled. We refer to Krüger and Rösch (2017) who calculate further quantile regressions for several quantiles.

Comparing the decision trees by MSE, the models with the linear regression in $(0, 1)$ outperform those with the quantile regression. Moreover, it is preferable to not include any information about a crisis for the decision tree with the linear regression in the unit interval, whereas the decision tree with quantile regression including the Crisis Probability outperforms the other decision trees with quantile regression. Additionally, the trees with the neural networks for classification result in lower MSEs than the models with the logistic regressions. This finding might be explained by the slightly lower prediction error of the neural network for the classification whether the RR attains values smaller than or equal to 0. However, the differences in the MSE are marginal and since the logistic regression gives some more insight in the determinants of the RR, it might be preferable to use them. We compare the results of the decision tree approach with the regressions on the entire dataset and recognize that the models on the entire data result in lower MSEs than the decision trees.

Regarding the results of the neural networks, we notice that the network including the Crisis Probability outperforms in-sample as well as out-of-sample the other neural networks. Furthermore, we recognize that the MSE is in-sample small, whereas the results are out-of-sample similar to the MSE of the logistic regression models. One reason for the difference in the MSE between the in-sample and out-of-sample subset might be overfitting.

Finally, we consider the results of the mixture regression models. The models excluding EAD as covariate are superior to the mixture regression models including the covariate EAD. In addition, the models which regress the densities as well as the probabilities outperform the mixture models with fixed probabilities. In conclusion, model M2C is the best model.

In-sample as well as out-of-sample, the mixture regression models outperform the regressions as well as the neural networks. One reason for this might be that the mixture regression model can display the different modes better than the other models.

5.6. Practical consequences from the best models

In the following, we compare the three best models: The mixture regression model M2C, the neural network including the Crisis Probability and the linear regression model with separate subsets for crisis and non-crisis.

We investigate the difference d_i between the predicted RRs $\hat{R}_i$ and the observed RRs $R_i^{obs.}$:

$$d_i = \hat{R}_i - R_i^{obs.},$$

for $i = 1, ..., n$ where n is the number of observations. From the risk-managers point of view, a situation in which the RR is conservatively underestimated is favorable compared to a situation in which the RR is overestimated.

Therefore, we are interested in the number of observations where the difference between the predicted RR and the observed RR exceeds a certain threshold $\theta \in \{0.1, ..., 0.9\}$ proportional to the overall number of observations:

$$\frac{\#\{d_i > \theta\}}{n}.$$

	$\frac{\#\{d_i>0.1\}}{n}$	$\frac{\#\{d_i>0.2\}}{n}$	$\frac{\#\{d_i>0.3\}}{n}$	$\frac{\#\{d_i>0.4\}}{n}$	$\frac{\#\{d_i>0.5\}}{n}$	$\frac{\#\{d_i>0.6\}}{n}$	$\frac{\#\{d_i>0.7\}}{n}$	$\frac{\#\{d_i>0.8\}}{n}$	$\frac{\#\{d_i>0.9\}}{n}$
LR	0.2975	0.2503	0.2117	0.1740	0.1310	0.0833	0.0316	0.0108	0.0013
NN	0.2859	0.2447	0.1960	0.1467	0.0937	0.0575	0.0267	0.0095	0.0015
M2C	0.1379	0.0386	0.0085	0.0009	0.0001	0.0000	0.0000	0.0000	0.0000

Table 7: In-sample results for the difference between the estimated and observed RR.

The results are presented in Table 7 for the in-sample and in Table 8 for the out-of-sample data.

	$\frac{\#\{d_i>0.1\}}{n}$	$\frac{\#\{d_i>0.2\}}{n}$	$\frac{\#\{d_i>0.3\}}{n}$	$\frac{\#\{d_i>0.4\}}{n}$	$\frac{\#\{d_i>0.5\}}{n}$	$\frac{\#\{d_i>0.6\}}{n}$	$\frac{\#\{d_i>0.7\}}{n}$	$\frac{\#\{d_i>0.8\}}{n}$	$\frac{\#\{d_i>0.9\}}{n}$
LR	0.2999	0.2613	0.2263	0.1858	0.1288	0.0929	0.0359	0.0110	0.0000
NN	0.3008	0.2613	0.2061	0.1527	0.0938	0.0607	0.0304	0.0092	0.0009
M2C	0.1398	0.0414	0.0101	0.0009	0.0000	0.0000	0.0000	0.0000	0.0000

Table 8: Out-of-sample results for the difference between the estimated and observed RR.

In-sample, the mixture regression model M2C overestimates the true RR by more than $\theta = 0.1$ in 14% of all cases, whereas the linear regression model as well as the neural network overestimates the RR even in 29% of all observations. Having a look at $\theta = 0.2$, we recognize that the mixture regression model M2C only overestimates the observed RR in 3.9% of all cases. The results of the linear regression as well as the neural network are worse, since 25% of all cases predict a RR which exceeds the true RR by more than $\theta = 0.2$. In addition, there is no observation where the predicted RR of the mixture regression model exceeds the true RR by more than $\theta = 0.6$. Exemplary, if we estimate a RR of 1, the true value is bigger than 0.4. Thus, for a case where a full recovery is predicted, we know that at most 40% of the exposure at default will be recovered. In case of the linear regression model as well as the neural network, there are cases where the predicted RR overestimates the true RR by more than $\theta = 0.9$. We refer to the same example as above. If the estimated RR is 1, the true RR can be smaller than 0.1, which is almost a total loss even though the model predicts a full recovery. Moreover, we notice that the behaviour of the linear regression model and the neural network is similar.

The out-of-sample results are similar to the in-sample results. The behaviour of the linear regression model equals the behaviour of the neural networks. Moreover, regarding the maximum difference, the out-of-sample results for the mixture regression model M2C are even slightly better than in-sample, since there is no prediction which overestimates the true RR by the value $\theta = 0.5$. Similar to the in-sample results, the mixture regression model overestimates the true RR by more than $\theta = 0.1$ in 14% of all cases, whereas the linear regression model as well as the neural network exceed the observed RR in 30% of all observations. In addition, the true RR is overestimated by more than $\theta = 0.2$ in only 4.1% of all observations in case of the mixture regression model and in 26% of all observations in case of the linear regression model as well as the neural network. Since

the maximum difference is smaller in case of the mixture regression model and the predicted RR exceeds the observed RR by 0.1 resp. 0.2 in only 14% resp. 4.1% of all cases instead of 30% resp. 26% of all cases, we conclude that the mixture regression model M2C outperforms the neural network as well as the linear regression model.

6. Summary and Conclusion

We compared different models to predict the RR; namely regression methods, decision trees, neural networks and mixture regression models. Additionally, we investigated how information on an economic crisis can be introduced into the models.

For our analysis, we considered a dataset of US-based SMEs obtained from GCD. We use the definition of the workout RR. Empirical RRs exhibit a multimodal structure with three modes at 0, 0.5 and 1. Since earlier studies in literature point out that an economic crisis during the time to resolution has an impact on the RR, we use a predictive Crisis Indicator (resp. Crisis Probability).

The best models are in-sample as well as out-of-sample the mixture regression models, especially the concomitant variable model which regresses the density as well as the probability that an observation belongs to a certain cluster. We find by model selection with the BIC that including the Crisis Probability is preferable compared to including the Crisis Indicator. The neural network outperforms in-sample the linear regression model, but the results are similar out-of-sample. The quantile regression models lead to higher MSEs than the linear regression models. Decision trees performed worst in our study.

Concluding, let us propose some areas for future research in predicting and modelling the RR. In the present study, the RR can take values greater than one as well as smaller than zero and we conclude that the mixture regression models outperform the other models. Since most of the studies consider observations with RR in the unit interval, the question arises whether the mixture regression models also outperform the other methods in such a restricted dataset. In addition, there are some parameters which can be modified. For regression models, we do not consider interactions, yet. Moreover, the number of hidden layers in neural networks can be adjusted. Additionally, different activation and error functions can be tested.

A. Attributes in this study

Table 9 gives an overview over all variables from GCD that were used.

Attributes	Type	Reference Category	Borrower (B) / Loan (L) Level?	Aggregation for Loan Level
Country of Business	categorical	Unknown	B	
Public Private Indicator	categorical	Unknown	B	
Primary Industry Code	categorical	Unknown	B	
Leveraged Finance Indicator	categorical	No	B	
Operating Company Indicator	categorical	Unknown	B	
Incomplete Portfolio	categorical	No	B	
Number of Loans	metric		B	
Nature of Default	categorical	Unknown	B	
Lender Issued Amount	categorical	No	B	
Entity Sales	metric		B	
EAD	metric		B	
Crisis Indicator resp. Crisis Probability	categorical	No	B	
Collateral Indicator	categorical	Unknown	L	at least one collateral
Guarantee Indicator	categorical	No	L	at least one guarantee
Default Amount	metric		L	Sum
Utilization Rate	metric		L	Median
Guarantor Rating Moodys	categorical	Unknown	L	worst rating
Rank of Collateral	categorical	Unknown	L	worst rank

Table 9: Attributes of the data.

References

Altman, E. I. and Kalotay, E. A. (2014). Ultimate recovery mixtures. *Journal of Banking & Finance*, 40:116–129.

Basel Comittee on Banking Supervision (2004). Basel II: International convergence of capital measurement and capital standards: A revised framework. *Bank for International Settlements, Basel*.

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

Bastos, J. A. (2010b). Predicting bank loan recovery rates with neural networks. *CEMAPRE Working Papers*, (1003).

Bellotti, A. and Crook, J. (2009). Loss given default models for uk retail credit cards. *Building*, 1–28.

Brumma, N. and Winckle, P. (2017). Gcd downturn lgd study 2017. *SSRN Electronic Journal*.

Calabrese, R. (2012). Predicting bank loan recovery rates with a mixed continuous-discrete model. *Applied Stochastic Models in Business and Industry*, 30:99–114.

Calabrese, R. (2014). Downturn loss given default: Mixture distribution estimation. *European Journal of Operational Research*, 237:271–277.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38.

Dermine, J. and Carvalho, C. N. D. (2006). Bank loan losses-given-default: A case study. *Journal of Banking & Finance*, 30:1219–1243.

Ernst, C., Grossmann, M., Höcht, S., Minden, S., Scherer, M., and Zagst, R. (2009). Portfolio selection under changing market conditions. *International Journal of Financial Services Management*, 4(1):48–63.

Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. (2013). *Regression: Models, Methods and Applications*. Springer-Verlag, Berlin.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models.* Springer-Verlag New York.

Günther, F. and Fritsch, S. (2010). Neuralnet: Training of neural networks. *R Journal*, 2.

Grün, B., Kepler, J., and Leisch, F. (2008). Flexmix version 2: Finite mixtures with concomitant variables and varying and constant parameters. *Journal of Statistical Software.*

Gupton, G. and Stein, R. (2005). Losscalc v2: Dynamic prediction of LGD - modeling methodology. *Moody's/KMV*

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning.* Springer Series in Statistics. Springer New York Inc., New York, NY, USA.

Höcht, S. and Zagst, R. (2008). Loan recovery determinants: A pan-european study. *Working Paper.*

Hosmer, D. W. and Lemeshow, S. (2013). *Applied logistic regression.* John Wiley and Sons.

Keijsers, B., Diris, B., and Kole, E. (2015). Cyclicality in Losses on Bank Loans. (15-050/III).

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

Lantz, B. (2015). *Machine Learning with R.* Packt Publishing, 2nd edition.

Leisch, F. (2004). Flexmix: A general framework for finite mixture models and latent class regression in r. *Journal of Statistical Software, Articles*, 11(8):1–18.

Loterman, G., Brown, I., Martens, D., Mues, C., and Baesens, B. (2012). Benchmarking regression algorithms for loss given default modeling. *International Journal of Forecasting*, 28:161–170.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective.* MIT Press.

Qi, M. and Zhao, X. (2011). Comparison of modeling methods for Loss Given Default. *Journal of Banking & Finance*, 35:2842–2855.

Riedmiller, M. and Braun, H. (1993). A direct adaptive method for faster backpropagation learning: The rprop algorithm. In *IEEE INTERNATIONAL CONFERENCE ON NEURAL NETWORKS*, 1:586–591.

Tomarchio, S. D. and Punzo, A. (2019). Modelling the loss given default distribution via a family of zero-and-one inflated mixture models. *Journal of the Royal Statistical Society Series A*, 182(4):1247–1266.

Wang, H., Forbes, C. S., Fenech, J.-P., and Vaz, J. (2018). The determinants of bank loan recovery rates in good times and bad – new evidence. *Monash Econometrics and Business Statistics Working Papers*, (7/18).

Yao, X., Crook, J., and Andreeva, G. (2015). Support vector regression for loss given default modelling. *European Journal of Operational Research*, 240(2):528–538.

Ye, H. and Bellotti, A. (2019). Modelling recovery rates for non-performing loans. *Risks*, 7:19.