Advanced Methods for Loss Given Default Estimation

Inauguraldissertation
zur
Erlangung des Doktorgrades
der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der
Universität zu Köln

2015

vorgelegt
von

Dipl.-Wirt.-Math. Eugen Töws

aus
Klimowka

Referent:	Univ.-Prof. Dr. Thomas Hartmann-Wendels, Universität zu Köln
Korreferent:	Univ.-Prof. Dr. Dieter Hess, Universität zu Köln
Vorsitz:	Univ.-Prof. Dr. Heinrich R. Schradin, Universität zu Köln

Tag der Promotion: 11. Januar 2016

Acknowledgments

The completion of this thesis would not have been possible without the support from many people, whom I would like to thank.

Foremost, I would like to express my sincere gratitude to my advisor Professor Dr. Thomas Hartmann-Wendels for his continuous support, patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I would also like to thank him for the opportunity to present my work at several international conferences. Further thanks go to Professor Dr. Dieter Hess for serving as co-referee and Professor Dr. Heinrich Schradin for chairing the dissertation committee.

I am greatly indebted to Patrick Miller, who partnered with me on the thesis' subject from the first day. I am especially grateful for the numerous highly productive discussions with him that resulted in two co-authored projects. I very much appreciate conducting joint research with him. I want to thank Dr. Hans Elbracht for his exceptional work on collecting the data employed in this thesis without which this thesis would not have been.

Many thanks go to my former colleagues at the University of Cologne. In particular, I would like to mention Patrick Wohl, Dr. Tobias Schlüter, Dr. Claudio Wewel, Benjamin Döring, David Fritz, and Elke Brand as well as Dr. Claudia Wendels and Dr. Wolfgang Spörk for their advice and support concerning matters beside this thesis. I am grateful for the wonderful time spent together both on and off campus.

Finally, I wish to thank my mother Irene and above all my girlfriend Genesis for her steady encouragement and support that made the completion of this work possible and who continuously reminded me of the really important things in life.

Cologne, January 2016 Eugen Töws

This thesis consists of the following works:

Thomas Hartmann-Wendels, Patrick Miller, and Eugen Töws, 2014. Loss given default for leasing: Parametric and nonparametric estimations. Journal of Banking & Finance 40, 364–375.

Eugen Töws, 2014. The impact of debtor recovery on loss given default. Working paper.

Patrick Miller and Eugen Töws, 2015. Loss given default-adjusted workout processes for leases. Working paper.

Contents

List of Abbreviations

AIC	Akaike information criterion
ALGD	Asset-related loss given default
AP	Asset proceeds
ARR	Asset-related recovery rate
BIC	Bayesian information criterion
CF	Cash flows
CRR	Capital requirement regulation
EAD	Exposure at default
FMM	Finite mixture model
ICT	Information and communications technology
IRBA	Internal ratings based approach
Is	In-sample
kNN	k-nearest neighbor
LC	Liquidation costs
LGD	Loss given default
MAE	Mean absolute error
MLGD	Miscellaneous loss given default
MRR	Miscellaneous recovery rate
MSE	Mean squared error
MURD	Moody's Ultimate Recovery Database
NN	Neural network

OLS	Ordinary least squares
Oos	Out-of-sample
Oot	Out-of-time
PD	Probability of default
RF	Random forest
RMSE	Root mean squared error
RR	Recovery rate
RT	Regression Tree
SVM	Support vector machine
TARGET	Tree analysis with randomly generated and evolved trees
TIC	Theil inequality coefficient
WC	Workout costs

List of Figures

List of Tables

1 Introduction

Credit risk is a major concern of financial institutions and their risk management. To the institutes, it is essential to identify and measure this risk in order to make economically reasonable credit decisions and to calculate regulatory capital. For the determination of credit risk of financial assets, banking regulation provides three essential components. These are the probability of default (PD), the loss given default (LGD), and the exposure at default (EAD). Moreover, two different approaches are provided to incorporate these components. Article 107 (1) of the capital requirement regulation (CRR) states that financial institutions shall apply either the Standardised Approach or the Internal Ratings Based Approach (IRBA) to calculate their regulatory capital requirements for credit risk. Depending on the chosen approach, these components can or must be determined for regulatory and economic purposes. With the IRBA, the Basel Committee on Banking Supervision (2003) intends to increase the sensitivity of risk factors to the risk of the assets of the applying institutions. A risk-adequate approach, such as IRBA, should ideally reduce the regulatory capital of these institutes. Consequently, capital could be released that is tied-up in backing financial assets.

So far, researchers studied the PD extensively and established sophisticated measurements, such as the value-at-risk. Implementations are available abundantly, e. g., CreditPortfolioView, CreditMetrics™, and CreditRisk$^+$. Frontczak and Rostek (2015) assume that the EAD is predictable to a large extent by means of amortization schedules. Hence, only for LGD robust estimation procedures are scarce.

In the following, we will consider LGD in more detail. LGD is that share of the outstanding claim of a defaulted contract that could not have been recovered. Several studies refer to its counterpart the recovery rate, which is $1 - \text{LGD}$.

Accurate estimates of potential losses are essential to allocate economic and regulatory capital and to price credit risk of financial instruments. Moreover, Gürtler and Hibbeln (2013) argue that accurately estimating the LGD should result in competitive advantages to the applying institution. Against this theoretical and practical background, this thesis contributes to the growing research area of LGD estimation. Particularly, it introduces new approaches and puts these into the context of the existing literature. Furthermore, several findings in related research can be confirmed empirically or put into perspective.

Recent studies on the estimation of LGD are mainly based on defaulted loans and bonds, such as Yao et al. (2015), Leow et al. (2014), Jankowitsch et al. (2014), Khieu et al. (2011), and Calabrese and Zenga (2010). Only little evidence exists on the LGD of leases apart from Hartmann-Wendels and Honal (2010), De Laurentis and Riani (2005), and Schmit and Stuyck (2002). However, there is at least one major peculiarity of leases when comparing their recovery risk to that of loans, which may reduce the LGD significantly. Eisfeldt and Rampini (2009) argue that for the legal owner of the leased asset, i. e. the lessor, its reposition is easier than foreclosure on the collateral for a secured loan. Moreover, the lessor may retain any value from disposing of the asset, even if the recoveries exceed the outstanding claim. Thus, the asset of a lease contract is a native collateral, which lessors are experts in disposing off. In fact, examining defaulted lease contracts from major European financial institutions, Schmit and Stuyck (2002) find that defaulted leases on vehicles and real estate on average exhibit lower LGDs than loans and bonds. Theoretically, this finding is plausible. However, there might be exceptions to this rule.

Calculation of LGD is a rather technical issue. There are two acknowledged methods to determine the LGD of financial instruments. One of which is the concept of market LGDs. The market LGD is calculated as one minus the ratio of the trading price of the asset some time after default to the trading price at the time of default. However, market LGDs are only available for bonds and loans issued by large firms. Moreover, Khieu et al. (2011) find evidence, that market LGDs are biased and inefficient estimates of the realized LGD. The second concept is the workout LGD. Workout LGDs are calculated as one minus the ratio of the discounted cash flows after default to the EAD.

The distribution of workout LGDs is often reported to be bimodal, e. g. by Li et al. (2014), Qi and Zhao (2011), and Hartmann-Wendels and Honal (2010). The distribution of a parameter is considered bimodal if it exhibits two local maxima. Particularly for the LGD of leases these maxima are located around zero and one. This shape is rather unusual because there is no single probability function coming close to bimodal distributions. The unusual shape raises the question whether standard econometric methods, such as ordinary least squares (OLS) linear regression, are appropriate for the estimation task. This thesis presents empirical evidence that complex approaches, such as regression trees and multi-step models, have a significant advantage over standard methods, given a sufficiently large data and information base. Furthermore, we find that economic consideration can be a key driver of estimation improvements.

All approaches developed in this work incorporate workout LGDs and several additional requirements of the CRR and its predecessor Basel II. In particular, we consider workout costs and the update of LGD estimates in case of default, which prior literature mostly neglects.

This thesis consists of three essays on the estimation of the risk parameter LGD. The first essay (Hartmann-Wendels, Miller, and Töws, 2014, Loss given default for leasing: Parametric and nonparametric estimations) fills a gap in LGD related literature by focusing on elementary differences of the examined estimation approaches. Three major German leasing companies provided a total of 14,322 defaulted leasing contracts. Based on this data, we compare parametric, semiparametric and nonparametric estimation methods. We use the historical average and the parametric OLS regression as benchmark and compare the semiparametric finite mixture model (FMM) to the nonparametric model tree M5'. We evaluate the performance of the used methods in an in-sample and out-of-sample validation.

The most elementary estimation method is a look-up table, which either bases on historical averages or expert opinions (see Gupton and Stein (2005)). Directly following is OLS, which is also easy to implement. Therefore, OLS dominates the used methods for estimating the LGD in recent literature. Most empirical studies find the LGD to have a bimodal or even multimodal shape. Given this finding, OLS may lead to inefficient estimates by estimating the conditional expectation of the LGD. Whereas it is econometrically reasonable to approximate the LGD distribution by a mixture of a finite number of standard distributions, e. g. normal distributions. Implementing FMM, we develop a multi-step model to cluster the data into distinct clusters first. The data then is classified to the found clusters employing different classification algorithms. Finally, we calibrate OLS models to the contracts of each cluster. Thereby, we allow for different influencing factors within these clusters. The last category of studied approaches is tree algorithms. These methods produce decision trees using if-then conditions to divide the data subsequently in order to reduce its inhomogeneity. The determined final subsamples then are averaged in terms of their LGD. Alternatively, regression models are built within these subsamples.

Our results show that a model's in-sample performance is a poor indicator of its out-of-sample estimation capability. We find that FMM is quite capable of reproducing the unusual shape of the LGD distribution in-sample as well as out-of-sample. Moreover, when measuring the models performance in terms of the deviation of estimated from realized LGDs, FMM produces very low in-sample errors. However, out-of-sample the error increases significantly and exceeds that of OLS in most cases. While OLS is mostly outperformed in-sample, the model tree produces robust in-sample and out-of-sample estimations exhibiting the lowest level of out-of-sample estimation errors. Furthermore, we find that the improvement of the model tree increases with an increasing dataset. Also, all models' performance level is highly dependent on the peculiarities of the underlying data. In order to account for a company's idiosyncratic characteristics, it is reasonable to consider the datasets separately.

The second essay (Töws, 2014, The impact of debtor recovery on loss given default) addresses the economic consideration of the workout process of defaulted contracts. It founds on the lessor's retainment of legal title to the leased asset and, consequently, his easy access to it in case of default. Dependent on the lessor's workout strategy, defaulted contracts may develop in two distinct ways. Either the default reason can be dissolved and the debtor recovers or the contract must be written off. This work uses the essential information of the contract's default end to study its influence on the LGD. We observe the recovery or write-off of 42,575 defaulted leasing contracts of three German leasing companies. In the data, we find that recovered contracts exhibit significantly lower LGD levels than contracts that were written off. According to the significant influence, we provide evidence that employing the default end in an LGD estimation approach is highly beneficial to the estimation accuracy.

Reminding ourselves that regression trees performed well in estimating LGD in the first essay, we compare the forecasting performance of three tree algorithms. These are J4.8, random forest (RF), and C5.0. Developing a two-step approach for estimating LGD, we first divide the data according to the contracts' default end in a classification. On each of the two classes, a regression model is calibrated in the second step, and every contract is assigned exactly two LGD estimations from both of the regression models. The final LGD estimation then is the linear combination of the estimated LGDs weighted with their respective classification probability.

Compared to direct estimation with OLS, we find our approach to improve the estimation accuracy of LGD. When we consider the coefficient of determination, the improvement is significant for each of the three datasets. The study indicates the benefits of establishing the lessor's expertise in assessing a defaulted contract's continuation worthiness. If successfully implemented, the resulting workout process should produce lower LGDs than before and thereby strengthen the lenders competitiveness.

The third essay (Miller and Töws, 2015, Loss given default-adjusted workout processes for leases) contributes to the LGD estimation literature considering unique features of leasing contracts. Based on a dataset of 1,493 defaulted leasing contracts, we economically account for leasing peculiarities and develop a particularly suited approach to estimate the contracts' LGD. To the best of our knowledge, we are the first to separate the LGD into two distinct parts. We ground this separation on the economic consideration that cash flows of the workout process of defaulted leasing contracts, in general, are coming from two distinct sources. The first part includes all asset-related cash flows, such as the asset's liquidation value and incurred liquidation costs. The second part comprises the remaining cash flows, such as overdue payments and collection costs.

In the course of the study, we find ALGD to be a theoretical upper bound to the LGD, given that the MLGD does not exceed a value of one. Assuming this is the case, an ALGD less than one directly indicates the overall LGD being below a value of one. As soon as ALGD reaches a value of zero, the disposal revenues cover EAD in full. In any case, if MLGD exceeds a value of one, the lessor should restrict the workout process to the asset's disposal. Under such circumstances, the collection of overdue payments is economically inefficient and causes monetary losses to the lessor.

Constructing a multi-step LGD estimation approach, we essentially compare the performance of two different methods: OLS regression and RF. The first step estimates the respective shares of the LGD. These are the asset-related LGD (ALGD) and the miscellaneous LGD (MLGD). Including these factors, we perform a classification of the data in the second step. We estimate whether a contract's ALGD exceeds its LGD. Similar to the approach of the second essay, we calibrate regression models for each class in step three. The final LGD estimation is the weighted linear combination of the estimated LGDs and the contract's classification probability.

Including the estimated ALGD and MLGD into the estimation approach increases the estimation accuracy. Most importantly, the relative performance improvement is independent of the method applied. It rather arises from the economic approach, which targets the specifics of leasing contracts. We find that the estimated values of ALGD and MLGD are sturdy indicators for the success or failure of the workout process. Thus, the lessor can benefit from the consideration of both these forecasts for his actions concerning the workout process.

2 Loss given default for leasing: Parametric and nonparametric estimations

The loss given default (LGD) and its counterpart, the recovery rate, which equals one minus the LGD, are key variables in determining the credit risk of a financial asset. Despite their importance, only a few studies focus on the theoretical and empirical issues related to the estimation of recovery rates.

Accurate estimates of potential losses are essential to efficiently allocate regulatory and economic capital and to price the credit risk of financial instruments. Proper management of recovery risk is even more important for lessors than for banks because leases have a comparative advantage over bank loans with respect to the lessor's ability to benefit from higher recovery rates in the event of default. In their empirical cross-country analysis, Schmit and Stuyck (2002) note that the average recovery rate for defaulted automotive and real estate leasing contracts is slightly higher than the recovery rates for senior secured loans in most countries and much higher than the recovery rates for bonds. Moreover, the recovery time for defaulted lease contracts is shorter than that for bank loans. Because the lessor retains legal title to the leased asset, repossession of a leased asset is easier than foreclosure on the collateral for a secured loan. Moreover, the lessor can retain any recovered value in excess of the exposure at default. Repossessing used assets and maximizing their return through disposal in secondary markets are aspects of normal leasing business and are not restricted to defaulted contracts. Therefore, lessors have a good understanding of the secondary markets and of the assets themselves. Because the lessor's claims are effectively protected by legal ownership, the high recoverability of the leased asset may compensate for the poor creditworthiness of a lessee. Lasfer and Levis (1998) find empirical evidence for the hypothesis that lower-rated and cash-constrained firms have a greater propensity to become lessees. To leverage their potential lower credit risk, lessors must be able to accurately estimate the recovery rates of defaulted contracts.

This paper compares the in-sample and out-of-sample accuracies of parametric and nonparametric methods for estimating the LGD of defaulted leasing contracts. Employing a large dataset of 14,322 defaulted leasing contracts from three major German lessors, we find in-sample accuracy to be a poor predictor of out-of-sample accuracy. Methods such as the hybrid finite mixture models (FMMs), which attempt to reproduce the LGD distribution, perform well for in-sample estimation but yield poor results out-of-sample. Nonparametric models, by contrast, are robust in the sense that they deliver fairly accurate estimations in-sample, and they perform best out-of-sample. This result is important because out-of-sample estimation has rarely been performed in other studies – with the notable exceptions of Han and Jang (2013) and Qi and Zhao (2011) – although out-of-sample accuracy is critical for proper risk management and is required for regulatory purposes.

Analyzing estimation accuracy separately for each lessor, our results suggest that the number of observations within a dataset has an impact on the relative performance of the estimation methods. Whereas sophisticated nonparametric estimation techniques yield, by far, the best results for large datasets, simple OLS regression performs fairly well for smaller datasets.

Finally, we find that estimation accuracy critically depends on the available set of information. We estimate the LGD at two different points in time, at the execution of the contract and at the point of contractual default. This procedure is of particular importance for leasing contracts because the loan-to-asset value changes during the course of a leasing contract. Furthermore, the Basel II accord requires financial institutions using the advanced internal ratings based approach (IRBA) to update their LGD estimates for defaulted exposure. To the best of our knowledge, an analysis of this type of update has been neglected in the literature thus far.

2.1 Literature review

There are two major challenges in estimating recovery rates for leases with respect to defaulted bank loans or bonds. First, estimates of LGD on loans or bonds take for granted that the recovery rate is bounded within the interval $[0, 1]$, which assumes that the bank cannot recover more than the outstanding amount (even under the most favorable circumstances) and that the lender cannot lose more than the outstanding amount (even under the least favorable circumstances). Although the assumption of an upper boundary is justified for bank loans, it does not apply to leasing contracts. As the legal owner of the leased asset, the lessor may retain any value recovered by redeploying the leased asset, even if the recoveries exceed the outstanding claim. In fact, there is some empirical evidence that recovery rates greater than 100% are by no means rare. For example, Schmit and Stuyck (2002) report that up to 59% of all defaulted contracts in their sample have a recovery rate that exceeds 100%. Using a different dataset, Laurent and Schmit (2005) find that recovery rates are greater than 100% in 45% of all defaulted contracts. The lower boundary of the recovery rate rests on the implicit assumption of a costless workout procedure. In fact, most empirical studies neglect workout costs (presumably) because of data limitations. Only Grippa et al. (2005) account for workout costs in their study of Italian bank loans and find that workout costs average 2.3% of total operating expenses. The Basel II accord, however, requires that workout costs are included in the LGD calculation. Thus, when workout costs are incorporated, there is no reason to assume that workout recovery rates must be non-negative. The second challenge in estimating recovery rates is the bimodal nature of the density function, with high densities near 0 and 1. This property of workout recovery rates is well documented in almost all empirical studies, whether of bank loans or leasing contracts (e. g., Laurent and Schmit (2005)).

Because of the specific nature of the recovery rate density function, standard econometric techniques, such as OLS regression, do not yield unbiased estimates. Renault and Scaillet (2004) apply a beta kernel estimator technique to estimate the recovery rate density of defaulted bonds, but they find that it is difficult to model its bimodality. Calabrese and Zenga (2010) extend this approach by considering the recovery rate as a mixed random variable obtained as a mixture of a Bernoulli random variable and a continuous random variable on the unit interval and then apply this new approach to a large dataset of defaulted Italian loans. Qi and Zhao (2011) compare fractional response regression to other parametric and nonparametric modeling methods. They conclude that nonparametric methods – such as regression trees (RTs) and neural networks – perform better than parametric methods when overfitting is properly controlled for. A similar result is obtained by Bastos (2010), who compares the estimation accuracy of fractional response to RTs and neural networks.

Despite the growing interest in the modeling of recovery rates, little empirical evidence is available on this topic. Several studies (e. g., Altman and Ramayanam (2007), Friedman and Sandow (2005), and Frye (2005)) rely on the concept of market recoveries, which are calculated as the ratio of the price for which a defaulted asset is traded some time after default to the price of that asset at the time of default. Market recoveries are only available for bonds and loans issued by large firms. Workout recoveries are used by Khieu et al. (2011), Dermine and Neto de Carvalho (2005), and Friedman and Sandow (2005). However, Khieu et al. (2011) find evidence that the post-default price of a loan is not a rational estimate of actual recovery realization, i. e., it is biased and/or inefficient. According to Frye (2005), many analysts prefer the discounted value of all cash flows as a more reliable measurement of defaulted assets because: (1) cash flows ultimately become known with certainty, whereas the market price is derived from an uncertain forecast of future cash flows; (2) the market for defaulted assets might be illiquid; (3) the market price might be depressed; and (4) the asset holder might not account for the asset on a market-value basis.

Schmit et al. (2003) analyze a dataset consisting of 40,000 leasing contracts, of which 140 are defaulted. Using bootstrap techniques, they conclude that the credit risk of a leasing portfolio is rather low because of its high recovery rates. Similar studies are conducted by Laurent and Schmit (2005) and Schmit (2004). Schmit and Stuyck (2002) find considerable variation in the recovery rates of 37,000 defaulted leasing contracts of 12 leasing companies in six countries. Average recovery rates depend on the type of the leased asset, country, and contract age. De Laurentis and Riani (2005) find empirical evidence that leasing recovery rates are inversely correlated with the level of exposure at default. However, recovery rates increase with the original asset value, contract age, and existence of additional bank guarantees. Applying OLS regressions to forecast LGDs in that study leads to rather poor results: the unit interval is divided into three equal intervals, and only 31–67% of all contracts are correctly assigned in-sample. With a finer partition of five intervals, the portion of correctly assigned contracts decreases even further. These results clearly indicate that more appropriate estimation techniques are needed to accurately estimate recovery rates.

Our study differs from the LGD literature in several crucial aspects. First, we calculate workout LGDs and consider workout costs. Second, we perform out-of-sample testing at contract execution and default, which meets the Basel II requirements for LGD validation. Third, by separately analyzing the datasets of three lessors, we gain insight into the robustness of the estimation techniques.

Company	# Contracts	# Lessees
A	9,735	5,811
B	2,995	2,344
C	1,592	964

Table 2.1: Numbers of contracts and lessees in the datasets of companies A–C in descending order of the number of contracts.

2.2 Dataset

This study uses datasets provided by three German leasing companies, which shall be referred to herein as companies A, B, and C. All three companies use a default definition consistent with the Basel II framework. According to Table 2.1, the dataset from lessor A contains 9,735 leasing contracts with 5,811 different customers and default dates between 2002 and 2010. The dataset from lessor B contains 2,995 leasing contracts with 2,344 different lessees who defaulted between 1994 and 2009, with the majority of defaults occurring between 2001 and 2008. The dataset for leasing company C consists of 1,592 leasing contracts with 864 different lessees who defaulted between 2002 and 2009.

For the defaulted contracts, we calculate the LGD as one minus the recovery rate. The recovery rate is the ratio of the present value of cash inflows after default to the exposure at default (EAD). For leasing contracts, the cash flows consist of the revenues obtained by redeploying the leased asset and other collateral combined with other returns and less workout expenses. The cash flows are discounted to the time of default using the term related refinancing interest rate.[1] The EAD is the sum of the present value of the outstanding minimum lease payments, compounded default lease payments, and the present residual value. All values refer to the time of default. A contract is classified as defaulted when at least one of the triggering events set out in the Basel II framework has occurred.

[1] Only a few studies (such as Gibilaro and Mattarocci (2007)) address risk-adjusted discounting. We use the term related refinancing interest rate to discount cash flows at the time of default, independently of the time span of the workout and the risk of each type of cash flow.

Before the data was collected, all three companies agreed to use identical definitions for all the elements that are entered into the LGD calculation, and for all details of the leasing contract, lessee, and leased asset. Thus, for every contract, we have detailed information about the type and date of payments that the lessor received after the default event. Moreover, we incorporate expenses arising during the workout into the LGD calculation, to meet Basel II requirements. Workout costs are rarely considered in empirical studies.

The workouts have been completed for all the observed contracts. Gürtler and Hibbeln (2013) recommend restricting the observation period of recovery cash flows to avoid the under-representation of long workout processes, which might result in an underestimation of LGDs. Because we do not see a similar problem in our data, we do not truncate our observations based on that effect.

All three companies also provide a great deal of information about factors that might influence the LGD, which we divide into four categories:

1. contract information;
2. customer information;
3. object information; and
4. additional information at default.

Contract information is elementary information about the contract, such as its type, e. g., whether it was a full payment lease, partial amortization, or hire-purchase; its duration; its calculated residual value or prepayment rents; and information about collateralization and/or purchase options. Customer information mainly identifies retail and non-retail customers. The category object information consists of basic information about the object of the lease, including its type, initial value, and supplementary information, such as the asset depreciation range. Whereas all the information in the first three groups is available from the moment the contract is concluded, the last category consists of information that only be-

Company	Mean	Std	P5	P25	Median	P75	P95
A	0.52	0.40	−0.11	0.19	0.52	0.88	1.05
B	0.35	0.42	−0.18	0.00	0.25	0.72	1.01
C	0.39	0.42	−0.23	0.03	0.32	0.77	1.03

Table 2.2: Loss given default (LGD) density information for companies A–C. Std is the standard deviation and P5–P95 are the respective percentiles.

comes available after the contract has defaulted, such as the exposure at default and the contract age at default.

Descriptive statistics

The LGD is clearly not restricted to the interval $[0, 1]$. As presented in Table 2.2 and Figure 2.1, negative LGDs are not only theoretically possible but also occur frequently in the leasing business. Hartmann-Wendels and Honal (2010) argue that such cases mainly occur if a defaulted contract with a rather low EAD yields a high recovery from the sale of the asset. Because we incorporate the workout expenses, LGDs greater than one are also feasible. Thus, we do not bound LGDs within the $[0, 1]$ interval, as is common for bank loans and as is done by Bastos (2010), by Calabrese and Zenga (2010), and by Loterman et al. (2012).

An LGD of 45%, as specified in the standard credit risk approach, is considerably higher than the median LGDs observed for companies B and C. In general, we emphasize that the shape of the LGD distribution varies significantly among these three companies. As presented in Figure 2.1, only the LGD distribution of company C exhibits the frequently mentioned bimodal shape, whereas those of companies A and B feature three maxima. These differences continue to prevail when we account for differences in the leasing portfolio. Thus, we trace these variations back to differences in workout policies. Because the requirements for the pooling of LGD data, set out in section 456 of the Basel II accord, are clearly vio-

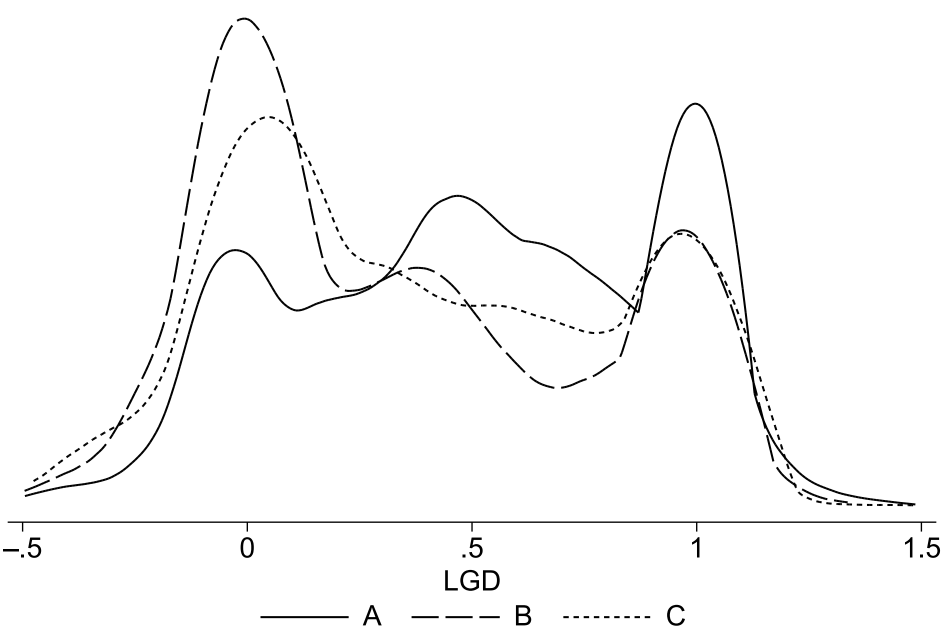


Figure 2.1: Density of the realized loss given default (LGD) by company. The realized LGD concentrates on the interval $[-0.5, 1.5]$. The figures describe a loss severity of -50% on the left end, which indicates that 150% of the exposure at default (EAD) was recovered. On the right end, the loss severity is 150%, indicating a loss of 150% of the EAD. Consequently, a realized LGD of 0 or 1 indicates the following: in case of 0, full coverage of the EAD (included workout costs); or, in case of 1, total loss of the EAD.

lated, we construct individual estimation models to account for institution-specific characteristics and differences in LGD profiles among the companies.

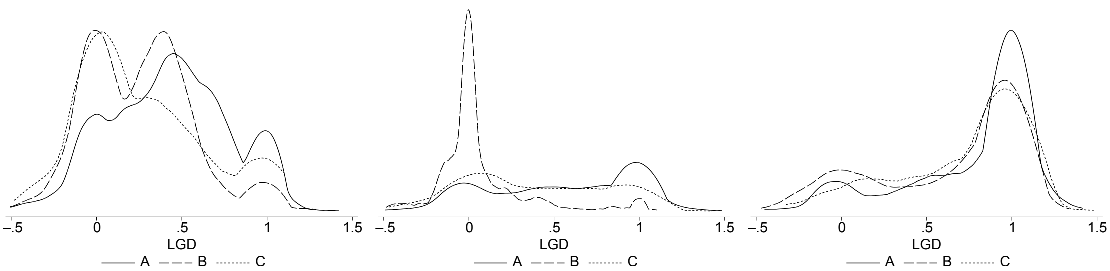
Previous studies on the LGD of defaulted leasing contracts consistently show that the LGD distribution depends largely on the underlying asset type. We categorize the contracts according to the underlying asset using five classes: vehicles, machinery, information and communications technology (ICT), equipment, and other. Table 2.3 summarizes the key statistical figures of the distributions for each company. We can unambiguously rank the three companies with respect to their mean LGD. Company B achieves the lowest average LGD for all asset types, company C is second best, and company A bears the highest losses. Contracts in ICT have the highest average LGD. Examining the median of ICT, we find that companies A, B, and C retrieve only 4%, 16%, and 13% of the EAD, respectively, in half of the cases. The key statistical figures for equipment and other assets are

Asset type	Company	# Contracts	Mean	Std	Median
	A	4,578	0.44	0.35	0.45
Vehicles	B	1,111	0.26	0.31	0.27
	C	599	0.28	0.37	0.21
	A	4,140	0.55	0.43	0.61
Machinery	B	779	0.06	0.27	0.00
	C	646	0.39	0.42	0.32
	A	606	0.77	0.38	0.96
ICT	B	1,062	0.64	0.43	0.84
	C	201	0.72	0.38	0.87
	A	353	0.61	0.44	0.74
Equipment	B	26	0.26	0.44	0.09
	C	26	0.38	0.41	0.15
	A	58	0.56	0.43	0.54
Other	B	17	0.39	0.44	0.26
	C	120	0.46	0.43	0.45

Table 2.3: Loss given default (LGD) density information by asset type for companies A–C. For each asset type, # Contracts is the number of contracts containing this type of asset, Mean is its mean, Std is its standard deviation, and Median is its median. ICT is information and communications technology. The displayed asset types vary in the numbers of their contracts and even further in the characteristics of their realized LGD.

seemingly less meaningful because of the small sample sizes for these classes, but the trends are consistent across all three companies.

Figure 2.2 presents the LGD distributions for vehicles, machinery, and ICT for each company. The shape of the LGD distributions differs tremendously with respect to the different asset types. Whereas for ICT, the LGD density in Figure 2.2c is right-skewed toward high LGDs with only weak bimodality throughout all of the companies, the density of machinery runs partly the opposite direction. For machinery, in Figure 2.2b, we see a higher concentration around 0, but for company A, larger LGDs again outweigh this effect. The LGD for contracts with vehicles varies greatly from company to company. We observe a strong multimodality for all of the companies with an additional peak at approximately 0.5, and most of the density lies in the lower LGD range.



(a) Vehicles. (b) Machinery. (c) ICT.

Figure 2.2: Densities of realized loss given default (LGD) by company for the three major asset types: vehicles, machinery, and information and communications technology (ICT). Depending on the asset type, the realized LGD density appears in completely different shapes. For machinery (Figure b), even the difference between companies is enormous.

2.3 Methods

This section describes the various approaches that we use to estimate the LGD and its density. According to section 448 of the Basel II regulations, institutes are required to base their estimations on a history of defaults and to consider all relevant data, information, and methods. Furthermore, a bank using the advanced IRBA must be able to break down its experience with respect to the probability of default (PD), LGD, and the IRBA conversion factor. This breakdown is to be based on the factors that are identified as drivers of the respective risk parameters.

The basic method used to identify these drivers is to partition the data according to a certain attribute (e. g., the type of object). Differences in the means of the partitions are then captured by setting the inducing factor as the driver. The average value is then the (naive) estimator of the LGD for the corresponding subclass. As Gupton and Stein (2005) note, this traditional look-up table approach is static and backward-looking, even if considerable variation is observed in the LGD distributions for different types of objects. An alternative method of verifying the impact of potential factors and developing an estimation model is to conduct a regression analysis. Linear regressions always estimate the (conditional) expectation of the target variable, but this average is not a reasonable parameter under mixed distributions, so it is not an adequate approach from a statistical perspective. However, regression analyses for LGD estimation are successfully implemented by Bellotti and Crook (2012) and by Zhang and Thomas (2012).

Table 2.2 reports the median LGDs as 52%, 25%, and 32% for companies A, B, and C, respectively. Considering the LGD distribution in Figure 2.1, its heterogeneity suggests that the overall portfolio is composed of several subclasses, which are less heterogeneous in terms of the LGD. This implies that each subclass has its own characteristic LGD distribution. We use FMMs to reveal these unknown classes (cluster analysis), to fit a reasonable model to the data and to classify the observations into these classes. Furthermore, we apply two different regression/model tree algorithms to the data. These tree-based models also have the basic function of dividing the portfolio into homogeneous partitions; by contrast to the FMMs, however, the number of subclasses is endogenously determined rather than exogenously specified.

At the end of this section, we present an overview of how to select the explanatory variables for tree-based methods. We also describe our methodology for out-of-sample testing.

2.3.1 Finite mixture models and classification

Modeling the probability density of realized LGDs as a mixed distribution allows us to use different potential LGD drivers for different clusters and to capture differences in the effects of these drivers on the LGD in various subclasses. We adapt an approach originally proposed by Elbracht (2011). FMMs are described by Frühwirth-Schnatter (2006).

The approach consists of three steps: (1) cluster the total dataset into finite classes by finite mixture distributions using all available information; (2) classify the dataset into the resulting classes using only the information available at the execution or default of the contract by the k-nearest neighbors (kNN) or the classification tree algorithm J4.8; and (3) perform OLS regressions for each class.

Step (1) can be adjusted between the two extremes of nonparametric and parametric modeling, thus providing a flexible method of data adaptation. We use normal distributions to construct the mixing distributions. We estimate unknown model parameters using the expectation maximization algorithm, which also provides a probabilistic classification of the observations. The accuracy of classification step (2) can be measured for in-sample testing. However, in out-of-sample testing, the goal is to classify observations that do not belong to any class initially – because these objects are not part of the training sample used to form classes – into exactly one of the given classes.

We compare two different approaches to classifying contracts into previously established classes. The nonparametric kNN approach assigns an observation to the class with the majority of its k nearest neighbors, whereas the distance between observations is determined as the Euclidean distance. This approach is described by Hastie et al. (2009). We also apply the tree algorithm J4.8 for classification.

The J4.8 algorithm generates pruned C4.5 revision 8 decision trees, as illustrated by Witten et al. (2011) and originally implemented by Quinlan (1993). The decision tree is constructed by dividing the sample according to certain threshold values. The optimal split in terms of maximized gain ratio is performed until additional splits yield no further improvement, or a minimum of instances per subset is reached. Every partition results in a node. To prevent overfitting, we prune back the fully developed tree to a certain level. According to Quinlan (1993), these deleted nodes shall not contribute to the classification accuracy of unseen cases.

2.3.2 Regression and model trees

RTs are classified as nonparametric and nonlinear methods. Similar to other regression methods, they can be applied to analyze the underlying dataset and to predict the (numeric) dependent variable. An essential difference between RTs and parametric methods, such as linear or logistic regressions, is that ex-ante no assumption is made concerning the distribution of the underlying data, and no functional relationship is specified.

These characteristics are particularly beneficial in case of LGD estimation because it is typically not possible to describe the distribution of the LGD suitably with a single distribution, such as the normal distribution. In addition, the distribution of the LGD varies significantly according to the underlying data. Thus, as described in Section 2.2, the LGD distributions of the three companies studied here are all multimodal, although there are appreciable differences between companies, such as the number of maxima. In particular, more types of distributions are observed for bank loans (for an overview, see Dermine and Neto de Carvalho (2005)).

The basic idea of regression and model trees is to partition the entire dataset into homogeneous subsets by a sequence of splits, which creates a tree consisting of logical if-then conditions. Starting with the root node of the tree that contains all instances of the underlying data, each leaf covers only a fraction of the data.

In an RT, the prediction of the dependent variable is given by a constant for all instances belonging to a leaf, typically defined as the average value of these instances. Model trees are an extension of RTs in the sense that the target variable of instances belonging to a leaf is estimated by a linear regression model. Therefore, model trees are hybrid estimation methods combining RTs and linear regression. Model trees are clearly applicable for LGD estimation because RTs are successfully used in previous studies such as Bastos (2010) and Qi and Zhao (2011). Linear regression models are also applied to analyze and predict LGDs, and these models may deliver comparable or better results than those of more complex models, as shown by Bellotti and Crook (2012) and Zhang and Thomas (2012).

For our LGD estimation, we apply the M5' model tree algorithm and the corresponding RT algorithm that is introduced by Wang and Witten (1997) and described by Witten et al. (2011). This algorithm is a reconstruction of Quinlan's M5 algorithm that was published in 1992. In the case of the M5' algorithm, the underlying dataset is divided step by step, each time using the binary split based on the explanatory variables with the greatest expected reduction in the standard deviation. The constructed tree is subsequently pruned back to obtain an appropriately sized tree to control overfitting, which can influence out-of-sample performance negatively.

The resulting tree essentially depends on the explanatory variables used, particularly with respect to the M5' model tree algorithm; selecting appropriate variables is a complex issue because of ex-ante relevance and effectiveness not always being known. In the first step, we consider the potential application of a large number of parameters. However, it might be preferable to include only a fraction of the available variables, which we account for in the second step.

There are various algorithmic approaches for variable selection; two frequently applicable greedy algorithms are forward selection and backward elimination. Bellotti and Crook (2012) use forward selection for their LGD estimations of retail credit cards with OLS regression. However, forward selection has a significant disadvantage neglecting variable interactions.

Instead of forward selection, we employ backward elimination, initiating all available variables and step by step eliminating the variables without which the best value in terms of the respective fit criterion is achieved. This procedure continues until a stop condition is reached, or all the variables are eliminated.

A typical fit criterion for regression models is the F-score. However, the Akaike information criterion (AIC) and Bayesian information criterion (BIC) are used for forecasting, both of which are based on the log-likelihood function.

Analogous to the approximation of the AIC used by Bellotti and Crook (2012), the BIC can be approximated by

$$\text{BIC} = n \cdot ln(\text{MSE}) + p \cdot ln(n), \tag{2.1}$$

where n denotes the number of observations, p is the number of input variables, and MSE is the mean squared error of the observations.

We use the BIC, which penalizes the complexity of the model more than the AIC. This complexity is measured by the number of input variables. In addition to the number of explanatory variables, regression and model trees offer another complexity feature: the number of leaves in the computed tree. This aspect is among those included by Gray and Fan (2008) when designing the TARGET RT algorithm. The more leaves that are present in the computed tree, the greater the risk is that a contract will be misclassified, which negatively influences the estimation.

We find that the number of leaves is determined not only by the pruning procedure but also by the input variables. Thus, we modify the BIC and penalize the size of the computed tree

$$\text{BIC}^* = n \cdot ln(\text{MSE}) + p \cdot ln(n) + |T| \cdot ln(n), \tag{2.2}$$

where $|T|$ denotes the number of leaves of the computed tree.

For BIC^*, lower values are preferred. As with our data, $\text{MSE} \in (0, 1)$, $n \gg p$ and $n \gg |T|$ holds; thus, we have $\text{BIC}^* < 0$. We set the stop condition for our backward elimination such that a variable in the i-th iteration can only be eliminated if the BIC* value increased by an absolute value of at least one, which implies that the following constraint must be fulfilled

$$\mathrm{BIC}^*_{i-1} - \mathrm{BIC}^*_i \geq 1. \tag{2.3}$$

2.3.3 Out-of-sample testing

We calibrate our models on randomly divided training sets of 75% and validate their performance on the remaining 25% of the total dataset. Division and calibration are repeated 25 times. The final results are averaged. Our out-of-sample validation combines the advantages of k-fold cross-validation and the approach of splitting the dataset into training and test sets, and is particularly suitable for large datasets.

Bastos (2010) and Qi and Zhao (2011) employ k-fold cross-validation – using $k = 10$ – to evaluate the out-of-sample performance of their models. This method relies on partitioning the dataset randomly into k equal-sized subsets. While the model is calibrated on $k-1$ subsets, the models predictive performance is validated on the remaining subset. This procedure is performed k times, with each of the k subsets used exactly once for validation. Therefore each observation contained in the total dataset is used exactly once for validation. By contrast, we draw the 25 divisions in training and test data randomly. With a small k in the k-fold cross-validation there are fewer performance estimates, but the size of the subsets, and therefore the amount of the total dataset which is used for each validation, is larger. As k increases, the number of performance estimates increases, however, the size of the validation subset decreases rapidly. Given larger datasets, the data can be split into some training and test sets. Here, the validation is restricted to the unseen cases of the test set. Gürtler and Hibbeln (2013) randomly shuffle and divide their data as 70% training and 30% validation. Consequently, our out-of-sample validation combines the advantages of these two approaches. In particular we make use of large test sets and still generate multiple estimations.

2.4 Results

We present both in-sample and out-of-sample results in terms of LGD estimation – using different error measurements – and compare the results. These error parameters reflect the performance of our methods. Naturally, a low parameter outcome is preferable. We calculate the mean absolute error (MAE) and root mean squared error (RMSE) for each applied method according to the following definitions

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\mathrm{LGD}_i - \mathrm{LGD}_i^*|, \tag{2.4}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\mathrm{LGD}_i - \mathrm{LGD}_i^*)^2}, \tag{2.5}$$

where LGD denotes the realized LGD, LGD^* is the predicted LGD, and n is the number of observations.

In addition to these measurements we calculate the Theil inequality coefficient (TIC), presented by Theil (1967)

$$\mathrm{TIC} = \frac{\frac{1}{n}\sum_{i=1}^{n}(\mathrm{LGD}_i - \mathrm{LGD}_i^*)^2}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\mathrm{LGD}_i^2} + \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\mathrm{LGD}_i^*)^2}}. \tag{2.6}$$

TIC sets the mean squared error relative to the sum of the average quadratic realized and estimated LGD and thereby accounts for both the model's goodness of fit and robustness. The factor is bound to $[0, 1]$ with $\mathrm{TIC} = 0$ being the perfect estimator. Theil finds that a useful forecast can be made up to $\mathrm{TIC} \approx 0.15$.

For a better interpretation of the results, we also show the results of the historical average and two simple OLS regression models as benchmarks. We use identical explanatory variables for OLS regression as for the M5' algorithm and RT before applying the variable selection procedure. Similar to M5' and RT, we further apply a backward elimination to the OLS regression according to the BIC criterion in Equation (2.1).

We estimate the LGD at two different points in time: once at the execution of the contract and once at the time of default. Typically, more information is available at default, which should theoretically yield better predictions.

The in-sample and out-of-sample results are evaluated by calculating the Janus quotient introduced by Gadd and Wold (1964)

$$\text{Janus} = \sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}(\text{LGD}_i - \text{LGD}^*_{i,\text{Oos}})^2}{\frac{1}{m}\sum_{i=1}^{m}(\text{LGD}_i - \text{LGD}^*_{i,\text{Is}})^2}}, \tag{2.7}$$

with the in-sample estimation LGD^*_{Is} in the denominator and the out-of-sample estimation $\text{LGD}^*_{\text{Oos}}$ in the numerator. Janus $= 1$ for equally large prediction errors for both estimations. A value close to 1 indicates a stable model and data structure.

At the end of the chapter, we also provide quality features of the identified finite mixture distributions and the error rates of classification for robustness reasons, and we interpret these results.

2.4.1 In-sample results

Beginning with the in-sample outcomes presented in Table 2.4, our models largely produce better estimations with the additional information available at default.

Our results clearly show the superiority of the FMMs for in-sample testing. The MAE, RMSE, and TIC of the FMM_{3NN} are mostly far from their counterparts

	Company A				Company B				Company C			
Method	Lv.	MAE	RMSE	TIC	Lv.	MAE	RMSE	TIC	Lv.	MAE	RMSE	TIC
Hist. avg.		0.3418	0.4018	0.1381		0.3646	0.4205	0.1988		0.3662	0.4195	0.1822
At execution												
OLS		0.3240	0.3868	0.1268		0.2706	0.3451	0.1235		0.3282	0.3889	0.1519
OLS_{BIC}		0.3246	0.3874	0.1272		0.2719	0.3471	0.1251		0.3307	0.3909	0.1537
FMM_{3NN}		0.2806	0.3713	0.1099		0.2209	0.3354	0.1044		0.2571	0.3601	0.1176
$\text{FMM}_{\text{J4.8}}$		0.2919	0.3916	0.1043		0.2589	0.3911	0.1192		0.3028	0.3914	0.1255
M5'	13	0.3142	0.3786	0.1209	1	0.2711	0.3459	0.1241	2	0.3272	0.3874	0.1504
$\text{M5'}_{\text{BIC}^*}$	17	0.3132	0.3774	0.1201	9	0.2640	0.3388	0.1185	9	0.3148	0.3751	0.1400
RT	34	0.3183	0.3817	0.1231	7	0.2726	0.3464	0.1248	9	0.3279	0.3871	0.1510
RT_{BIC^*}	26	0.3197	0.3829	0.1240	11	0.2687	0.3423	0.1217	7	0.3314	0.3898	0.1531
At default												
OLS		0.3114	0.3761	0.1191		0.2692	0.3435	0.1211		0.3238	0.3858	0.1490
OLS_{BIC}		0.3123	0.3768	0.1195		0.2709	0.3451	0.1234		0.3238	0.3858	0.1527
FMM_{3NN}		0.2550	0.3468	0.0955		0.2148	0.3280	0.1001		0.2432	0.3437	0.1091
$\text{FMM}_{\text{J4.8}}$		0.2588	0.3594	0.0900		0.2408	0.3693	0.1056		0.2835	0.3723	0.1190
M5'	6	0.3014	0.3680	0.1134	2	0.2650	0.3399	0.1193	1	0.3274	0.3883	0.1513
$\text{M5'}_{\text{BIC}^*}$	12	0.2997	0.3666	0.1126	12	0.2539	0.3277	0.1101	3	0.3244	0.3858	0.1490
RT	49	0.3032	0.3689	0.1143	25	0.2642	0.3373	0.1181	13	0.3247	0.3844	0.1487
RT_{BIC^*}	39	0.3046	0.3699	0.1150	10	0.2674	0.3422	0.1215	7	0.3294	0.3886	0.1523

Table 2.4: In-sample estimation errors at the execution and default of contracts by company. The best results are underlined for each company and type of error. Hist. avg. is the historical average loss given default (LGD) used as estimation of the LGD. OLS represents the ordinary least squares regression, and FMM is the finite mixture model in combination with 3-nearest neighbors (3NN), or J4.8. OLS is also performed with the variable selection BIC algorithm and the M5' algorithm and the RT are performed with the variable selection BIC^* algorithm. Lv. defines the number of leaves on the tree. MAE is the mean absolute error defined in Equation (2.4) and RMSE is the root mean squared error defined in Equation (2.5). TIC is the Theil inequality coefficient defined in Equation (2.6). For MAE, RMSE, and TIC, lower outcomes are preferable.

of the other models and even farther from the historical averages. The OLS regressions are outperformed in all cases.

Upon closer inspection, we note a large gap between the MAE and RMSE of the FMMs of approximately 10 percentage points, which is thus much larger than for the OLS regressions and tree-based models. We discuss this effect in more detail in Section 2.4.3. Our findings are consistent with Elbracht (2011) and with the discrepancy between MAE and RMSE noted by Loterman et al. (2012).

Proceeding with the tree-based models, we determine that by application of the variable selection procedures, $\text{M5'}_{\text{BIC}^*}$ strictly outperforms all RT models and the

OLS regressions, except for company C at default. However, the variable selection is beneficial because, without it, the algorithm partly divides the contracts into only one or two classes, leading to estimation errors close to those of the OLS regression models. Furthermore, by applying the variable selection, we observe that more underlying contracts tend to be associated with more classes. Compared to the FMMs, the M5' models yield significantly higher MAEs, but they are somewhat competitive in terms of the RMSE.

The RTs tend to divide the contracts into significantly more classes than the M5' models. Nonetheless, the results are predominantly worse than those for the M5' models. The RTs, with all available explanatory variables RT and in combination with variable selection $\mathrm{RT}_{\mathrm{BIC}^*}$, outperform the OLS regressions for most companies. As expected, we notice that punishing the number of classes in $\mathrm{RT}_{\mathrm{BIC}^*}$ results in a model with fewer classes and thereby reduces the prediction quality. Unlike the M5' algorithm, the RT can reduce its error only by increasing the number of classes. Likewise, OLS regression performs better when using all available variables.

The TIC of all considered models remains well below its values of the historical average and mainly less than the value of OLS. Additionally, all of the values are within the range of the suggested threshold value of TIC ≈ 0.15 or less, which confirms that the methods used are worth being considered for estimating LGDs.

2.4.2 Out-of-sample results

Most of the studies in this field report in-sample findings but not out-of-sample results, although the latter are crucial for proper risk management and are required for regulatory purposes. Certainly, our out-of-sample findings, summarized in Table 2.5, differ significantly from the in-sample results. Accordingly, to evaluate the method's efficiency and robustness, out-of-sample testing is essential because in-sample results can be misleading.

	Company A			Company B			Company C		
Method	MAE	RMSE	TIC	MAE	RMSE	TIC	MAE	RMSE	TIC
Hist. avg.	0.3437	0.4022	0.1383	0.3657	0.4221	0.1999	0.3679	0.4200	0.1828
At execution									
OLS	0.3257	0.3891	0.1282	0.2722	0.3469	0.1246	0.3348	0.3959	0.1576
OLS_{BIC}	0.3262	0.3893	0.1285	0.2734	0.3479	0.1256	0.3369	0.3962	0.1583
FMM_{3NN}	0.3539	0.4479	0.1600	0.2917	0.4178	0.1621	0.3593	0.4755	0.2056
$FMM_{J4.8}$	0.3424	0.4422	0.1544	0.2749	0.4004	0.1453	0.3313	0.4193	0.1720
M5'	0.3235	0.3879	0.1271	0.2723	0.3475	0.1250	0.3365	0.3957	0.1576
$M5'_{BIC^*}$	0.3215	0.3873	0.1264	0.2711	0.3467	0.1242	0.3384	0.4004	0.1607
RT	0.3245	0.3890	0.1280	0.2751	0.3490	0.1266	0.3386	0.3961	0.1587
RT_{BIC^*}	0.3243	0.3888	0.1278	0.2746	0.3480	0.1259	0.3386	0.3961	0.1595
At default									
OLS	0.3132	0.3786	0.1206	0.2702	0.3447	0.1229	0.3319	0.3949	0.1560
OLS_{BIC}	0.3143	0.3790	0.1209	0.2730	0.3494	0.1260	0.3334	0.3945	0.1565
FMM_{3NN}	0.3204	0.4214	0.1410	0.2832	0.4085	0.1549	0.3345	0.4504	0.1870
$FMM_{J4.8}$	0.2958	0.3988	0.1253	0.2741	0.3996	0.1464	0.3297	0.4260	0.1732
M5'	0.3100	0.3767	0.1191	0.2678	0.3433	0.1220	0.3332	0.3951	0.1565
$M5'_{BIC^*}$	0.3069	0.3757	0.1178	0.2659	0.3428	0.1206	0.3341	0.3977	0.1579
RT	0.3136	0.3804	0.1216	0.2710	0.3462	0.1244	0.3370	0.3958	0.1583
RT_{BIC^*}	0.3142	0.3811	0.1220	0.2727	0.3474	0.1252	0.3384	0.3979	0.1598

Table 2.5: Out-of-sample estimation errors at the execution and default of contracts by company. The best results are underlined for each company and type of error. Hist. avg. is the historical average loss given default (LGD) used as estimation of the LGD. OLS represents the ordinary least squares regression, and FMM is the finite mixture model in combination with 3-nearest neighbors (3NN), or J4.8. OLS is also performed with the variable selection BIC algorithm and the M5' algorithm and the RT are performed with the variable selection BIC* algorithm. MAE is the mean absolute error defined in Equation (2.4) and RMSE is the root mean squared error defined in Equation (2.5). TIC is the Theil inequality coefficient defined in Equation (2.6). For MAE, RMSE, and TIC, lower outcomes are preferable.

Our findings indicate that, in general, $M5'_{BIC^*}$ generates the best out-of-sample results, although the performance seems to depend on the size of the underlying dataset. Consistent with the in-sample results, we observe predominately more accurate estimations using the additional information available at default.

Concerning the FMMs, we first note that the in-sample favorable FMM_{3NN} is now outperformed by all of the other models and also partly by the historical averages. This outcome is unexpected because the in-sample results are good and sturdy. The $FMM_{J4.8}$ generates isolated good MAE values but the RMSE and TIC values are worse than their counterparts from the tree-based models and OLS regressions. As we did in-sample, we continue to note a large gap between the MAE and RMSE for both FMMs; furthermore, the TIC values exceed the suggested value of 0.15 by several times.

Our results clearly demonstrate that by application of the variable selection procedure, the model tree $\mathrm{M5'}_{\mathrm{BIC}^*}$ is the best choice for companies A and B. For these two companies, applying the variable selection procedure to the model tree algorithm is beneficial without exception. Whereas both model tree methods outperform the RT models, $\mathrm{M5'}_{\mathrm{BIC}^*}$ also generates consistently better MAE, RMSE, and TIC values than both OLS regressions. For company C, we obtain a slightly different picture. Considering the performance measures in total, the OLS regression – particularly using all available explanatory variables – is favorable now. At the very least, at execution, the M5' algorithm generates equally good or even slightly better RMSE and TIC values as the OLS regressions. However, the results of the $\mathrm{M5'}_{\mathrm{BIC}^*}$ and both RT models are worse for company C. Consistent with the in-sample results, we find that the variable selection procedure almost throughout worsens the results of the OLS regression and RT for all companies.

The out-of-sample results suggest to some extend a link between the numbers of observations and the relative performances of the estimation methods considered. Containing the LGD data from three different companies, our dataset provides us with a particularly good opportunity to analyze this relationship in greater detail. Bearing in mind the ranking of the dataset sizes, company A delivered the largest number of observations (9,735), followed by company B (2,995) and company C (1,592).

First of all, we note that the TIC exceeds the suggested value of 0.15 for all of the methods in the case of company C. For companies A and B, TIC values remain well below 0.15, at least with respect to the tree-based models and the OLS regressions. This finding indicates that the prediction accuracy in general becomes weaker if the underlying dataset contains fewer observations. Furthermore, we note that the performances of the model trees relative to the regression model OLS improve with an increasing dataset size. At default of the contracts, M5' performs 0.51% worse than OLS for company C concerning the MAE. But, for company B, M5' performs 0.89% better than OLS and the improvement increases to 1.02% for company A. Analogously, this tendency applies to the RMSE. At execution of the contract the relative performance improves with an increasing sample size only regarding the MAE. Actually, for M5'$_{\mathrm{BIC}^*}$, the estimation accuracy relative to OLS improves throughout monotonically with an increasing sample size. Moreover, the link between the performance relative to OLS and the number of observations included in the underlying dataset is even more distinctive for M5'$_{\mathrm{BIC}^*}$. At default M5'$_{\mathrm{BIC}^*}$ performs 0.66% worse than OLS for company C concerning the MAE, but 1.59% (2.01%) better than OLS for company B (A).

With respect to the RT models we cannot establish an unambiguous link between the performances relative to OLS and the sample size. Compared with OLS, the performances of RT and RT$_{\mathrm{BIC}^*}$ improve with an increasing dataset size only with regard to the MAE. Whereas, regarding the RMSE the results of RT deteriorate relative to OLS with an increasing dataset size at default of the contract and RT$_{\mathrm{BIC}^*}$ obtains the relatively worst outcomes at execution and default of the contract on the dataset of company B. Also concerning the FMMs and OLS$_{\mathrm{BIC}}$ we could not identify a link between the performances relative to OLS and the number of observations contained in the underlying dataset. Both FMMs and OLS$_{\mathrm{BIC}}$ obtain relative to OLS the worst results almost throughout for company B.

We further compare the results for each of the random divisions of the respective dataset used for out-of-sample testing. The findings support the link between the performances of M5' and particularly M5'$_{\mathrm{BIC}^*}$ relative to OLS and the sample size. For company C, M5'$_{\mathrm{BIC}^*}$ yields better results than OLS on only about 5 partitions.

With an increasing sample size, M5'$_{\mathrm{BIC}^*}$ performs better than OLS significantly more often, to be precise, for company B on at least 60% of the divisions and for company A on more than 90%. Actually, at execution of the contract, M5'$_{\mathrm{BIC}^*}$ yields throughout better MAE values than OLS for company A.

Although there might be several factors influencing the estimation accuracy of the models, such as idiosyncratic firm characteristics, we find the sample size to be of particular importance. We apply an additional test to confirm the link between estimation accuracy and sample size. Pooling the three datasets generates a large sample that contains 14,322 contracts (100%). We randomly draw 7,161 contracts (50%) out of the large sample to generate a medium sized sample. For a small sample, we randomly draw 1,432 contracts (10%) out of the large sample. We repeat these random drawings ten times, leaving us with a total of 21 datasets. Again, for out-of-sample testing we split the datasets randomly into 75% training sample and 25% test sample. This step is also done ten times. We showed before, that the M5'$_{\mathrm{BIC}^*}$ seems to be particularly sensible to small sample sizes. Also, bearing in mind that M5'$_{\mathrm{BIC}^*}$ and the regression model OLS perform best for companies A and B, respectively for company C, we focus on testing the impact of sample size for these two models. All results are averaged with respect to the sample size.

We see in Table 2.6 that both methods perform better on larger datasets. Unlike OLS regression, the estimation accuracy of M5'$_{\mathrm{BIC}^*}$ increases almost monotonically with increasing sample size. In particular, the degree of accuracy improvement is clearly higher for the M5'$_{\mathrm{BIC}^*}$. The M5'$_{\mathrm{BIC}^*}$ performs consistently better than OLS regression. The improvement of M5'$_{\mathrm{BIC}^*}$ over OLS regression increases with increasing sample size. At default of the contract M5'$_{\mathrm{BIC}^*}$ performs 1.5% better than OLS regression on small datasets concerning the MAE. The improvement increases to 5.7% and 7.4% on medium sized and large datasets. With regard to

	100% (large)		50% (medium)		10% (small)	
Method	MAE	RMSE	MAE	RMSE	MAE	RMSE
At execution						
OLS	0.3327	0.3977	0.3311	0.3943	0.3359	0.4012
$M5'_{BIC^*}$	0.3186	0.3877	0.3194	0.3869	0.3323	0.4002
At default						
OLS	0.3316	0.3964	0.3285	0.3943	0.3334	0.4001
$M5'_{BIC^*}$	0.3069	0.3783	0.3098	0.3814	0.3285	0.3998

Table 2.6: Out-of-sample estimation errors at the execution and default of contracts by sample size. The sample sizes are 100% (large), 50% (medium), and 10% (small) of all contracts. OLS is the ordinary least squares regression without variable selection and $M5'_{BIC^*}$ is the M5' algorithm with the variable selection BIC^* algorithm. MAE is the mean absolute error defined in Equation (2.4) and RMSE is the root mean squared error defined in Equation (2.5). For MAE and RMSE lower outcomes are preferable.

the RMSE the improvement over OLS regression is 0.1% (2.8%, 4.6%) for small (medium, large) datasets.

Furthermore, we compare the results of $M5'_{BIC^*}$ and OLS regression for each of the 210 randomly drawn subsamples. We find that $M5'_{BIC^*}$ outperforms OLS regression in all drawings on large and medium sized datasets, whereas OLS regression achieves better results on small samples in 28% of the drawings concerning the MAE and in over 40% concerning the RMSE.

We conclude that the $M5'_{BIC^*}$ should be based on an adequately large dataset to process the information more efficiently than the OLS regression. Moreover, the performed test confirms the link between prediction accuracy and sample size. Considering that the dataset of company C contains the fewest observations, and the $M5'_{BIC^*}$ improves with additional observations, we conclude that the $M5'_{BIC^*}$ in general is the best choice for out-of-sample predictions.

At first glance, the differences between the values of the accuracy measurements of the sophisticated estimation methods compared to OLS regression without variable selection seem to be negligible in most cases and are consistent with the results of Zhang and Thomas (2012). This finding raises the question as to whether

Method	Company A		Company B		Company C	
	Exec.	Dflt	Exec.	Dflt	Exec.	Dflt
Hist. avg.	1.0011	1.0011	1.0037	1.0037	1.0013	1.0013
OLS	1.0059	1.0066	1.0052	1.0035	1.0180	1.0236
OLS_{BIC}	1.0049	1.0058	1.0023	1.0125	1.0136	1.0226
FMM_{3NN}	1.2062	1.2151	1.2456	1.2453	1.3202	1.3105
$\text{FMM}_{\text{J4.8}}$	1.1291	1.1095	1.0238	1.0818	1.0713	1.1442
M5'	1.0246	1.0236	1.0046	1.0100	1.0214	1.0175
$\text{M5'}_{\text{BIC}^*}$	1.0262	1.0248	1.0233	1.0461	1.0674	1.0308
RT	1.0191	1.0312	1.0075	1.0264	1.0232	1.0297
RT_{BIC^*}	1.0154	1.0303	1.0167	1.0152	1.0162	1.0239

Table 2.7: Janus quotient for in-sample and out-of-sample estimations of loss given default (LGD) for each method and company at execution (Exec.) and default (Dflt) of contracts. The quotient is calculated according to Equation (2.7) and is constant for the historical average. A Janus quotient greater than 1 indicates that the error for the out-of-sample estimation is greater than the error for the in-sample estimation. OLS represents the ordinary least squares regression, FMM is the finite mixture model in combination with 3-nearest neighbors (3NN), or J4.8. OLS is also performed with the variable selection BIC algorithm and the M5' algorithm and the RT are performed with the variable selection BIC* algorithm.

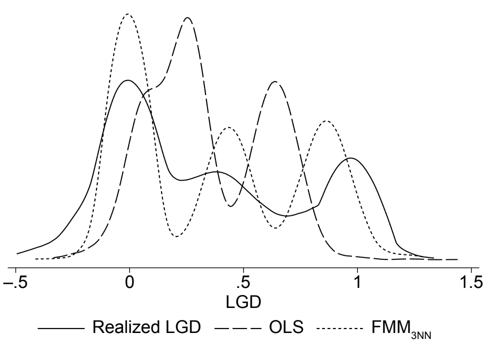
it is worth the effort to implement more demanding estimation methods. For a more illustrative interpretation of our results, we use the average aggregated EAD of our test sample, which is €133,671,554 (€34,762,061) for company A (B) to estimate the total loss for the test sample. Using the $\text{M5'}_{\text{BIC}^*}$ yields an estimation that is in expectation up to €220,000 more accurate than the OLS regression for company B and for company A, the estimation is even up to €1,340,000 more accurate. Thus, improvements of an even few percentage points matter in terms of the parameter outcomes.

Our results indicate that in-sample results are an insufficient indicator of a method's out-of-sample performance. In particular, for the in-sample outperforming FMM_{3NN}, the results are obviously misleading because the out-of-sample predictions are worst. Hence, we study the stability of our models using the Janus quotient, as shown in Table 2.7. According to the Janus quotient, we can partition our methods into stable and unstable methods. A Janus quotient close to 1 indicates a stable model and data structure, which holds for the tree-based models, the OLS regressions, and the historical averages mainly with quotients less than 1.05. Exclusively taking into account the stable models, we observe more or less the same order concerning the estimation accuracy in-sample and out-of-sample. In particular, the $M5'_{BIC^*}$ performs in-sample conspicuously better than the other stable methods for companies A and B. This finding remains valid for the out-of-sample results without exception, only the advantage is smaller. As expected, for the FMMs, a Janus quotient that is mainly considerably greater than 1 indicates that these models are unstable. For the FMM_{3NN}, the quotient consistently exceeds 1.20. Hence, if out-of-sample testing is impossible, e. g. due to an insufficiently large dataset, the in-sample results can be used as a prime indicator of the out-of-sample performance for stable methods, but this relationship obviously does not apply for unstable methods.

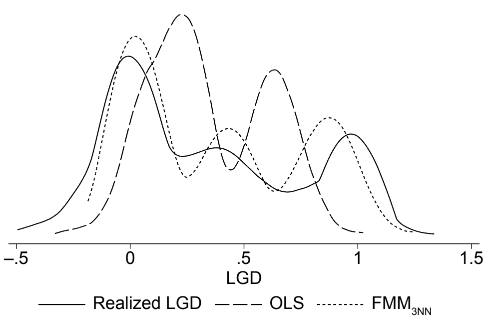
2.4.3 Validation and interpretation

To analyze the models' performances in detail and to elaborate on the several steps of FMMs, we present some key figures of our methods in this section.

FMMs produce accurate in-sample results by aiming to reproduce the distribution density. This relationship is true for both of our FMMs and is independent of the choice of the classification method in step (2). Figure 2.3a displays the realized and estimated LGDs for company B. Whereas OLS regression is not capable of properly accounting for the multimodality of the realized LGD distribution, the FMM's estimation is a good approximation. However, such density representations could be misleading because they are not capable of showing the deviation of an estimate from its realized value. This effect becomes particularly clear when we consider the out-of-sample results of the FMMs. For misclassified observations



(a) In-sample LGD density.



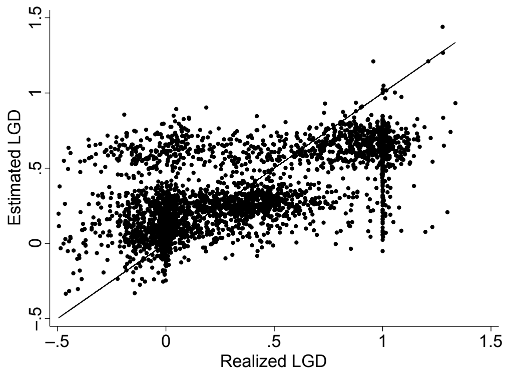
(b) Out-of-sample LGD density.

Figure 2.3: Densities of realized loss given default (LGD), LGD estimated by ordinary least squares (OLS) regression without variable selection, and LGD estimated by finite mixture combined with 3-nearest neighbors (FMM_{3NN}) for company B. The in-sample approximation of the realized LGD distribution by FMM is already good (Figure a) and it even improves in the out-of-sample estimation (Figure b). OLS regression, by contrast, is visibly only slightly changing and is not necessarily improving from in-sample to out-of-sample estimation.

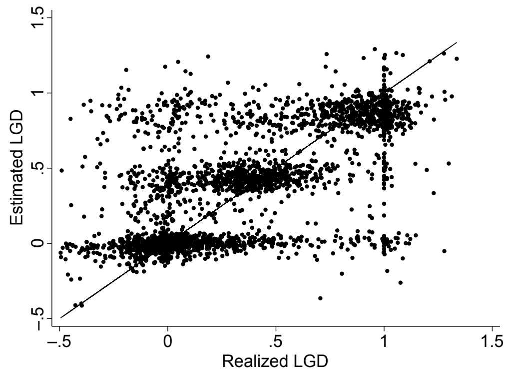
during either the clustering or classification process, the RMSE increases rapidly, while the approximation of the density remains accurate (Figure 2.3b).

The effect can also be observed regarding the scatter plots in Figure 2.4. For both OLS regression and FMM, the in-sample estimation of LGD is rather concentrated around the diagonal in Figures 2.4a and 2.4b. Out-of-sample, we notice for OLS regression in Figure 2.4c that the LGD estimates are thinned out uniformly, which leaves most of its density close to the diagonal. The FMM, by contrast, retains a relatively large amount of its estimates that are far from the diagonal, thus far from the realized value of the LGD. These large deviations consequently result in a larger RMSE. The MAE remains at an acceptable level because most of the density stays on the diagonal.

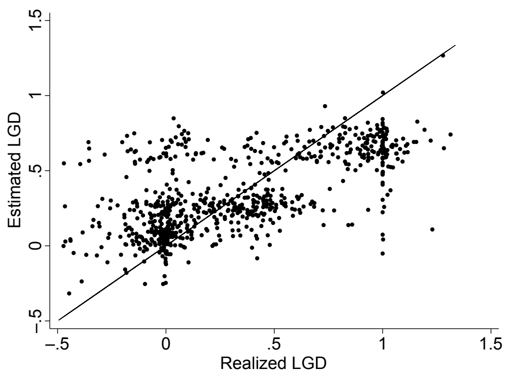
We analyze the quality of FMMs by examining the density of the a posteriori probability of belonging to a certain class, as proposed by Grün and Leisch (2007). The classification becomes more unambiguous as the probability approaches one, which indicates the quality of the adaptation.



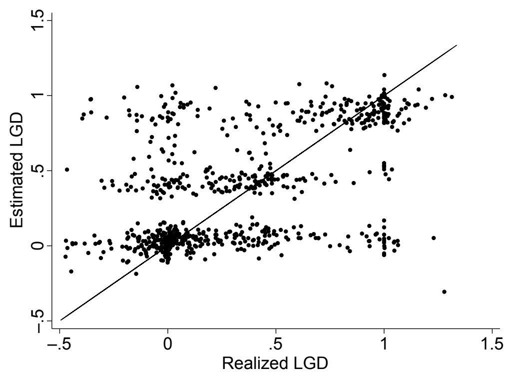
(a) In-sample: realized LGD versus estimated LGD by OLS regression.



(b) In-sample: realized LGD versus estimated LGD by FMM_{3NN}.



(c) Out-of-sample: realized LGD versus estimated LGD by OLS regression.



(d) Out-of-sample: realized LGD versus estimated LGD by FMM_{3NN}.

Figure 2.4: In-sample and out-of-sample: realized loss given default (LGD) versus estimated LGD by ordinary least squares (OLS) regression without variable selection (Figures a and c) and finite mixture combined with 3-nearest neighbors (FMM_{3NN}) (Figures b and d) for company B. Each figure has a simple diagonal line to illustrate the deviation.

For mixing distributions with two clusters, the average in-sample probability that observations are classified into a particular class is at least 88%, whereas the median is close to one. Poorer performance is observed with three clusters because of the larger overlap caused by additional clusters, resulting in lower classification probabilities. However, three-quarters of all observations are classified during the clustering process in step (1), with a minimum probability of 58%.

Validating the classification methods is even more important than validating the clustering in step (1) of the procedure. Although clustering works well for all of the companies, classifying the observations with the information available

	Company A		Company B		Company C	
Method	Exec.	Dflt	Exec.	Dflt	Exec.	Dflt
3NN	0.2176	0.1925	0.2018	0.1834	0.1943	0.1808
J4.8	0.3740	0.3371	0.3693	0.3085	0.2833	0.2634

Table 2.8: In-sample classification errors for the 3-nearest neighbors (3NN) and J4.8 methods at execution (Exec.) and default (Dflt) of the contracts. Given the clustering of the finite mixture model in step (1) (see Section 2.3.1), a contract is classified incorrectly if the classification algorithm (3NN or J4.8) in step (2) assigns this contract to a different cluster. The classification error is then the relative number of falsely assigned contracts.

at the contract execution and default is critical. By reviewing the classification errors of our classification methods, we analyze the performance of these methods. Thus, we can assess the percentage of incorrectly assigned observations. This process only works in-sample because, for unseen cases the true class is unknown. Table 2.8 demonstrates an improvement when we compare classification errors at the execution and default of the contract for both methods. The 3NN approach clearly results in a more accurate classification, which is attributable to the 2-clustered mixing distribution. J4.8 distinguishes among three clusters. These clusters naturally overlap to a significant extent, which results in higher classification errors. The errors are in line with the MAE and RMSE in Table 2.4.

The number of mixing distributions is an exogenous parameter. Our model with three mixing normal distributions constantly produces the smallest AIC and BIC. However, the lower classification error in step (2) might suggest a model with two mixing normal distributions. In terms of MAE and RMSE, neither the parameters, such as the AIC and BIC of the mixing models, nor the in-sample classification error is a consistently good performance indicator for the composed method. Our results in Section 2.4.2 show that the in-sample classification error and out-of-sample MAE and RMSE do not behave proportionally. By contrast, AIC and BIC work well with the out-of-sample results of the $\mathrm{FMM}_{\mathrm{J4.8}}$.

Overall, the large difference between the MAE and RMSE arises from the entire procedure of FMMs, which are focused on accurately mapping the LGD density.

Out-of-sample in particular, the classification is problematic, which becomes obvious in Figure 2.4d, as a large number of estimations is far from the realized value. Therefore, reproduction results in comparatively robust MAEs, but the RMSE rises quadratically and penalizes these outliers.

Reproducing the LGD distribution to yield accurate estimations is proposed by Hlawatsch and Ostrowski (2011). Qi and Zhao (2011), however, conclude that mapping the density is only of minor importance for precisely predicting the LGD. Using transformation regressions under different parameters, they cannot establish a link between the ability to map the density properly and the estimation accuracy, neither in-sample nor out-of-sample. To some extent the results of the FMMs support this conclusion. Nonetheless, there is a significant difference. For instance, the $\text{FMM}_{3\text{NN}}$ generates accurate predictions in-sample and only performs worse out-of-sample. This finding suggests that the $\text{FMM}_{3\text{NN}}$ adapts well to the training data by reproducing the density, but it also indicates that overfitting might be a severe problem.

With regard to out-of-sample predictions, a high level of adaptation to the underlying training data is only reasonable if the training and test data are exceedingly homogeneous. Given inhomogeneous datasets, a good adaptation to the training data basically involves potential overfitting. This relationship is also supported by the results of the model trees. The dataset of company C contains notably fewer observations than those of companies A and B. Moreover, the TIC for company C exceeds the suggested value of 0.15 out-of-sample for all of the methods. Thus, it can reasonably be concluded that the training and test data are comparatively inhomogeneous. $\text{M5'}_{\text{BIC}^*}$ performs strictly better than M5' for all three companies in-sample; in other words, the $\text{M5'}_{\text{BIC}^*}$ attains a superior adjustment to the underlying dataset. Out-of-sample, however, $\text{M5'}_{\text{BIC}^*}$ yields better results only for companies A and B whereas M5' is beneficial for company

C. Hence, given inhomogeneous training and test data, the superior adaptation to the training data is not transferred into sturdy out-of-sample predictions.

For the FMMs, the classification is obviously of prime importance, and out-of-sample in particular, it is problematic. However, classification is also relevant for the tree-based models because the observations are also partitioned into different classes. Certainly, by contrast to the FMMs, the tree-based models use more classes.[2] This increased number of classes indicates that in case of the M5' models, the different classes considerably overlap with one another. For the RTs, the classes are spread over the entire observation interval. In-sample, the number of misclassified contracts is manageable for both the tree-based models and the FMMs. As a result, the latter method mainly yields accurate in-sample estimations (Figure 2.4b), whereas the predictions of the tree-based models, particularly in terms of the MAE, are not as accurate. Naturally, it is more difficult to classify unseen observations correctly. This fact also holds for the tree-based models, although the out-of-sample predictions are significantly better than those of the FMMs. However, based on the tree model's class structure, a misclassified contract tends to be placed into an adjacent class; thus, the resulting error remains low. By contrast, the classes of the FMMs are largely disjointed; thus, the error for a misclassified observation tends to be more significant.

2.5 Conclusion

We use contracts of three leasing companies separately to evaluate various models in-sample and out-of-sample at two different points in time. Our findings prove that out-of-sample testing is essential for evaluating a model for LGD estimation. In-sample results might be significantly misleading when estimating out-of-sample LGDs, which are crucial for proper risk management and are required for regulatory purposes.

[2]The number of classes is chosen by the algorithm and is not defined ex-ante.

FMMs account for the multimodality of the LGD density. Combined with the classification algorithm 3NN, this method achieves the lowest in-sample MAE, RMSE, and TIC values. In particular, it outperforms the historical average and the OLS regressions, which were used as benchmarks. Along with the FMMs, the model tree with variable selection $M5'_{BIC^*}$ yields the best results for in-sample estimation.

Out-of-sample, a clear trend can be observed that model trees and particularly $M5'_{BIC^*}$ generate the best results. Compared with OLS regression the performance of $M5'_{BIC^*}$ improves notably with an increasing dataset size. We confirm this result by applying an additional test, in which we eliminate idiosyncratic features by pooling the three datasets. Furthermore, for the company with the fewest observations, the TIC values indicate that all applied methods have difficulties predicting the LGD of unseen contracts accurately. As opposed to in-sample results, FMMs now are outperformed even by the OLS regression; in particular, FMM_{3NN} performs worst.

The Janus quotient determines the stability of our models, dividing them into stable and unstable methods. In particular, the in-sample results of unstable methods, namely the FMMs, cannot be used as indicators for out-of-sample estimation errors.

3 The impact of debtor recovery on loss given default

The risk of a debtor's default is one of the main risks financial institutions take. Its extent and complexity require considerable expertise and resources in the risk management. If this risk becomes effective, a workout process launches. Particularly, this process intends to limit the financial damage to the institute. However, defaulted debtors can recover. In such cases, the institute often suffers only small losses or no loss at all.

The recovery of defaulted customers is an interesting event in the credit risk management of financial assets. Financial contracts basically develop in one of two ways after the default of the customer. Either any collateral is being liquidated, the remaining amount is written off, and the contractual relationship comes to an end or the customer recovers and the contract can continue properly. We find that the development of the contract is particularly associated with the level of its loss given default (LGD).

According to the capital requirement regulation (CRR) Article 178, the default of a customer can be triggered by the following events: the lender considers that the debtor is unlikely to pay his credit obligations to the lender in full, or the debtor is past due more than 90 days on any material credit obligation to the lender. A formally defaulted contract recovers if these triggered default reasons no longer exist.

So far, few studies have put effort into accounting for recovered customers and their impact on LGD. This circumstance may have several reasons: the lack of information about contracts' recoveries; lenders either do not account for recovery; lenders write-off defaulted contracts timely in the majority of cases; or lenders principally do not consider continuation of such contracts. Indeed, at the time of default it is not clear whether the write-off or recovery of a defaulted contract leads to higher returns. Although, the LGD of recovered contracts on average is lower compared to written off contracts, the influence of contracts' recovery cannot simply be generalized. In fact, recovery or write-off of a contract is an endogenous event primarily reflecting the lender's workout policy and other latent influences. While writing off defaulted contracts requires competencies in asset disposal, a rather different capability is needed when considering to continue the contract. Lenders, who in principal consider the recovery of defaulted debtors, should be able to evaluate the worthiness of continuing their contracts properly.

In the literature of LGD estimation, Han and Jang (2013) analyze the effects of debt collection practices and find that foreclosure and seizure of credit loans reduce the LGD on average while individual rehabilitation increases it. By accounting for various actions during the workout process, they can improve the estimation of LGD significantly. Carried out by the lessor, these practices and other lessee specific circumstances may be aggregated in the single information of the contract's recovery or write-off. The outcome depends on how useful these practices are to resolving the lessee's default reason. A contract's default outcome can be determined as soon as its workout is completed or, in case of recovery, after its conclusion.

Recently a number of studies set their focus on the estimation of LGD, which is required in the advanced internal ratings based approach. An accurate estimation of the LGD is important for the appropriate allocation of regulatory and economic capital. We find that the accuracy of LGD estimation can be improved by the distinction between recovered and written off contracts during the workout process.

Similar to the event of default considered in CRR Article 181 (1)(h), the recovery of the contract changes the contract's current economic circumstances. In this study, we confirm a significant effect of recovery on the LGD. Hence, when forecasting the LGD, we account for a major event in the contractual relationship between lender and debtor other than the default of the contract.

By now, a set of different approaches to estimate LGD has been applied in the literature. Primarily, these focus on loans and bonds. Several studies try to establish a direct link between the LGD and the available predictors, such as Han and Jang (2013), Bastos (2013), and Altman and Kalotay (2014). Using Moody's Ultimate Recovery Database, Altman and Kalotay (2014) find that a mixture of Gaussian distributions outperforms the chosen parametric and nonparametric estimation methods. Employing the same database, Bastos (2013) presents a bagging based algorithm, which combines a number of models to an ensemble learner. Thereby, the estimation accuracy improves compared to the models' single versions. Other authors, such as Loterman et al. (2012), use multi-step models. In the first step Loterman et al. (2012) cluster the data by its LGD, separating contracts at a threshold of zero by logistic regression (Logit). In the second step, they estimate the LGD by linear regression and adjust these estimates with an additional estimation of the residuals. Contracts with LGD equal to zero are assigned an estimated LGD of zero.

Similar to Loterman et al. (2012) we compare a series of two-step models in this study. The models' purpose is to classify recovered and written off contracts and to estimate the LGD for both classes separately. Therefore, we use several advanced classification tree methods, among others Breiman's random forest (RF) and Quinlan's C5.0. By classifying the contracts, we determine the lender's workout policy measured in terms of recovery or write-off of the contracts. We find that the distinction between recovered and written off contracts, as well as the advanced techniques used in the study, increase the explanatory power of LGD variation significantly. The increase in LGD estimation precision should be particularly beneficial to the risk-adjusted calculation of contract prices and appropriate allocation of capital. For both steps of the model, we discuss out-of-sample and out-of-time results at the contracts' execution and default. The unique dataset comprises more than 42,000 contracts in total. The three datasets are inhomogeneous in terms of size, available information, leased assets, and distribution and level of LGD. As indicated by Hartmann-Wendels et al. (2014), these features should proof the robustness of our methods.

We find that recovered contracts on average have a low LGD. This finding should be of particular interest to the lender, reducing his capital tied-up in backing the contract. In addition, the accuracy of LGD estimates increases when considering the different contracts' default ends. Moreover, LGDs of recovered and written off contracts have different drivers and contract characteristics. If separated, we can better account for these characteristics. Finally, with an increasing number of successfully recovered customers, the lender might be able to improve his general customer satisfaction and consequently his reputation.

3.1 Dataset

This study uses a dataset provided by three major German leasing companies. It contains 1,106 defaulted leasing contracts with 670 different lessees from company D, 2,376 contracts with 1,294 lessees from company E, and 39,093 contracts with 23,748 lessees from company F. Table 3.1 displays that the contracts defaulted between 1993 and 2010. The workout process of all contracts has been completed.

The datasets contain numerous information about the contract, customer, leased asset, additional information at the default of the contract, and further information after default. Information about the contract, customer, and asset provides a detailed description of the observations that are already available at the execution of the contract. At the default of the contract, additional information becomes

Company	# Contracts	# Lessees	Year of default
D	2,376	1,294	2002–2010
E	1,106	670	2000–2005
F	39,093	23,748	1993–2004

Table 3.1: Number of contracts and lessees in the datasets of companies D–F in descending order of the most current default year.

available. In particular, this information includes the exposure at default (EAD) and mostly the reason of default, which corresponds to the default trigger events set out in the CRR framework. The category further information provides essential information about the contract during the workout process, such as cash flows and the date of recovery.

Table 3.2 summarizes the described categories. Accordingly, company D provides the most comprehensive information about its contracts, assets, and lessees. Company E holds detailed information about its assets and the contract specifics. Finally, company F provides the asset type, its initial value, and the EAD.

The datasets differ not only in the time range of their contracts' defaults, but in particular in their leased assets and the quantity of information of each contract. Company D's contracts show no particular specification in single asset types. It finances vehicles, machinery, information and communications technology (ICT), and other equipment. Company E, however, exclusively leases passenger cars. Company F provides contracts on passenger cars and ICT.

Employing the EAD and the further information, we calculate the LGD of each contract. We discount all cash flows to the time of default using the term related refinancing interest rate. The EAD is the sum of the present value of contractually outstanding lease payments. The LGD then is the ratio of the discounted cash flows to the EAD.

Over time, default and recovery of a single contract can occur multiple times. The contract's first default and recovery, if any, is assigned to the contract. Af-

At execution			At default	At completion
Contract	Asset	Lessee	Additional	Further
Type$^{\mathrm{D,E,F}}$	Type$^{\mathrm{D,E,F}}$	Type$^{\mathrm{D}}$	EAD$^{\mathrm{D,E,F}}$	Overdue pay.$^{\mathrm{D,F}}$
Calc. interest rate$^{\mathrm{D,E}}$	Purchase price$^{\mathrm{D,E,F}}$	Industry$^{\mathrm{D}}$	Date$^{\mathrm{D,E,F}}$	Asset dis. pay.$^{\mathrm{D,F}}$
Assessment base$^{\mathrm{D,E}}$	Calc. residual value$^{\mathrm{D,E}}$	Internal rating$^{\mathrm{D}}$	Reason$^{\mathrm{D,E}}$	Coll. dis. pay.$^{\mathrm{D,F}}$
Maturity$^{\mathrm{D,E}}$	Manufacture date$^{\mathrm{D}}$	Legal form$^{\mathrm{D}}$	Remaining term$^{\mathrm{D,E}}$	Workout costs$^{\mathrm{D,F}}$
Leasing rate$^{\mathrm{D,E}}$	Useful life$^{\mathrm{D}}$		Asset age$^{\mathrm{D,E}}$	Recovery date$^{\mathrm{D}}$
Market interest rate$^{\mathrm{D}}$	Depreciation range$^{\mathrm{D}}$		Asset value$^{\mathrm{D}}$	Completion date$^{\mathrm{D}}$
Conn. agreements$^{\mathrm{D}}$	Second hand$^{\mathrm{D}}$		Collateral value$^{\mathrm{D}}$	LGD$^{\mathrm{E}}$
Collateral$^{\mathrm{D}}$	Car specifics$^{\mathrm{E}}$		Add. car specifics$^{\mathrm{E}}$	
Rent prepayment$^{\mathrm{D}}$				
Payment cycle$^{\mathrm{D}}$				

Table 3.2: Categorized information contained in the datasets of companies D–F. Contract, asset, and lessee information are available at the execution of the contract. Additional information becomes available upon the default of the contract, and further information becomes available at its completion. Type is the type of the contract, asset, or lessee, e. g. full payment lease, car, or retail respectively. Calc. abbreviates calculated, conn. abbreviates connecting, coll. abbreviates collateral, add. abbreviates additional, dis. abbreviates disposal, and pay. abbreviates payments. The letters D–F next to each information indicate whether this information is contained in the respective dataset.

ter the completion of the workout process, the LGD can always be calculated using the EAD and contract related incoming and outgoing cash flows. If the contract recovers during the workout process, we use the EAD determined at its first default. In this case, we can ultimately calculate the LGD after the contract's conclusion. This way, LGD can be determined for every defaulted contract, independent of whether the workout process ended in its recovery or write-off. In the case of company E, the LGD has been provided along with the dataset.

Table 3.3 provides a brief overview of the distribution parameters of LGD. For each company, we see that more than one-third of all defaulted contracts recovered during the workout process. These recovered contracts have a mean LGD, which is substantially lower than that of written off contracts and even lower than 0 for company D. That means the lessor regains more than the EAD. On average, this gain amounts to a profit of €4,225 per contract for company D.

There are several reasons for such low LGDs of recovered contracts. Hitting only the first default trigger of Article 178 CRR, the lessee might recover without

Status	# Ctrcs	Mean	Std	Min.	P25	Median	P75	Max.
Company D								
Recovered	842	−0.08	0.45	−1.15	−0.31	−0.07	0.08	1.06
Written off	1,534	0.34	0.48	−1.36	0.00	0.28	0.74	1.50
Overall	2,376	0.19	0.51	−1.36	−0.09	0.08	0.57	1.50
Company E								
Recovered	541	0.13	0.12	−0.45	0.07	0.13	0.19	0.65
Written off	565	0.23	0.14	−0.23	0.14	0.21	0.31	0.87
Overall	1,106	0.18	0.14	−0.45	0.10	0.17	0.25	0.87
Company F								
Recovered	15,345	0.00	0.01	−0.49	0.00	0.00	0.00	0.00
Written off	23,748	0.44	0.41	−0.50	0.00	0.41	0.85	1.50
Overall	39,093	0.27	0.39	−0.50	0.00	0.00	0.57	1.50

Table 3.3: Loss given default (LGD) density information of recovered and written off contracts for companies D–F. # Ctrcs is the number of contracts, Std is the standard deviation, and P25 and P75 are the respective percentiles. Min. and Max. are the minimum and maximum LGD values.

causing any financial damage. In case, payments are overdue more than 90 days, recovery still might lead to small losses because outstanding payments are only delayed rather than omitted. Likewise, arranging a new payment plan restricts losses to manageable amounts.

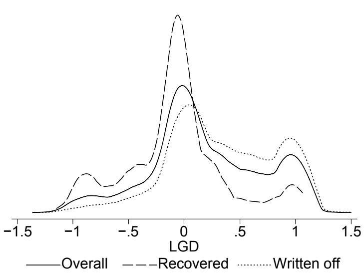
Contrary to recovered contracts, defaulted and written off contracts have a much higher mean LGD. They lose more than 23% of the EAD. Again, in case of company D this loss amounts to €17,584 per contract on average. Mainly, high LGDs of written off contracts result from low cash inflows from overdue payments and asset disposal. High costs of the asset's disposal and payments collection additionally increase the LGD.

We see in Table 3.3 that the range of LGD values considerably exceeds the unit interval. In fact, for over 50% of the cases LGD is less than 0 or equal to 0. 71% of these cases are contracts that recovered after default. Likewise, the LGD of recovered contracts rarely exceeds 1. In contrast, written off contracts have LGDs larger than 1 frequently.

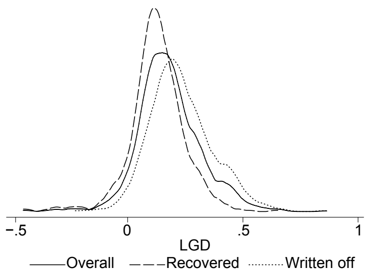
Section 272 of the Committee of European Banking Supervisors (2006) indicates that defaulted positions may generate no loss, or even positive outcomes, i. e., negative LGDs. Hence, no loss might occur, if an exposure recovers with no associated costs and no loss due to discount effects. Negative LGDs in loan and leasing contracts are well known and have been observed by Laurent and Schmit (2005) and Loterman et al. (2012). The latter argue that the reasons for negative LGDs include paid penalties and gains in collateral sales. The authors further state that workout costs can increase the LGD to more than 1 if considered. Especially in leasing Hartmann-Wendels and Honal (2010) find that a negative LGD results from a rather small EAD, which is smaller than the proceeds from the asset's sale.

With the last argument in mind, it is not intuitively clear how recovered contracts can achieve a negative LGD because the asset cannot be liquidated. A reasonable explanation might be that in case the defaulted lessee is granted an extended payment maturity by restructuring his payment plan, the lessor increases his interest income while the EAD remains unchanged. The total income then would exceed the EAD. Furthermore, only the contract's first default and recovery are captured in the data. Hence, any contract may have defaulted a second time and the leased asset could have been sold afterward. In this case, again, the argument of Hartmann-Wendels and Honal (2010) is applicable.

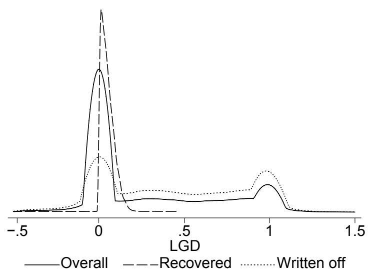
Figure 3.1 presents the densities of LGD for the whole dataset, for recovered, and for written off contracts. The overall LGD, as well as the LGD of written off contracts of companies D and F is bimodally shaped with high concentrations around 0 and 1. However, in all cases the LGD of recovered contracts is rather normally distributed around 0. In case of company D, it also has heavy tails with peaks around -1 and 1. Company E's LGD is most dense on the interval $(0, 0.5)$. Having only one peak, the LGD is rather normally distributed. The difference in company E's distribution arises from its specialized asset portfolio, which contains



(a) Company D.



(b) Company E.



(c) Company F.

Figure 3.1: Density of loss given default (LGD) of the total dataset and recovered and written off contracts of companies D–F. The LGD concentrates on the interval $(-0.5, 1)$. The figures describe a loss severity of -150% (-1.5) on the left end, which indicates that 250% of the EAD is recovered. On the right end, the loss severity is 150% (1.5), indicating a loss of 150% of the EAD. Consequently, a realized LGD of 0 or 1 indicates the following: in case of 0, full coverage of the EAD; or, in case of 1, total loss of the EAD.

passenger cars only. Extracting vehicles from the datasets of companies D and F results in a similar distribution of LGD of these contracts. On the one hand, the large overlap of recovered and written off LGDs of company E might be challenging when it comes to classification. On the other hand, the multimodality found for companies D and F might be challenging in terms of LGD estimation.

The distribution of LGD of company E in Figure 3.1b demonstrates that different assets yield different levels of LGD. This effect mostly depends on the marketability of the assets and the disposal competence of the lessor. Therefore, it would be reasonable if recovery or write-off of a defaulted contract depended on the type of asset. Likewise, it is possible that macroeconomic factors influence the rate of recovery or write-off. However, accounting for different asset types as well as unemployment rate and gross domestic product, we find no empirical evidence providing any of these connections.

Gürtler and Hibbeln (2013) argue that the recovery of loans is mainly liquidity-based and not linked to the value of collateral. Unfortunately, we lack information about the lessors financial situation. However, unlike bank loans, lessors retain legal title to the leased asset. Hence, we attribute the recovery event mainly to the lessor's workout policy.

For complementary analysis of the workout policy, we regard three additional contract portfolios provided by distinct German leasing companies. These portfolios are used for descriptive purposes only because all 35,476 defaulted contracts have been written off. The mean LGDs range from 0.48 to 0.56 and, thus, exceed those of companies D–F by far. This outcome shows that the workout policy of exclusively writing off defaulted contracts has a strong influence on the level of LGD.

3.2 Methods

Evidence on an optimal procedure to estimate LGD is yet scarce. Also, the factors favoring or deteriorating the recovery of defaulted contracts in the existing literature often vary and are rarely considered in detail. However, in accordance with Gürtler and Hibbeln (2013), we find that accounting for differences in recovered and written off contracts is beneficial to the accuracy of LGD estimation.

To be able to distinguish between recovered and written off contracts and measure the accuracy improvement, the used estimation models consist of two steps. We illustrate these in Figure 3.2: (1) classification into recovered and written off contracts; and (2) estimation of LGD. Meeting the requirements of CRR Article 181 (1)(h), we perform both steps at two points in time: execution and default of the contract. We validate the models' performance by in-sample, out-of-sample, and out-of-time testing. This procedure shall provide confidence in the accuracy and robustness of the estimates, which is required by CRR Article 179 (1)(d). Kaastra and Boyd (1996) claim that both out-of-sample and out-of-time testing are particularly important to assess a model's generalization ability, i. e., its practical applicability. Furthermore, Hartmann-Wendels et al. (2014) find that in-sample estimation accuracy is not a reliable indication of a model's capability to handle unseen cases.

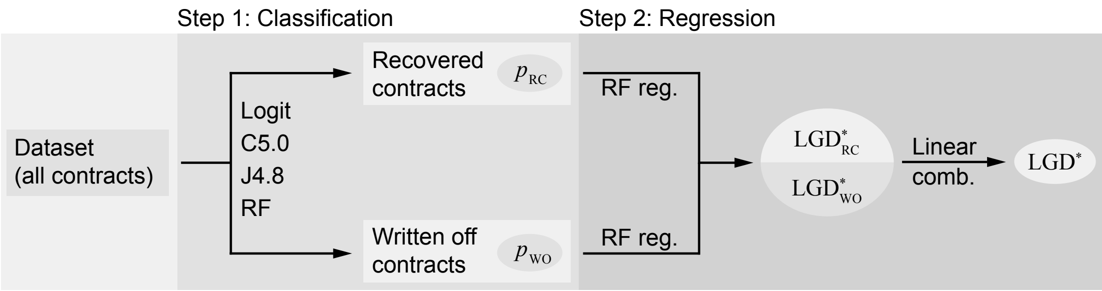


Figure 3.2: Procedure of the two-step model. In step (1), the data of each company is classified into recovered (RC) and written off (WO) contracts. This allocation is done by the classification methods logistic regression (Logit), J4.8, C5.0, or random forest (RF). In addition, the contracts' class probability p is issued. In step (2) we train two RF regression models separately, one on each subset of recovered and written off contracts. The LGD then is estimated twice for each contract, once with the regression model trained on recovered contracts and additionally with the regression model trained on written off contracts. Using the class probability from step (1), we weight the estimated LGDs in a linear combination to calculate the final LGD estimation LGD* according to Equation (3.1).

We compare different models to forecast the LGD of defaulted contracts. For classifying whether a contract will recover we use logistic regression (Logit), as well as a series of tree-based classification models. These are J4.8, C5.0, and random forest (RF).

The estimation model in step (2) incorporates the classification of contracts from step (1) to train the model and estimate the LGD. In addition to the contract's class, we estimate the particular class probability of each contract in step (1). This continuous variable contains more information about the estimated recovery of a contract than the categorical class variable. In simple words, to estimate the LGD, we use the variables available at the respective point in time (see Table 3.2) enhanced with the prediction of recovery from step (1). This approach should be beneficial to the accuracy of LGD estimation if the classification of the contracts is successful.

3.2.1 Tree algorithms

In order to get an impression of the functionality of the used tree algorithms, we provide a brief description. Additionally, we discuss parameter settings.

J4.8 and boosted J4.8

J4.8 is a modification of the C4.5 algorithm implemented by Quinlan (1993). Witten et al. (2011) describe the former in detail. The algorithm generates classification trees by partitioning the dataset to produce subsets with increased homogeneity. Every partitioning results in a node. At each node, only the split with maximized gain ratio is performed until a minimum of instances per node is reached, or the gain ratio does not reach a set minimum value. Such tree structure, which shall not influence the classification accuracy, is pruned back to prevent overfitting. This pruning is especially beneficial to the out-of-sample and out-of-time estimation accuracy of the model.

To enhance the performance of J4.8, we use AdaBoost.M1, a boosting algorithm introduced by Freund and Schapire (1996). The algorithm generates a sequence of weak classifiers on an evolving training sample. In this sample, misclassified cases are successively assigned a higher weight to increase their consideration in subsequent classifications.

C5.0 and boosted C5.0

C5.0 is Quinlan's latest version of the tree algorithm C. Until he released the algorithm to the public in 2011, it was almost exclusively used for commercial purposes but rarely in scientific research. In May 2013, Kuhn et al. (2013) implemented C5.0 into the R project for statistical computing. The algorithm is the successor of its previous version C4.5, which Quinlan (1993) describes in detail. C5.0 inherits the basic function of classifier trees: repeated splitting of the dataset. However, splitting criteria and calibration options have been revised or added. One of many advantages over the previous version is boosting, which we are using in this study. Kuhn and Johnson (2013) provide further information on the C5.0 algorithm and its features.

The native boosting algorithm is similar to AdaBoost.M1 by Freund and Schapire (1996). However, it is suited to C5.0 classification trees. According to Kuhn and Johnson (2013), it has some notable differences compared to its predecessor. These are: C5.0 creates about equally sized trees with respect to the number of terminal nodes; it combines the weak classifiers differently; and it has automatic stop conditions for very effective and very ineffective models.

Random forest

Random forest is a tree algorithm introduced by Breiman (2001). It bases on the idea of bootstrap aggregated predictions of many models and combines these in a voting. Similar to the ensemble model of Bastos (2013) each tree is trained on a random subsample of the data. Additionally, at each node of the tree a random set of the available variables is drawn. The algorithm selects the best split among these variables. A contract's class then is determined by majority vote, counting the frequency of assignments in the individual trees. In the case of regression instead of classification, the terminal node's average value is assigned to each contract in the node. We use $\sqrt{m}$ randomly chosen variables for classification and $m/3$ variables for regression with RF, where m is the number of available variables. Throughout the study, we train 2,000 trees, which is double the number proposed by Hastie et al. (2009). In our case, growing a forest that is even larger than 2,000 trees increases the computational time noticeably. However, it improves the model's performance only marginally.

AdaBoost.M1, C5.0's boosting, and RF, which uses bagging, pursue a similar idea of voted decision finding. Therefore, they should be most comparable in contract classification. Quinlan (1996) compares boosting with bagging and analyzes the procedure and performance of both methods in detail.

3.2.2 Regression model

After classifying the dataset into two distinctive classes, we calibrate two separate models to estimate the LGD. Thereby we take into account the different LGD distributions and contract characteristics of recovered and written off contracts. The first model exclusively bases on recovered contracts, the second on written off contracts. All contracts of the test sample then are assigned two LGD estimations. We weight these estimations with the probability of recovery from the classification in step (1) of the analysis

$$\mathrm{LGD}^* = p_{\mathrm{RC}} \cdot \mathrm{LGD}^*_{\mathrm{RC}} + p_{\mathrm{WO}} \cdot \mathrm{LGD}^*_{\mathrm{WO}}, \tag{3.1}$$

where p_{RC} is the estimated probability of recovery and $p_{\mathrm{WO}} = 1 - p_{\mathrm{RC}}$ is the estimated probability of a write-off. $\mathrm{LGD}^*_{\mathrm{RC}}$ is the LGD estimated by the regression model based on recovered contracts and $\mathrm{LGD}^*_{\mathrm{WO}}$ is the LGD estimated by the model based on written off contracts.

Following the suggestion of Hartmann-Wendels et al. (2014) for multimodally shaped response distributions, we fit a tree model to both recovered and written off contracts, namely an RF regression model. In fact, this tree regression model outperforms simple linear regression in step (2) of the method in all cases.

3.2.3 Model testing

Most empirical studies on LGD estimation provide in-sample results of their fitted models. Recently numerous studies started to provide out-of-sample performance results and LGD estimation errors, such as Bastos (2013) and Li et al. (2014). Still few studies are testing out-of-time, such as Han and Jang (2013). Here a sufficiently large dataset is crucial and requires several time periods to train and test a model's performance. Reasons for not testing the models out-of-sample or out-of-time might be insufficiently large datasets in order to establish an adequate training and testing base or a data history that is too short.

In-sample the models are trained and tested on the full dataset. Certainly this method accounts best for most realized contract specifics but might suffer from an overfitting to the data and, thereby, distort the estimations for unseen cases.

For the out-of-sample testing, we use a method proposed by Hartmann-Wendels et al. (2014). We partition the dataset randomly into 75% training set and 25% test set. The random partitioning is repeated 25 times. Consequently, 25 models are fitted with the training sets and validated on the test sets. Finally, the results of the error measurements are averaged. Other methods for out-of-sample testing are k-fold cross-validation (see Qi and Zhao (2011)) and a single partitioning of the dataset (see Gürtler and Hibbeln (2013)).

The objective of out-of-time testing is to evaluate the model's ability to forecast the next period based on observations of prior periods. In their popular LossCalc^{TM} version 2 model Gupton and Stein (2005) use walk-forward testing to assess the generalization ability of their model. While the original idea of walk-forward testing is a sliding window that uses the last s periods to forecast the next period, Gupton and Stein (2005) append their training set with each period. Thereby, they produce an ever increasing window.

Our datasets cover periods of at least five consecutive years with defaulted contracts between 1993 and 2010. We adopt the walk-forward approach used by Gupton and Stein (2005) to determine an adequate model in the out-of-time testing. Hence, we meet the requirement of CRR Article 181 (1)(j) for companies D and F, which demands a data observation period of seven years for LGD estimates.

3.3 Results

For the two steps of this study's procedure, we use different error measurements. Step (1) classifies the observed contracts into two classes: after default the contract either recovers or is written off. Given that we already know the recovery outcome because all considered contracts are completed, we can always verify the correctness of the classification.

Correctly classified contracts are such, which recovered or were written off and are estimated to recover or to be written off respectively. The probability of correctly classifying recovered contracts is the model's sensitivity. The probability of correctly classifying written off contracts is the model's specificity. Consequently, all other contracts are misclassified, such as contracts that were written off during the workout process but were estimated to recover and vice versa. We calculate the classification error as the ratio of misclassified contracts with all contracts.

To evaluate our results in step (2), the estimation of LGD, we find the coefficient of determination R^2 to be adequate for measuring the quality of LGD prediction. The out-of-sample R^2 statistic proposed by Campbell and Thompson (2008) and used by Gürtler and Hibbeln (2013) for evaluating LGD estimation models, is computed as

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(\mathrm{LGD}_i - \mathrm{LGD}_i^*\right)^2}{\sum_{i=1}^{n} \left(\mathrm{LGD}_i - \overline{\mathrm{LGD}}_{\mathrm{Is}}\right)^2}, \tag{3.2}$$

where LGD denotes the realized LGD, LGD^* is the predicted LGD, $\overline{\mathrm{LGD}}_{\mathrm{Is}}$ is the average in-sample LGD, and n is the number of contracts.

While a low classification error is preferable, the LGD prediction model's determination improves with higher values of the coefficient of determination R^2.

Method	Company D Is	Company D Oos	Company D Oot	Company E Is	Company E Oos	Company E Oot	Company F Is	Company F Oos	Company F Oot
At execution									
Logit	0.2942	0.3224	0.4187	0.3671	0.3895	0.6039	0.3925	0.3905	0.4300
C5.0	0.1023	0.2561	0.4168	0.0940	0.2261	0.4380	0.3814	0.3805	0.4215
C5.0 Boosted	0.0758	0.2328	0.4056	0.0000	0.1668	0.4318	0.3815	0.3800	0.4202
J4.8	0.0791	0.2608	0.4314	0.0542	0.2145	0.4468	0.3790	0.3806	0.4199
J4.8 Boosted	0.0093	0.2305	0.4125	0.0054	0.1851	0.4341	0.3790	0.3806	0.4199
RF	0.0093	0.1900	0.3878	0.0000	0.1355	0.3877	0.3785	0.3797	0.4193
At default									
Logit	0.2858	0.2975	0.3566	0.3092	0.3212	0.5824	0.3926	0.3905	0.4300
C5.0	0.0509	0.1823	0.3538	0.0497	0.1906	0.4253	0.3788	0.3788	0.4183
C5.0 Boosted	0.0051	0.1600	0.3281	0.0000	0.1682	0.4219	0.3791	0.3784	0.4213
J4.8	0.0539	0.1948	0.3083	0.0506	0.1929	0.4070	0.3744	0.3787	0.4163
J4.8 Boosted	0.0000	0.1604	0.3046	0.0000	0.1823	0.4055	0.3744	0.3787	0.4163
RF	0.0000	0.1341	0.2895	0.0000	0.1239	0.3848	0.3686	0.3778	0.4141

Table 3.4: Classification errors at execution and default of the contracts of companies D–F. The listed methods classify the contracts into recovered and written off. The error is calculated according to Section 3.3. We validate the estimates in-sample (Is), out-of-sample (Oos), and out-of-time (Oot). Logit is the logistic regression. The tree classifiers C5.0 and J4.8 are performed in single and boosted version. RF is the random forest algorithm. In all cases, lower outcomes are preferable. We underline the best results for each testing method, both points in time, and each company.

3.3.1 Recovery classification

In step (1) of the analysis, we predict a contracts recovery or write-off. In order to classify the contracts into these two groups, we use Logit and the tree algorithms J4.8, C5.0, and RF. The models' classification errors are calculated according to Section 3.3.

In Table 3.4 we find Logit to be strictly dominated by the other methods. For companies D and E, this dominance becomes particularly obvious in the in-sample testing. Here Logit produces an error that is about three times as large as the error of any other method. Except for Logit, we notice that the in-sample error is substantially lower than the out-of-sample error, which in turn is lower than the out-of-time error. This ranking comes at no surprise because clearly the in-sample model accounts for all realized contract information, but the out-of-sample model classifies unseen cases. Moreover, out-of-time the impact and level of particular contract information may change considerably. This change complicates classification even more.

The classification trees J4.8 and C5.0 perform similarly well while J4.8 is slightly advantageous. Both boosted versions somewhat improve the error compared to their single versions. The improvement of boosting J4.8 and C5.0 is most evident in the in-sample classification. Here, it often reduces the classification error to zero. We observe the same effect in the out-of-sample and out-of-time testing. However, the level of improvement is lower than before.

RF is most successful in classifying the contracts. Despite the other tree methods, it builds rather small trees, using only a fraction of the available variables. These variables are drawn randomly at each node of a tree. Therefore, various variable interactions occur, which take into account even seemingly non-influential variables. The model tree's randomly changing training set reduces the risk of overfitting the RF model to the data. Thereby, its generalization ability increases.

We find the classification accuracy of each method to rise significantly by adding the additional variables available at the default of the contract. The performance increase in out-of-time estimation on company D's dataset is particularly striking. The error reduces by up to 28% when adding this information. However, there are a few exceptions to this rule, where a particular algorithm cannot enhance its accuracy. Such is the case, if the additional variables have no explanatory power in the respective model and, therefore, are not included. Still, the classification accuracy of the models remains in order, demonstrating the superiority of RF over the other models.

The LGD distribution of recovered and written off contracts of company E overlap largely. However, the classification error with respect to out-of-time estimation is mostly only slightly above its counterparts of company D and F at contracts' execution. While this small difference might be surprising, we see a much wider gap between the out-of-time results of company D and E at the default of the contracts. Here, company E's classification error can hardly improve because essential information is missing, such as collateral and asset value at default. The reason is that a specialized leasing company usually monitors and approximates its asset's fair value very precisely at any time during the lease. Moreover, it can dispose of the asset at this or close to this price. The market for vehicles is particularly liquid. Therefore, the disposal would also be completed shortly after the default of the contract. Hence, to know the asset's value at default is very beneficial to the LGD forecast.

Company F's classification results meet all former arguments, but the methods produce a classification error on a higher level. The main reason for the increased level of LGD is that company F provides a very small and, in this case, the smallest set of information (see Table 3.2). This little information complicates the distinction between the contracts. All methods struggle to fit an appropriate classification model because initially only a few contract characteristics are available. These are the type of contract and asset as well as the asset's purchase price. Furthermore, at the default of the contract the only additional information is EAD. However, EAD hardly improves the classification results because the majority of contracts are still being misclassified.

3.3.2 Loss given default estimation

In Table 3.5 we compare our two-step models based on their explanatory power of LGD variation according to Equation (3.2). The LGD estimation results are in line with the classification accuracy in Table 3.4. In particular, RF mostly outperforms the other models in all testing methods for companies D and E. It also confirms the difficulties in handling company F's dataset. As before, boosting has a slight advantage over the single versions of J4.8 and C5.0.

Negative R^2 values indicate that the model cannot explain the variation of the LGD. The out-of-time models for companies D and E are particularly affected by

	Company D			Company E			Company F		
Method	Is	Oos	Oot	Is	Oos	Oot	Is	Oos	Oot
At execution									
Direct OLS	0.1807	0.1045	−0.0445	0.1727	0.1166	−0.0487	0.0154	0.0153	0.0157
Direct RF	0.7672	0.2202	−0.0943	0.7127	0.1316	−0.0876	0.0154	0.0153	0.0157
Logit	0.2937	0.0524	−0.0742	0.4246	0.1420	−0.0760	0.1881	0.1821	0.1118
C5.0	0.7311	0.2284	−0.0791	0.6300	0.1743	−0.0721	0.2565	0.2484	0.1120
C5.0 Boosted	0.6311	0.1929	−0.0474	0.8023	0.2834	−0.0036	0.1489	0.1657	0.1119
J4.8	0.7677	0.2354	−0.0735	0.6004	0.2250	−0.0073	0.2584	0.2493	0.1120
J4.8 Boosted	0.8430	0.2282	−0.0721	0.7844	0.1950	−0.0641	0.2574	0.2483	0.1120
RF	0.8131	0.2939	−0.0575	0.8182	0.2938	0.0047	0.1334	0.1191	0.1120
At default									
Direct OLS	0.2095	0.1251	−0.0500	0.2177	0.1794	0.0086	0.0154	0.0148	0.0151
Direct RF	0.8587	0.3180	0.0541	0.7718	0.2362	0.0369	0.0154	0.0148	0.0151
Logit	0.2281	0.0106	0.0455	0.4104	0.1729	0.0498	0.2007	0.1843	0.1116
C5.0	0.8419	0.3591	0.0476	0.8238	0.3111	0.0451	0.2757	0.2498	0.1117
C5.0 Boosted	0.8233	0.3466	0.0531	0.8217	0.3373	0.0486	0.1797	0.1394	0.1118
J4.8	0.8383	0.3633	0.0621	0.7687	0.2539	0.0479	0.2870	0.2501	0.1117
J4.8 Boosted	0.8871	0.3688	0.0666	0.8234	0.2527	0.0509	0.2779	0.2507	0.1117
RF	0.8808	0.4185	0.0732	0.8459	0.3570	0.0752	0.2014	0.1443	0.1119

Table 3.5: Coefficient of determination R^2 of the one-step and two-step models for companies D–F. The table lists the used classification methods (step (1)). Random forest (RF) regression produces the estimation of LGD (step (2)) described in Section 3.2.2. The listed coefficients are calculated according to Equation (3.2). We validate the estimates in-sample (Is), out-of-sample (Oos), and out-of-time (Oot). Direct OLS is the direct ordinary least squares regression of the LGD, and direct RF is the direct regression of the LGD with RF. Both are one-step models. Logit is the logistic regression. The tree classifiers C5.0 and J4.8 are performed in single and boosted version. In all cases, higher outcomes are preferable. We underline the best results for each testing method, both points in time, and each company.

such R^2. This result demonstrates that the available information at the contract's execution is insufficient to adequately forecast LGD. Again, the estimation results improve with the additional information at the default of the contracts.

Along with this study's two-step models, we fit two estimation models as a benchmark for LGD estimation: direct ordinary least squares (OLS) regression and direct RF regression. In most cases and most importantly in the out-of-sample and out-of-time estimation, we find that the two-step models exceed the explanatory power of direct estimation.

The in-sample results are remarkably good throughout all methods. Particularly the tree methods generate determination coefficients as high as 88%. The distinction between recovered and written off contracts seems negligible because direct RF regression yields similarly high R^2. Although R^2 in general decreases out-of-sample, the gap between R^2 of the one-step and two-step model increases significantly. This finding rewards the consideration of recovery and write-off of the contracts. The coefficient of determination in the out-of-time estimation is indeed negative in most cases at the execution of the contracts. However, accounting for the additional information at default, the two-step models can explain up to 11% of the variation of LGD. Again, the difference in R^2 between one-step and two-step models is significant.

The reason for the large deviation between direct and multi-step estimation of LGD is the control for the different default ends. Any regression missing this information is biased because recovery is a key driver of LGD as we show in Section 3.3.3. Concerning Table 3.3 it is clear that recovered and written off contracts yield significantly different LGD values and distributions. Accounting for this single information, the distributions in Figure 3.1 can already explain the variation of the respective LGD to a large extent.

Due to the large set of available information in company D and E's data, our methods perform similarly well on both. Company F's results can hardly compete with those of companies D and E in the in-sample and out-of-sample estimation. However, company F's out-of-time R^2 is surprisingly high. The out-of-time coefficient of determination for company F is higher than most of its counterparts for companies D and E. It is also positive throughout the methods at the execution of the contracts. We attribute this effect to the very different datasets in terms of contract numbers and LGD distribution. In particular, the LGD of company F is very dense around 0 with 66% of the observations lying in the small interval of $(-0.3, 0.3)$ (see Figure 3.1c). This density facilitates the forecasting of LGD in case of out-of-time estimation because here a less volatile response is particularly beneficial to the estimation accuracy. In the case of companies D and E, the disadvantage in contract numbers can be compensated by additional significant variables in the in-sample and out-of-sample estimation.

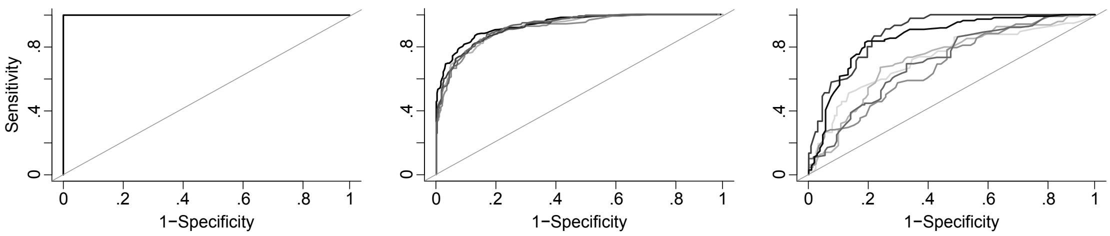
The results of the two-step model with Logit yield R^2 values that are comparable to those of Gürtler and Hibbeln (2013). Since our Logit-R^2 are only average compared to other classification models used in this study, we see a large opportunity for improvement by choosing advanced classification techniques.

Although R^2 is about equal throughout all models for company F, we recognize the large gap between the performances of one-step and two-step models. This difference is a direct result of the consideration of recovered and written off contracts. Still, RF is slightly superior in the out-of-time testing.

3.3.3 Validation and robustness

To validate our results of the classification, we use receiver operating characteristic (ROC) curves. Similar to the classification error in Section 3.3, the ROC curve displays the discriminatory capacity of the model. However, in contrast to the classification error, ROC curves preserve the classification probability of the cases without classifying according to a set threshold. Thereby, the models' tradeoff between sensitivity and specificity can be assessed. The models' discriminatory capacity then is measured as the area under the curve (AUC). For details on ROC curves see Hastie et al. (2009).

Figure 3.3 displays the ROC curves of the RF models for company D at contracts' default. The in-sample ROC curve is optimal with AUC = 1 because the classification error equals zero. Out-of-sample 5 out of 25 ROC curves of the respective RF models are chosen randomly and plotted. The individual RF models perform about equally good. They have an AUC of at least 0.9. The out-of-time AUC in Figure 3.3c is slightly smaller but is still located in an acceptable range of $(0.7, 0.9)$. We may also see, that the classification accuracy increases with the growing dataset, used for the model's training.



(a) In-sample. (b) Out-of-sample. (c) Out-of-time.

Figure 3.3: Receiver operating characteristic curves for random forest classification of company D's dataset at contracts' default. Figure b displays 5 of 25 randomly chosen models in the out-of-sample estimation. The transparency of the curves in the out-of-time estimation of Figure c reduces with increasing size of the training set and number of years (2002–2010).

The value of the models' AUCs is remarkably high, when we consider the in-sample, out-of-sample, and later (less transparent) out-of-time ROC curves. This finding proofs that forecasting the contracts' default outcome can be carried out very successfully even out-of-time, given an adequately large data history.

We argue in Sections 3.3.1 and 3.3.2 that the small set of information provided by company F is responsible for the models' poor classification and estimation performance. To proof this argument, we reduce the information sets of companies D and E to that of company F. The remaining information then is: type of contract; type of asset; asset purchase price; and EAD. As expected, for company D and E our two-step models with the reduced set of information produce a significant increase in classification error. The error rises to an average of 0.39 in out-of-sample and even 0.57 in out-of-time estimation at contracts' default for company D. That is an increase of 107% out-of-sample and 76% out-of-time compared to the full information set. We find similar results for company E. Here the classification error averages 0.39 and 0.40 respectively. Likewise, the coefficient of determination decreases drastically for these models in in-sample and out-of-sample estimation but less so in out-of-time estimation.

Still, there is a second influencing feature in company F's dataset. It yields the strong out-of-time results displayed in Table 3.5. In order to investigate this outcome, we randomly reduce the dataset of company F to 25% and 10% of the observations. Then, we run the same two-step models as before. R^2 in the in-sample and out-of-sample estimation remains as good as with the total dataset. However, we see a significant and nearly linear decrease in the out-of-time results. Here two-step RF yields a coefficient of determination of 0.09 on 25% and 0.07 on 10% of the original dataset. Hence, R^2 decreases by 20% and 37% respectively. The original dataset of company F has about ten to twenty times the number of observations per year compared to companies D and E. We conclude that this advantage in size leads to robust R^2 in the out-of-time estimation. It also reduces the dependence on concrete estimation models. The in-sample and out-of-sample R^2 on company F's dataset could be comparable to those of the other companies if it provided additional information.

Furthermore, we measure the importance of recovery in the RF regression models of step (2) to emphasize the influence of realized, as well as estimated recovery. According to Liaw and Wiener (2012), this importance is the total decrease in node impurities from splitting on the variable, averaged over all trees. The node impurity is measured by the Gini index.

Table 3.6 presents the five most important variables in the RF models of company D's dataset, which originally use 53 variables. As indicated before and serving as a benchmark, the realized recovery RC is most important in the LGD estimation. In Section 3.3.1 RF predicts the recovery of defaulted contracts very successfully in in-sample and out-of-sample testing. This outcome corresponds to the importance of the in-sample and out-of-sample predicted recoveries RC^*. The importance of the out-of-time predicted recovery is significantly lower than its counterpart of the other testing methods. We attribute the reduction in importance to the less efficient out-of-time classification in step (1) of the analysis. This inefficiency leads to false recovery assumptions and, thereby, reduces the intensity of the link between LGD and out-of-time RC^*.

Realized		In-sample		Out-of-sample		Out-of-time	
Var.	INI	Var.	INI	Var.	INI	Var.	INI
RC	78.68	RC*	79.42	RC*	65.80	RC*	7.15
RAT	24.76	RAT	24.73	RAT	25.47	RAT	30.16
RCA	22.60	RCA	23.10	EAD	21.39	DRE	26.02
EAD	21.60	TTM	21.20	RCA	21.33	AVE	22.61
TTM	21.59	EAD	21.08	EDB	20.12	EAD	21.25

Table 3.6: Variable importance in random forest (RF) for company D's contracts at their default. The variables' importance is calculated according to Liaw and Wiener (2012). We truncate the results according to the five highest values including the realized recovery (RC) or estimated recovery (RC*) of the contracts. Var. is the variable name. INI is the increased node impurity. RF estimates the recovery in-sample, out-of-sample, and out-of-time. The remaining variables are: internal rating (RAT); relative contract age at default (RCA); exposure at default (EAD); exposure at default per unit of the assessment base (EDB); time to maturity at default (TTM); default reason (DRE); and asset value per unit of the exposure at default (AVE).

Finally, to determine the optimal number of trees for the RF model, we calculate the out-of-sample classification error for RF models with 1 to 10,000 trees. Figure 3.4 shows that the classification error stabilizes toward larger forests and the error reduces by almost 50% compared to a single tree forest. Hence, as a precaution, we choose to train forests comprising 2,000 trees. These forests should produce robust results on an adequate level of the classification error.

3.4 Conclusion

The default of a debtor often entails monetary losses for the lender. The loss amount may be reduced by prepayments, various collateral, or concerning lease contracts by residual value, and reselling capability of the leased asset. However, the realized loss after the debtor's default may be affected by the lender's and the debtor's actions during the workout process. Other circumstances may be beyond their control. The actions taken lead to recovery, in case continuation of defaulted contracts is considered in principal, or non-recovery of the debtor. The latter would mean the contract's write-off.

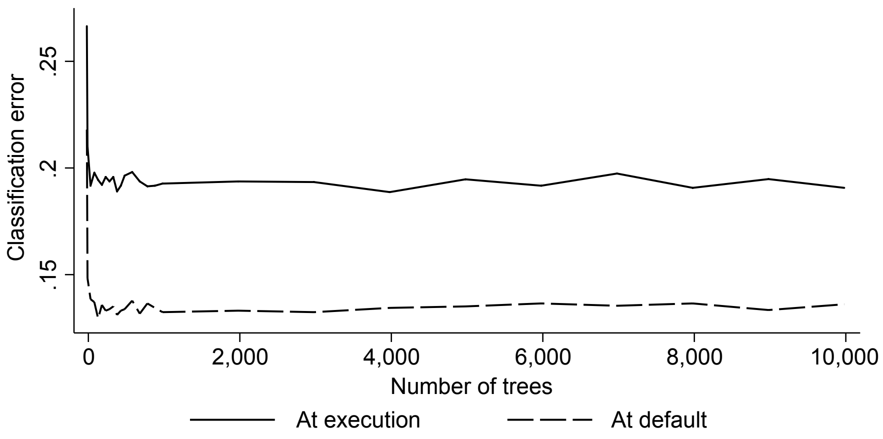


Figure 3.4: Out-of-sample classification error for random forest classification of company D's recovered and written off contracts as a function of the forest's size. The error is calculated according to Section 3.3 with the available information at the contract's execution or default.

It should be in the lender's best interest to keep the LGD of defaulted contracts as low as possible for several reasons. A low LGD means a high return of the EAD and thereby reduces the investment risk. Consequently, a lower risk premium can be applied to customers or particular contracts with low LGD. Moreover, contracts with low LGD require less equity to absorb unexpected losses.

In this study, we analyze the effect of recovery of defaulted contracts on the LGD. We find recovery is that piece of information with the highest explanatory power in LGD variation. Moreover, we confirm this finding by using datasets with different levels of LGD. Our estimation results show that the two-step models' improvement in explaining the variation of LGD is remarkably high compared to direct OLS and RF regression. Overall, Logit is mostly outperformed, J4.8 and C5.0 perform comparable, but seldom reach the accuracy of RF. Accounting for the contract's recovery causes the most difference in estimation accuracy. Other factors also have an influence, particularly when testing out-of-sample and out-of-time. These are information availability, number of observations, and LGD distribution characteristics.

Analyzing three datasets of defaulted leasing contracts, we find that the lessor benefits from the recovery of defaulted contracts in two ways. On the one hand, recovered contracts on average yield an LGD which is significantly lower than the LGD of written off contracts, granting the lender all advantages of low LGDs mentioned above. On the other hand accounting for the recovery of contracts is particularly beneficial to the estimation accuracy when forecasting the LGD of executed or defaulted contracts. This higher accuracy results in risk adequate pricing of the lender's services and consequently strengthens his competitiveness.

For companies, the key to successfully continue and recover defaulted contracts is to identify those contracts that are worth being continued, primarily in terms of reduced LGD. Therefore, developing expertise in evaluating the different default outcome values is very advantageous to the lessor. This expertise would support the lessor's decision between write-off and continuation of a defaulted contract. Our findings indicate that verifying the continuation worthiness of these contracts may particularly improve their LGD.

4 Loss given default-adjusted workout processes for leases

Credit risk modeling is an essential assignment of risk management in financial institutes. One of the major drivers of credit risk is the loss given default (LGD). The knowledge of potential losses is crucial for an efficient allocation of regulatory and economic capital and also for credit risk pricing. According to Article 107 (1) of the capital requirement regulation (CRR), financial institutions shall apply either the Standardised Approach or the Internal Ratings Based Approach (IRBA), in order to calculate their regulatory capital requirements for credit risk. When implementing the advanced IRBA, it is mandatory to develop internal models for estimating the probability of default (PD), exposure at default (EAD), and LGD. One of the main objectives of the IRBA is to achieve risk-adjusted capital requirements (see Basel Committee on Banking Supervision (2003)). Accurate forecasts of PD, EAD, and LGD may result in competitive advantages for the applying financial institution, in general, such as Gürtler and Hibbeln (2013) indicate.

While the procedure of calculating the PD might be almost identical for loans and leases, models for estimating the LGD should consider specific characteristics of leasing contracts. In contrast to loans, the collateralization of a lease by its leased asset is obligatory. Particularly, being the legal owner of the leased asset, the lessor can retain any recovered value of the leased asset's disposal. Thus, contrary to loans, the lessor has legal access to this additional source of payments, in case a contract defaults. Eisfeldt and Rampini (2009) argue that the main benefit of leasing is that repossession of a leased asset is easier than foreclosure on the collateral of a secured loan. During the workout process of a defaulted loan, the lender exclusively receives payments from the debtor and the liquidation of collateral. These incomes also occur during the workout process of leases. Consequently, considering the additional incomes from disposing of the leased asset, the cash flows of the leasing workout process consist of two parts. Han and Jang (2013), Töws (2014), and Frontczak and Rostek (2015) argue that the level of LGD critically depends on the actions taken during the workout process. Hence, the peculiarities of the workout process of leases should be taken into account when estimating LGD.

Although, various advanced approaches for estimating LGD have been analyzed, as yet, no single approach could be established, neither for loans nor leases. Nevertheless, Bastos (2010), Hartmann-Wendels et al. (2014), and Yao et al. (2015) find that complex models can generate robust and precise estimations. Moreover, Qi and Zhao (2011) and Loterman et al. (2012) argue that the consideration of nonlinear effects is important when estimating LGD. However, a prerequisite for the good performance of such models, is a correspondingly large database, both in terms of observations and associated information.

The results of Qi and Zhao (2011) and Hartmann-Wendels et al. (2014) indicate that overfitting is a common concern of complex models, which may negatively affect forecasting accuracy. Presumably due to the lack of data and issues with controlling overfitting, so far, linear regression is the most frequently used method for estimating LGD in the literature. Nevertheless, when regarding the peculiarities of the LGD distribution, linear regression seems to be at least econometrically inappropriate for the estimation task. Typically the workout LGD of loans and leases is bimodally or even multimodally distributed (compare Laurent and Schmit (2005), Zhang and Thomas (2012), Hartmann-Wendels et al. (2014), and Li et al. (2014)). This unusual shape of the density supports the hypothesis that LGD estimation requires the use of advanced methods. In order to produce accurate and comprehensible estimations, these methods should be able to approximate the complex relationships between the available information and the LGD as precisely as possible.

Against this theoretical and practical background, a number of different methods have been investigated in the literature. In particular, these studies examine the models' suitability and predictive accuracy to LGD estimation.

Several studies focus on reproducing the LGD's density function in order to extrapolate accurate estimations in this manner. For this purpose, Calabrese and Zenga (2010) use a mixed random variable to model LGD on the unit interval. They employ their concept to a large set of defaulted Italian loans. Altman and Kalotay (2014) pursue a similar approach based on the mixture of Gaussian distributions. They report successful estimations using Moody's Ultimate Recovery Database (MURD). Hartmann-Wendels et al. (2014) also apply an approach based on finite mixture models to estimate the LGD of leases. However, out-of-sample, their approach performs poorly. The authors conclude that reproducing the LGD density is only of secondary importance to the estimation accuracy.

Further studies examine the suitability of various parametric and nonparametric methods for LGD estimation. Applying several regression techniques to the data of six different banks, Loterman et al. (2012) conclude that nonlinear methods perform better than linear methods. Qi and Zhao (2011) obtain a similar result. They compare different parametric and nonparametric methods using MURD. The authors argue that nonparametric methods can generate more accurate LGD estimations due to their ability to model nonlinear relationships between the LGD and continuous explanatory variables. In particular, they find regression trees to be a suitable nonparametric method for estimating LGD. Bastos (2010) obtains a similar outcome when he uses regression trees on Portuguese bank loans. Likewise, Hartmann-Wendels et al. (2014) successfully apply model trees to estimate the LGD of German leases.

Recently, a couple of studies applied ensemble learning techniques to estimate LGD. These are an extension of the analysis of single procedures. Bastos (2013) improves the estimation accuracy significantly by using regression trees in an ensemble approach on MURD. On a set of leases, Töws (2014) finds that random forests achieve higher coefficients of determination than linear regression.

In addition to single-stage models, some studies implement two-stage models to forecast LGD. Typically, these models split the observations ex-ante according to a specific key feature. To predict the LGD of mortgage loans, Leow and Mues (2012) first estimate the probability of mortgage accounts undergoing repossession. Then, they subsequently calculate the loss in case of repossession using a certain haircut value. The latter is the ratio of the forced sale price and the valuation of the repossessed property. Concerning the LGD of leases, Töws (2014) successfully introduces a two-stage approach. He distinguishes between recovered and written off contracts and then estimates the respective LGD.

While several studies show that complex models can generate more accurate LGD estimations than linear regression, some works demonstrate the practical suitability of the latter. Zhang and Thomas (2012) apply linear regression and survival analysis to a dataset of defaulted personal loans from the UK. They find that linear regression generates the best LGD estimates in general and outperforms more advanced estimation techniques. Bellotti and Crook (2012) obtain a similar result when estimating the LGD of UK credit cards.

So far, all introduced studies have in common that they regard LGD as a holistic measure of risk. Concerning the LGD of loans, such an approach is reasonable. However, according to the specific characteristics of leasing contracts, the LGD of leases typically consists of cash flows from two distinct sources. Thus, a holistic approach to estimate the LGD of leases might be inappropriate.

Therefore, we present an entirely new approach to forecast leasing LGDs. In our study, we consider the specific characteristic of leases and, consequently, we suggest an economically motivated separation of the LGD into an asset-related and a miscellaneous share. Coming from different payment sources, both shares should be driven by different factors. Particularly, the loan-to-value ratio should be most important to the asset-related share, but of less importance to the miscellaneous share.

In the course of this paper, we describe the development of a multi-step estimation model, which is built upon the economic composition of the LGD of leasing contracts. Estimating the asset-related and miscellaneous share, we derive an estimation of the overall LGD. Our easily traceable model results in a significant advantage in terms of estimation accuracy.

Moreover, the estimated asset-related and miscellaneous LGD can be used to support decisions concerning the accomplishment of the workout process. In fact, the separation of LGD entails extensive practical implications for handling a defaulted contract's workout process. The derived shares of LGD are indicators for the success of both the asset's disposal and the effort of collecting further payments. Consequently, we find that our inferred suggestions for the actions to be taken by the lessor during the workout process lead to significant improvements in the resulting LGD value of the respective contracts.

For our study, we use a real life dataset provided by a major German lessor. The data features a high quality of details which is particularly important to our approach. We compare the performance of our procedure to that of traditional holistic methods for LGD estimation, e. g. carried out by ordinary least squares (OLS) regression. In particular, to measure the accuracy and robustness of the models, we use in-sample, out-of-sample, and out-of-time validation. Moreover, considering the economical context and the obtained estimation errors, we discuss theoretical and practical advantages and disadvantages of each step of our approach.

4.1 Dataset

The dataset consists of 1,493 defaulted leasing contracts with 907 lessees from a large German leasing company. The contracts were executed between 1996 and 2009. Their default occurred between 2002 and 2009. The default status of any contract was determined by the default events outlined in Article 452 of Basel II. These events correspond to Article 178 (1) of the CRR. The contracts default after an average of 50% of their maturity. That is approximately 2.5 years after the execution of the average contract. The mean workout lasts about 2 years. The workout of all contracts has been completed. The last of which was completed in 2010. Further data has not been provided.

Our data is extremely valuable with respect to its high level of detail, particularly regarding the workout process. Similar to Hartmann-Wendels et al. (2014), a large amount of information is available. These are contract, leased asset, and customer specific information as well as additional information about the contract's default and its workout process. The breakdown of cash inflows and outflows during the workout process is of particular importance to the derivation and economic interpretation of the approach we present in this study. The carefully documented costs concerning the disposal of the leased asset and the collection of overdue payments, allow for the precise and economically sensible separation of asset-related and miscellaneous revenues.

Before any separation or estimation of the LGD, we briefly discuss its calculation. The LGD is that portion of EAD that could not have been recovered in case of a contract's default. Its counterpart is the recovery rate (RR). The workout RR is the ratio of the amount recovered and EAD, it is equivalent to $1 - \text{LGD}$. In line with Article 5 (2) CRR, we use the term related refinancing interest rate to discount all incurred cash flows (CF) and workout costs (WC) to the time of default. The EAD is the present value of the defaulted contract's outstanding exposure, calculated as the sum of outstanding payments at the time of default.

The detailed breakdown of incoming and outgoing cash flows during each contract's workout enables us to determine LGD very precisely. Formally, we calculate the LGD as

$$\mathrm{LGD} = 1 - \frac{\mathrm{CF} - \mathrm{WC}}{\mathrm{EAD}} = 1 - \mathrm{RR}. \tag{4.1}$$

Beyond the pure determination of LGD, we calculate component LGDs. The asset-related LGD (ALGD) summarizes all asset-related payments, such as the asset's liquidation proceeds and incurred liquidation costs. We call the remaining share of the LGD miscellaneous LGD (MLGD). The MLGD comprises revenues from capital services, such as interest rates and customer payments, the costs of collateral, such as recovery costs and maintenance costs, and proceeds of collateral, other indirect costs, and other payments. Both component LGDs refer to the overall EAD. However, they particularly differ in terms of the lessors influence on the respective cash flows. While repossession of the leased asset, as well as its disposal, is in the responsibility of the lessor completely, miscellaneous cash flows depend on several factors outside his control. For instance, in case the defaulted lessee goes through an insolvency proceeding, the insolvency estate is distributed pro rata between all relevant creditors. Basically, the MLGD of a leasing contract is the equivalent of a loan's LGD.

We derive the two component LGDs from Equation (4.1) by identifying the asset proceeds (AP) within the incoming cash flows and the related asset liquidation costs (LC) within the workout costs. This splitting results in

$$\begin{aligned} \mathrm{LGD} &= 1 - \frac{(\mathrm{CF_M} + \mathrm{AP}) - (\mathrm{WC_M} + \mathrm{LC})}{\mathrm{EAD}} \\ &= 1 - \frac{\mathrm{AP} - \mathrm{LC}}{\mathrm{EAD}} - \frac{\mathrm{CF_M} - \mathrm{WC_M}}{\mathrm{EAD}} \\ &= 1 - \mathrm{ARR} - \mathrm{MRR}, \end{aligned} \tag{4.2}$$

with $\mathrm{CF_M}$ and $\mathrm{WC_M}$ denoting the remaining miscellaneous incoming cash flows and workout costs respectively.

Subsequently, we derive the asset-related RR (ARR) and the miscellaneous RR (MRR). As usual, we obtain the LGD as the counterpart of the RR

$$\text{ALGD} = 1 - \text{ARR}, \qquad \text{MLGD} = 1 - \text{MRR}. \tag{4.3}$$

In terms of ALGD and MLGD, the LGD then is calculated as

$$\text{LGD} = \text{ALGD} + \text{MLGD} - 1. \tag{4.4}$$

Descriptive statistics

In contrast to various studies, we do not restrict LGD to the unit interval, such as Chalupka and Kopecsni (2009), Bastos (2010), and Zhang and Thomas (2012) do. For leases, LGDs outside the unit interval are frequently observed. Hartmann-Wendels and Honal (2010) argue that LGDs less than 0 may occur in cases where the asset disposal covers more than the amount of EAD. Additionally, incorporating workout costs may cause the LGD to rise beyond 1. Table 4.1 provides a brief overview of the LGDs' distribution parameters. The overall LGD averages near 35%, and we observe an average ALGD of 69% and MLGD of 65%. The standard deviation of ALGD is notably lower than that of MLGD and LGD. Minimum and maximum of ALGD and MLGD are consequently higher than those of the LGD.

We find that the ratio of asset value at default to EAD is 54% on average. Although, the lower quartile of ALGD is quite high, for more than 10% of the contracts the asset value even exceeds EAD. For these contracts, ARR is higher than 1. While the default value of the leased asset is not an explicit part of the EAD, this value qualifies for incoming cash flow during the workout process in case of asset disposal. Such as any other cash income, the disposed asset value reduces the LGD. Moreover, in contrast to the liquidation of a loan's collateral, as

Share of LGD	Mean	Std	Min.	P25	Median	P75	Max.
ALGD	0.69	0.41	−0.99	0.41	0.98	1.00	2.03
MLGD	0.65	0.51	−1.04	0.22	0.89	1.01	2.68
LGD	0.35	0.48	−1.36	0.00	0.30	0.76	1.50

Table 4.1: Distribution parameters of the loss given default (LGD). Std is the standard deviation, Min. is the minimum, and Max. is the maximum LGD value. P25 and P75 are the respective quartiles. ALGD is the asset-related LGD and MLGD is the miscellaneous LGD. We derive both component LGDs from Equation (4.3).

the legal owner of the leased asset the lessor can keep any surpluses from disposing of the leased asset even if the resulting ARR exceeds 1.

For a lessor's internal risk management, determination of ALGD is useful. If interpreted as a stand-alone parameter, ALGD is theoretically an upper limit to the LGD. This is true if the MLGD does not exceed a value of 1, which implies the success of the workout process. Therefore, depending on the amount of ALGD, the lessor can determine whether the asset sales proceeds already cover the EAD or if further workout actions should be taken to collect overdue payments.

Frontczak and Rostek (2015) argue that knowledge about the effect of disposal efficiency and related costs on the LGD may affect a lender's disposal policy. Consequently, for the lessor it would be useful to know ex-ante if the MLGD will exceed 1. In case it does, the lender loses more than the full amount of EAD. Strictly speaking, MLGDs > 1 indicate that the incurred collection costs will exceed the payments collected. In such cases, even if the asset sales proceeds cover only a small portion of EAD, the workout should be restricted to the disposal of the leased asset because collecting further payments is inefficient from an economic standpoint.

Theoretically, it is also possible that the ALGD exceeds 1. Nevertheless, in our data we find that asset sales proceeds exceed the incurred disposal costs in 99% of all cases. This outcome could have been expected because leasing companies are experts in disposing of their leased assets. Hence, the disposal is economically

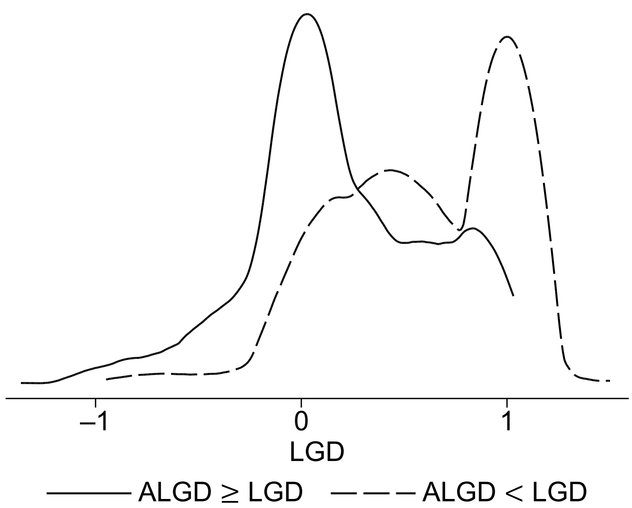


Figure 4.1: Densities of loss given default (LGD), after separating the contracts according to their relationship of LGD to asset-related LGD (ALGD). The full amount of the exposure at default (EAD) is recovered in case of 0. -1 defines an EAD recovery of 200% while an LGD value of 1 means the loss of 100% of the EAD.

reasonable in almost any case. Interestingly, for about 35% of the examined contracts, the MLGD exceeds 1. This implies that the ALGD as an upper limit of the LGD holds for only about 65% in practice. Nevertheless, as can be seen in Figure 4.1, this upper limit is an important feature to distinguish the contracts. More precisely, categorizing the contracts according to this upper limit leads to LGD distributions that are disjointed to a large extent. The realized LGDs of the contracts that satisfy ALGD $\geq$ LGD concentrate around 0 with a mean of 0.20. In contrast, for contracts with ALGD exceeding the LGD, LGDs are particularly located around 0.5 and 1 with a mean of 0.59.

Figure 4.2 visualizes the density of the calculated overall LGD of the underlying dataset. In addition, ALGD and MLGD densities are plotted. Both overall LGD and MLGD exhibit a pronounced bimodal shape, with concentrations around an LGD level of 0 and 1. The LGD's mean of 35% in Table 4.1 indicates that the overall LGD is rather small in most cases. Its median of 30% confirms this finding. ALGD and MLGD, however, have more density around 1. From the perspective of a regular lender, a high ALGD is of less concern than a high MLGD. Because the asset's fair value is not included in the EAD, the cash inflow from

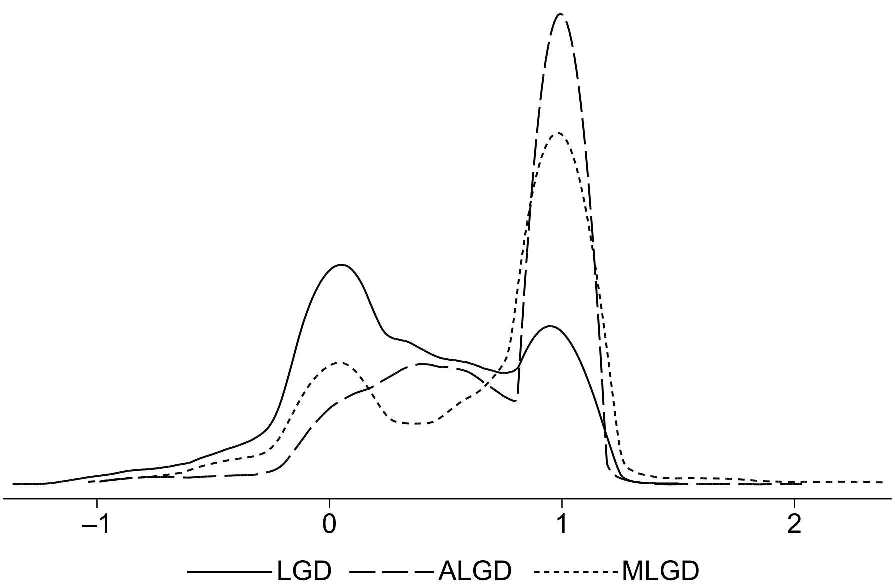


Figure 4.2: Density of loss given default (LGD), asset-related LGD (ALGD), and miscellaneous LGD (MLGD). The full amount of the exposure at default (EAD) is lost in case of 1. -1 defines an EAD recovery of 200% while an LGD value of 2 means the loss of 200% of the EAD.

the asset's disposal has an unexpected reducing effect on the LGD. In contrast, the cash inflows considered by MLGD are fully accounted for in the EAD. A high MLGD reflects a poor outcome from the workout process. However, the ALGD is important to lessors because revenues from disposing of the leased asset in case of default are a substantial aspect of a lessor's business model.

To be precise, the average revenue from disposal of the leased asset amounts to €15,322 per contract. The miscellaneous payments during the workout process sum up to €13,616 on average. This allocation of cash inflows emphasizes the importance of both sources of revenues for a leasing company. It confirms that the workout process of defaulted leases is quite different from that of loans. Consequently, for leasing contracts it is essential to consider both ALGD and MLGD when estimating the overall LGD. Moreover, the share of revenues from disposing of the leased asset is indeed slightly higher on average than the remaining share. However, in particular for less valuable assets, the traditional payments collection during the workout process is substantial.

	LGD		ALGD		MLGD	
Year	Mean	Std	Mean	Std	Mean	Std
2002	0.3723	0.5283	0.7386	0.4076	0.6337	0.5652
2003	0.2995	0.4597	0.6978	0.4360	0.6017	0.5344
2004	0.3068	0.4341	0.6637	0.4267	0.6431	0.5302
2005	0.3245	0.4516	0.6640	0.4505	0.6604	0.4603
2006	0.3218	0.4557	0.6506	0.4218	0.6712	0.4750
2007	0.3303	0.4225	0.6503	0.4221	0.6799	0.4555
2008	0.3413	0.4218	0.6167	0.3968	0.7247	0.4062
2009	0.3999	0.4179	0.6654	0.2914	0.7345	0.3826

Table 4.2: Distribution parameters of the loss given default (LGD) for each default year. Std is the standard deviation. ALGD is the asset-related LGD and MLGD is the miscellaneous LGD. We derive both LGDs from Equation (4.3).

In Table 4.1 we observe higher standard deviations of MLGD and LGD compared to ALGD. Thus, the latter is less volatile, Miller (2015) notes similar. Therefore, ALGD might be easier to estimate. In addition, Table 4.2 displays the key figures of the realized LGD, ALGD, and MLGD values over the default years of the observation period. At this level of aggregation, ALGD is still less volatile than MLGD and LGD in a year by year comparison. Concerning LGD, we observe rather small fluctuations during the period of 2003 to 2008. Only for the years 2002 and 2009, the realized LGD is in comparison noticeably higher on average. Particularly, the higher average LGD in 2009 might be a result of the global financial crisis. Regarding ALGD, the means fluctuate around 65%. However, we find it interesting that ALGD does not increase unusually in 2009. This finding indicates that the fluctuations of ALGD are driven by each year's asset disposals but are not driven by the economy. Apparently, ALGD does not increase during the financial crisis. We attribute this effect to the lessor's excellent knowledge of secondary markets. Obviously, there is a difference in the course of MLGD. It also fluctuates only rarely between the years 2002 and 2007. However, we observe a markedly but manageable increase in 2008 and 2009, which might be a result of the financial crisis.

The evolution of the three LGD ratios supports our hypothesis that ALGD might be easier to estimate for the lessor than MLGD or LGD. However, we find no empirical evidence, that the economy, e. g. accounting for gross domestic product and unemployment rate, has an impact on ALGD. The economy might influence MLGD and LGD slightly. Nevertheless, the potential effect seems to be minor. Moreover, Miller (2015) shows that the LGD estimation at the default of a lease benefits only slightly if at all, of considering the economic situation. Consequently, we do not include macroeconomic factors into our approach. In fact, for the estimation of the LGD ratios we focus on contract related factors, such as the type of the leased asset, of the customer, and the default reason.

4.2 Methods

Contrary to recent studies on LGD estimation, we do not focus on the comparison of very complex or even black box methods, such as support vector machines or neural networks. We rather develop an economically based and consistent technique for estimating LGDs. Instead of regarding LGD as a holistic measure of risk, we separate the LGD into an asset-related and a miscellaneous share and, hence, take into account the specific characteristics of leases. In order to provide evidence that the increase in estimation accuracy does not solely arise from particularly suited methods but sophisticated economic consideration, we essentially apply two distinct methods to our proposed multi-step approach. These are OLS and as an advanced estimation method, the tree algorithm RF. Throughout the study, we set the traditional direct estimation of LGD by OLS and RF as a benchmark to compare the performance of our multi-step estimation model and to measure the improvement.

As OLS is a common estimation method, we will only give a brief overview of the RF model. The RF tree algorithm was constructed by Breiman (2001). It has many similarities to regular regression and classification trees. These trees subsequently divide the initial dataset according to a series of if-then conditions. At every node of the tree, the best split is performed according to an appropriate split criterion, e. g., the greatest expected reduction in standard deviation. Each contract terminates in one leave of the final tree. Each leave's estimation value then is the average value of the contracts of the respective leave. In case of classification, the contracts' realized class in each leave determines the leave's class estimation.

RF differs from regular regression trees in three important ways. First, instead of building only one tree, a series of trees, thus, a forest is built. Second, each tree is calibrated with a random sample of the dataset. Third, at each node the available set of splitting variables is a random sample of all available variables. The final estimation of a contract is the average of the single tree estimations. For classification, the majority vote determines a contract's class. We use the RF standard parameters suggested by Breiman (2001). For classification, these are $\sqrt{m}$ randomly chosen variables for each split and $^{m}/_{3}$ variables for regression out of a total of m variables. The size of the forest is fixed to 1,000 trees, as proposed by Hastie et al. (2009).

Beside the frequently used OLS, numerous studies have shown that tree-based algorithms are particularly suited for estimating LGD. While Bastos (2010) and Hartmann-Wendels et al. (2014) find that regression and model trees generate robust and accurate LGD estimations, Töws (2014) reports similar outcomes for RFs explicitly.

4.2.1 Direct estimation

To begin with, we take a look at direct estimation methods. Direct estimation is easy to implement and, therefore, the most elementary and common method for estimating LGD. In this study, direct estimation by OLS and RF serves as a useful

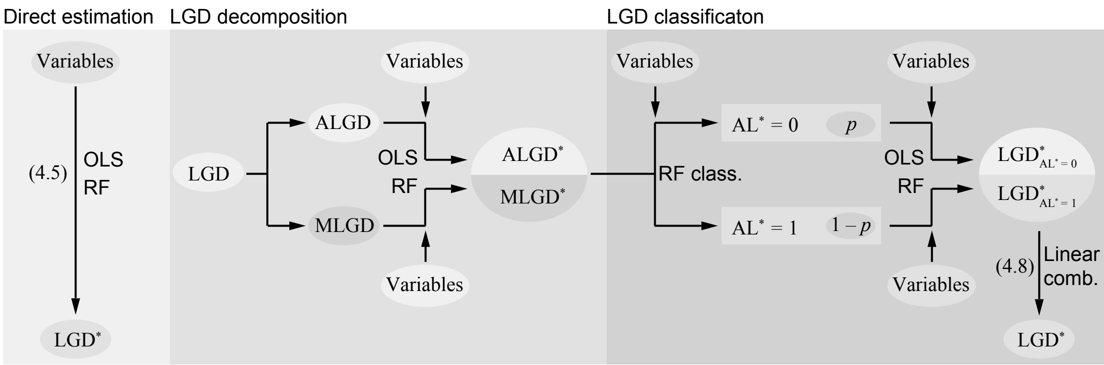


Figure 4.3: Procedure of the developed models. Our approach consists of three consecutive parts. Direct estimation determines the loss given default (LGD*) according to Equation (4.5) using the variables available, both in an ordinary least squares (OLS) and in a random forest (RF) regression model. In the LGD decomposition, we divide the realized LGD into an asset-related LGD (ALGD) and a miscellaneous LGD (MLGD). Then again, using the available variables, two OLS or RF models are calibrated to estimate ALGD* and MLGD*. Subsequently, the contracts of our dataset are classified into two classes. An RF classification model uses the available variables including ALGD and MLGD and their estimated values to perform the classification of Equation (4.7). It aims to assign AL$^* = 0$ correctly to contracts with an ALGD exceeding its LGD, and AL$^* = 1$ in case ALGD falls short of LGD. Based on these two disjoint datasets, we calibrate an OLS or RF model on each to estimate the two $\text{LGD}^*_{\text{AL}^*}$. Using the linear combination of Equation (4.8), we calculate the final LGD estimation by weighting $\text{LGD}^*_{\text{AL}^*}$ with their classification probabilities p and $1-p$.

benchmark when we compare it to the respective multi-step model by measuring the models' performance improvements. The left-hand side of Figure 4.3 visualizes a simplified direct estimation method. The method uses a set of variables to produce estimations of the LGD.

Using OLS, we model the LGD dependent on the available and relevant variables (VAR) in a linear combination

$$\text{LGD} = \alpha + \sum_{i=1}^{m} \beta_i \cdot \text{VAR}_i + \varepsilon, \tag{4.5}$$

with α the regression's constant, β_i the slope coefficient of variable i, ε the residual, and m the number of included variables. For RF regression, we train a forest based on the same information set, estimating the dependent variable directly.

The advantage of direct estimation is the plain analysis of the influence of the independent variables. In case of OLS, significance and slope of single influencing factors are rather easy to measure and have economic interpretations. The importance measure of RF allows for similar conclusions. However, from a methodological perspective, OLS comes with a major disadvantage. It only models linear relationships between dependent and independent variables. This way, many latent influences and changes of influences according to independent variable values cannot be considered. Although, OLS has been successfully used for estimating LGD, e. g. by Bellotti and Crook (2012) and Zhang and Thomas (2012), theoretically, RF should be much more suited to the estimation task. The latter can particularly consider nonlinear dependencies between the LGD and its explanatory variables by generating homogeneous subsets of the data. Still, both methods can only process the plain information available.

4.2.2 Loss given default decomposition

From an economic point of view, the LGD is a linear combination of cash flows relative to EAD. With leasing contracts, this relationship plays a particularly important role because, unlike with loans, the cash flows are typically issued from very different sources. Observing the cash flows in detail, we attempt to provide additional information to the estimation of LGD by breaking down the LGD to ALGD and MLGD. Equations (4.2) and (4.3) provide the necessary mathematical steps of this calculation. Figure 4.3 outlines the procedure of the LGD decomposition. Similar to LGD, neither ALGD nor MLGD are available at the time of contract's default. Therefore, the idea is, instead of estimating LGD directly, we estimate ALGD and MLGD and combine these parameters to a new LGD estimation. Again, for estimating ALGD and MLGD, we apply OLS and RF models. In principal, any other method can be utilized.

From a mathematical and economic perspective, the separate estimation is reasonable in three ways. First, obviously, asset-related cash flows depend on different influencing factors than miscellaneous cash flows. For instance, we expect the loan-to-value ratio to be a significant driver of ALGD but not of MLGD. Second, according to the different density shapes outlined in Figure 4.2, the estimation of the two LGDs might vary in its accuracy. In particular, the markedly lower standard deviation of ALGD compared to MLGD highlighted in Table 4.1, indicates that the estimations of ALGD might be significantly more precise. Third, both of these estimated components of the LGD provide decision support concerning the actions that should be taken during the workout process in order to achieve LGDs as low as possible. The last argument is particularly important from an economic point of view.

The gain of information by estimating ALGD and MLGD may be used in different ways to enhance the accuracy of LGD estimation. Theoretically, the LGD can be calculated reversely by using Equation (4.4)

$$\mathrm{LGD}^* = \alpha \cdot \mathrm{ALGD}^* + \beta \cdot \mathrm{MLGD}^* - \varepsilon, \tag{4.6}$$

where ALGD^* and MLGD^* are the estimated ALGD and MLGD. α and β are slope coefficients and ε is the constant. In the theoretical calculation, these three parameters are set to 1. However, for practical usage it might be suitable to set up an OLS regression to find the optimal values for these parameters. Nevertheless, a large disadvantage of this procedure is, that the full estimation error of both estimated ALGD and MLGD enters the estimated LGD. Consequently, we do not pursue this approach any further.

4.2.3 Loss given default classification

Instead of deriving an LGD estimation from ALGD^* and MLGD^*, we use the estimated values to classify the contracts into two classes. In Section 4.1, we show that the ALGD is a theoretical upper boundary to the LGD. By generating a dummy variable

$$\mathrm{AL} = \begin{cases} 0 & \text{if } \mathrm{ALGD} \geq \mathrm{LGD} \\ 1 & \text{if } \mathrm{ALGD} < \mathrm{LGD}, \end{cases} \tag{4.7}$$

we identify contracts, which realize an LGD exceeding their ALGD. According to Figure 4.1 this categorization leads to a largely disjointed separation of the contracts in terms of the LGD distributions. Moreover, the two resulting distributions of the LGD feature less distinctive bimodal shapes than the LGD distribution of all contracts, illustrated in Figure 4.2. Consequently, we expect that estimating LGD separately in each class is easier than estimating LGD without this separation. On this account, we calibrate an RF classification model with AL the dependent variable to predict whether a contract's LGD is expected to be below or above its ALGD. This model uses the relevant and available information at contract's default. Expanding this information set, we additionally use ALGD and MLGD determined according to Equation (4.3) to calibrate the classification model. For predictive classification, we consequently use the respective estimates of ALGD and MLGD from Section 4.2.2, such as is indicated by the right-hand side of Figure 4.3.

Theoretically, it is possible to classify the contracts directly by only using the estimates of ALGD and MLGD. However, in this case, the estimation error of these estimates would directly impact the classification accuracy negatively. Therefore, we do not rely on these two ratios but rather calibrate a classification model using a set of information.

For each contract, we obtain the classification probability p of the respective contract being in class 0, and its estimated class AL^*. Based on the contracts of these two classes, we calibrate two separate LGD regression models. In the estimation step, every contract receives exactly two LGD estimations, one from each of the two models calibrated. Finally, we calculate the estimated LGD in a linear combination

$$\mathrm{LGD}^* = p \cdot \mathrm{LGD}^*_{\mathrm{AL}^*=0} + (1-p) \cdot \mathrm{LGD}^*_{\mathrm{AL}^*=1}, \tag{4.8}$$

using the classification probability p to weight the single LGD estimates.

The additional classification step enriches the overall LGD estimation by interpreting ALGD as an upper limit to the LGD. Economically, the classification of a contract indicates, which actions the lessor should take during its workout process. In particular, if the LGD is likely to exceed its ALGD, the lessor should consider restricting the workout process to the disposal of the leased asset. Because, in this case, the miscellaneous workout costs are expected to exceed the miscellaneous cash inflows. Considering that MLGD > 1 for about 35% of the contracts of the studied lessor, the proper implementation of the workout process such as we suggest, would lower its mean realized LGD by nearly 10% to 0.32. This reduction of the LGD would lead to lower losses to the leasing company of about €2,250,000.

4.2.4 Validation techniques

In order to validate the estimation accuracy and to verify the robustness of our models, we use three fundamentally different validation techniques outlined in Figure 4.4. Beside common in-sample and out-of-sample validation, we also use out-of-time validation. The last simulates an estimation scenario that is as close to reality as possible. In the course of the study, we are estimating different

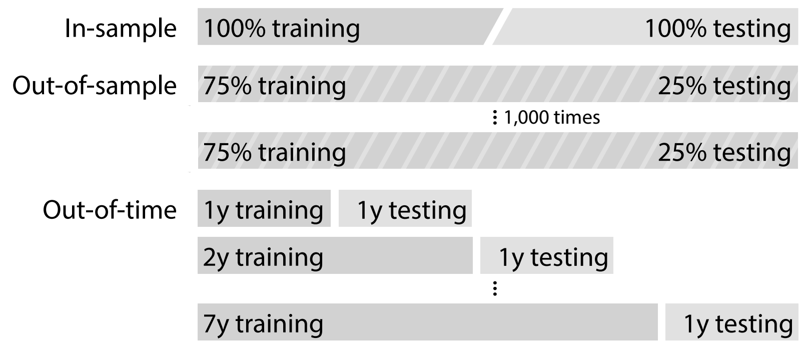


Figure 4.4: Validation techniques. Each validation approach divides the total dataset into $x\%$ training set to calibrate the estimation model, and $(1-x)\%$ validation set. Out-of-sample, we divide the data randomly 1,000 times. In case of out-of-time validation, the data is divided by the contracts' year of default. The first model then is calibrated on contracts that default in the first year and validated on contracts that default in the following year.

parameters, such as LGD, ALGD, and MLGD. Furthermore, we perform a classification to predict whether the ALGD is greater or less than the LGD. Since the following validation techniques apply to all of these parameters, we will use a uniform synonym and call them dependent variable.

For the in-sample model calibration, all observations and available information at the time of contracts' default are used. The estimation of the dependent variable then is carried out on the same data. Consequently, the estimation accuracy is expected to be relatively high. On the one hand, this effect is based on the particularly large dataset used for the model's calibration. On the other hand, when estimating the dependent variable, each combination of information that occurs in the validation set is already known to the model. A problem, however, is that a high in-sample estimation accuracy frequently results from the overfitting of the model to the underlying data. In fact, in reality, most validation sets consist of unknown observations and combinations of information.

Therefore, it is reasonable, and for the estimation of LGD, it is required by the regulator, to calibrate the estimation model on a sample of the data. Article 179 (1)(d) CRR states that this sample shall be sufficient to provide the performing institution with confidence in the accuracy and robustness of its estimates.

For out-of-sample validation, these samples can be implemented by k-fold cross-validation. While earlier studies on LGD estimation used this method frequently, Kohavi (1995) employs different validation methods, such as cross-validation, leave-one-out, and random subsampling. The last divides the data into training and validation set and is run l times. In their recent study Hartmann-Wendels et al. (2014) use random subsampling to validate their regression results. Dividing the data into 75% training and 25% validation set, they repeat the procedure 25 times. Yao et al. (2015) perform a similar out-of-sample validation using 70% and 30% randomly chosen observations as training and validation set respectively. Their procedure is repeated 100 times.

We randomly draw subsamples of 75% for the training set without returning the observations. The remaining 25% form the validation set. On each training set, an estimation model is calibrated. Subsequently, we estimate the dependent variable for the corresponding validation set. We perform this step 1,000 times and average the resulting performance measures. The estimation error obtained out-of-sample is usually greater than that of in-sample validation. However, the error reflects a much more realistic allocation of the model's predictive accuracy.

The final step in validating the predictive accuracy of estimation models is out-of-time validation. In recent literature on LGD, models are rarely validated out-of-time due to special requirements to the underlying data. In particular, a comprehensive dataset and time information are necessary. For out-of-time validation, Gupton and Stein (2005) propose a growing window, subsequently using observations prior to a fixed year for the training set. The following year serves as the validation set. Most recently Altman and Kalotay (2014) conduct an out-of-sample, out-of-time simulation experiment. They calibrate a model on observations prior to 2002 and randomly draw 100 observations from the period between 2002 and 2011 for the validation set. This step is repeated 50,000 times.

	2002	2003	2004	2005	2006	2007	2008	2009
# Contracts	575	191	116	116	144	122	127	102

Table 4.3: Year of default and frequency of contracts.

However, we are convinced, that an adequate and realistic out-of-time estimation model should be built upon the available historical data and should forecast the forthcoming period. Considering a period of several years for the validation set, as proposed by Altman and Kalotay (2014), might dilute specific characteristics of single years. Hence, the estimation model would produce out-of-time results that might be too optimistic.

Employing the method of Gupton and Stein (2005), we divide our dataset according to the contracts' time of default. To calibrate a solid first model, built upon a sufficiently large dataset, we use the contracts that defaulted in 2002. Table 4.3 shows that a total of 575 contracts defaulted in the first year of the observation period.[3] The trained model then is validated on those 191 contracts, which defaulted in the following year, in this case, 2003. Calibrating the next model, we expand the training set by one year. Thereby the first two years are used for the model's calibration. Subsequently, further models are built by expanding the training set. Validation is always performed on the contracts of the year following the training period. Consequently, the final model is based on the contracts that have defaulted between 2002 and 2008. This model's predictive accuracy is validated by contracts that defaulted in 2009. Finally, we weight the outcomes of each year with the respective number of observations.

[3]The large amount of defaults in 2002 arises from an inaccuracy in the default date provided. Some of these contracts may have defaulted before 2002 but were uniformly assigned to this specific first default year. However, this inaccuracy has no impact on the out-of-time validation.

4.2.5 Performance measurements

In order to compare the results of our different estimation models, we use four performance measurements. These are: mean absolute error (MAE); mean squared error (MSE); normalized area under the regression error characteristic curve (NAREC); and Theil inequality coefficient (TIC). Each of these performance measurements focuses on the evaluation of specific aspects of the estimation.

MAE and MSE are common measures to evaluate the performance of estimation methods. With LGD and LGD* denoting the realized and estimated LGD respectively and n being the number of observations, we calculate MAE and MSE according to the following definition

$$\text{MAE} = \frac{1}{n}\sum_{j=1}^{n} |\text{LGD}_j - \text{LGD}_j^*|, \tag{4.9}$$

$$\text{MSE} = \frac{1}{n}\sum_{j=1}^{n} \left(\text{LGD}_j - \text{LGD}_j^*\right)^2. \tag{4.10}$$

MSE punishes larger deviations between predicted and realized values harder. In general, a low parameter outcome is preferable for both measurements.

NAREC can be used to evaluate the performance of regression models in total. This measure bases on the regression error characteristic (REC) curve developed by Bi and Bennett (2003). For a regression model, the REC curve draws the error tolerance δ against the models accuracy acc(δ). The latter computes as

$$\text{acc}(\delta) = \frac{\#\left\{\text{LGD}^* \colon |\text{LGD}_j^* - \text{LGD}_j| \leq \delta,\ j = 1, \ldots, n\right\}}{n}. \tag{4.11}$$

It specifies the percentage of observations whose estimates do not exceed the error tolerance. A higher outcome of NAREC implies more accurate estimations produced by the estimation model in total.

TIC is introduced by Theil (1967) and sets the mean squared error in relation to the sum of the quadratic realized and estimated LGD. It aims to quantify the goodness of fit and robustness of a model. We use

$$\mathrm{TIC} = \frac{\frac{1}{n}\sum\limits_{i=1}^{n}(\mathrm{LGD}_i - \mathrm{LGD}_i^*)^2}{\sqrt{\frac{1}{n}\sum\limits_{i=1}^{n}(\mathrm{LGD}_i^*)^2} + \sqrt{\frac{1}{n}\sum\limits_{i=1}^{n}\mathrm{LGD}_i^2}}, \tag{4.12}$$

to calculate the TIC. A low parameter outcome is preferable.

To measure the performance of classification methods, we use the classification error (CE). It is the ratio of misclassified cases to all cases, which is the relative frequency of misclassification. We calculate CE as

$$\mathrm{CE} = \frac{1}{k}\sum_{i=1}^{k} I(\mathrm{AL}_i^* \neq \mathrm{AL}_i) = \frac{\text{Misclassified cases}}{\text{All classified cases}}, \tag{4.13}$$

where AL_i is the realized class of case i defined in Equation (4.7), AL^* is the estimated class, I is the indicator function, and k is the number of classified cases.

4.3 Results

The performance results of our applied models for the three different validation techniques are contained in Tables 4.4–4.6. Overall, the results of direct estimation are as expected. The complex RF model produces accurate and robust estimates in general. It outperforms OLS, particularly in-sample and also notably out-of-sample. However, the out-of-time validation reveals problems of complex models in predicting unseen data that are known to the literature (compare Hartmann-Wendels et al. (2014) and Töws (2014)). Although RF yields stable out-of-time results, OLS estimates are slightly more accurate. We attribute this outcome to the overfitting of the RF model to the calibration data. This model's exceedingly good in-sample errors indicate an overly close adaption to the training data.

On the one hand, from a methodological perspective our multi-step approach entails an increased complexity by splitting the LGD into ALGD and MLGD. On the other hand, this splitting founds on the economic consideration of leasing specifics. The results point out that our multi-step approach indeed improves the estimation accuracy compared to direct estimation of LGD. This outcome is independent of whether the multi-step model bases on OLS or RF. Moreover, it is particularly true for out-of-sample and out-of-time validation. Consequently, the increase in model complexity does not lead to overfitting. In fact, all used validation techniques confirm the robustness and accuracy of the multi-step estimates.

The results of the different validation techniques emphasize that it is beneficial to implement our multi-step approach instead of using direct estimation. However, it is challenging to decide whether the multi-step approach should be used with OLS or RF. The latter is clearly favorable in-sample and out-of-sample. Nevertheless, for practical usage it should be taken into account that the multi-step approach with OLS generates the most accurate out-of-time estimates.

4.3.1 In-sample validation

The in-sample results that we present in Table 4.4 show that RF produces more accurate estimates than OLS throughout the performance measurements. Independent of whether the direct or the multi-step approach is used, the RF model strictly outperforms both OLS models. In particular, the exceptionally low errors in terms of MAE and MSE indicate a close adaption of RF to the training data.

In general, the multi-step approach is beneficial to both methods. We observe a distinctive advantage of the multi-step model compared to direct estimation. According to the additional information presented in Table 4.8, we attribute this outcome to the almost perfect classification. Upon closer inspection, we further

Method	MAE	MSE	NAREC	TIC
Direct estimation				
OLS	0.3436	0.1821	0.6593	0.1835
RF	0.1484	0.0362	0.8325	0.0340
Multi-step estimation				
OLS	<u>0.2757</u>	<u>0.1184</u>	<u>0.7252</u>	<u>0.1159</u>
RF	<u>0.0671</u>	<u>0.0095</u>	<u>0.8849</u>	<u>0.0084</u>

Table 4.4: In-sample loss given default (LGD) estimation results. The used methods are ordinary least squares (OLS) and random forest (RF) regression. The determined performance measurements are mean absolute error (MAE), mean squared error (MSE), normalized regression error characteristic curve area (NAREC), and Theil inequality coefficient (TIC). These are calculated according to Equations (4.9), (4.10), (4.11), and (4.12) respectively. Comparing the direct to the multi-step estimation approach, we underline the better results per method used.

note that the classification error of RF is half of that of OLS. Consequently, the reduction in estimation error from the direct to the multi-step approach is even larger for RF than for OLS.

Additionally, the estimation results of ALGD and MLGD, presented in Table 4.7, show that estimating ALGD is easier than estimating MLGD. This confirms our expectations outlined in Section 4.1. Moreover, as we hypothesize in Section 4.2.2, ALGD and MLGD are driven by different factors. Particularly, asset related factors have a significant influence on ALGD, such as asset type and initial value of the leased asset. Contrary, for MLGD mainly contract and customer related factors are important. Again, we note a markedly higher estimation accuracy for RF compared to OLS. Consequently, the advantage of RF when forecasting ALGD and MLGD improves the classification accuracy.

4.3.2 Out-of-sample validation

The out-of-sample results that we report in Table 4.5 mostly confirm the findings of the in-sample validation. However, the performance gaps between the models

Method	MAE	MSE	NAREC	TIC
Direct estimation				
OLS	0.3505	0.1894	0.6538	0.1907
RF	0.3272	<u>0.1722</u>	0.6768	0.1725
Multi-step estimation				
OLS	<u>0.3387</u>	<u>0.1777</u>	<u>0.6655</u>	<u>0.1782</u>
RF	<u>0.3233</u>	0.1772	<u>0.6813</u>	<u>0.1708</u>

Table 4.5: Out-of-sample loss given default (LGD) estimation results. The used methods are ordinary least squares (OLS) and random forest (RF) regression. The determined performance measurements are mean absolute error (MAE), mean squared error (MSE), normalized regression error characteristic curve area (NAREC), and Theil inequality coefficient (TIC). These are calculated according to Equations (4.9), (4.10), (4.11), and (4.12) respectively. Comparing the direct to the multi-step estimation approach, we underline the better results per method used.

are now less pronounced. In particular, the benefit of RF turns out to be less distinctive.

Considering the out-of-sample outcomes more closely, we again note that the RF models strictly outperform both OLS models. Moreover, in line with the in-sample results, we find that the multi-step approach is beneficial to both methods out-of-sample. With OLS, the multi-step approach outperforms direct estimation notably for all used performance measurements. With RF the multi-step approach is also advantageous in terms of MAE, NAREC, and TIC, but not concerning MSE.

Although, out-of-sample each multi-step approach has a notable advantage over the respective direct estimation model, in absolute terms, this advantage is not as significant as in-sample. The reason is that the classification error in Table 4.8 increases similarly for RF and OLS in the out-of-sample validation compared to in-sample. Surprisingly, the classification error indeed increases slightly more for RF. Consequently, in contrast to the in-sample results, out-of-sample the improvement of using the multi-step approach is more distinctive for OLS than for RF.

Typically, an increased classification error particularly affects the MSE. Compared to MAE, NAREC, and TIC, the MSE penalizes large deviations of estimates from their realized value stronger. Consequently, for the multi-step approach already few falsely classified observations might increase the MSE significantly. That might be the case, even if the estimates are more accurate in general compared to direct estimation. Interestingly, we observe this effect only for RF but not for OLS. Apparently, OLS corrects for the bias of incorrectly classified contracts by estimating conditional expectations.

Again, as in-sample, we note more accurate estimations of ALGD than of MLGD. Moreover, Table 4.7 shows that RF once more produces lower errors than OLS. However, in line with the estimation of the overall LGD, out-of-sample the advantage of RF over OLS is less pronounced than in-sample.

4.3.3 Out-of-time validation

Our results of the most realistic scenario, the out-of-time validation, are presented in Table 4.6. Concerning the multi-step approach, we find that the outcomes confirm our in-sample and out-of-sample findings. However, the results differ concerning the direct estimation models.

Considering these, we find RF not to be strictly advantageous anymore. We rather observe better outcomes for OLS in terms of MAE, NAREC, and TIC. We attribute this finding to the overly good adaptation of RF to the training data, which becomes obvious when we regard the in-sample accuracy discussed in Section 4.3.1. Therefore, RF seems to experience difficulties with validation sets that differ significantly from the training sets. Related literature frequently observes relatively poor out-of-sample and particularly poor out-of-time estimates of complex models with an excellent in-sample performance compared to OLS. Hartmann-Wendels et al. (2014) address this phenomenon concerning finite mixture models and Töws (2014) observes similar results in particular for RF.

Method	MAE	MSE	NAREC	TIC
Direct estimation				
OLS	0.3451	0.1876	0.6632	0.1959
RF	0.3457	<u>0.1830</u>	0.6605	0.2003
Multi-step estimation				
OLS	<u>0.3372</u>	<u>0.1778</u>	<u>0.6694</u>	<u>0.1897</u>
RF	<u>0.3412</u>	0.1858	<u>0.6611</u>	<u>0.1963</u>

Table 4.6: Out-of-time loss given default (LGD) estimation results. The used methods are ordinary least squares (OLS) and random forest (RF) regression. The determined performance measurements are mean absolute error (MAE), mean squared error (MSE), normalized regression error characteristic curve area (NAREC), and Theil inequality coefficient (TIC). These are calculated according to Equations (4.9), (4.10), (4.11), and (4.12) respectively. Comparing the direct to the multi-step estimation approach, we underline the better results per method used.

A closer examination of the out-of-time outcomes shows that the multi-step approach still has a general advantage over direct estimation. Being most accurate out-of-time, the multi-step approach with OLS clearly outperforms the respective direct estimation, independent of the regarded performance measurement. However, the advantage of the OLS multi-step model over its direct estimation is slightly smaller than out-of-sample. This result could have been expected because classification is even more difficult out-of-time. The reason is that validation data might differ significantly from training data. This effect is documented by the somewhat increased classification error in Table 4.8.

Analyzing the outcomes of the multi-step approach and direct estimation with RF, we observe many similarities to out-of-sample. As in the out-of-sample validation, the multi-step approach with RF outperforms the respective direct model in terms of MAE, NAREC, and TIC. However, the multi-step approach produces a higher MSE. With respect to the classification errors in Table 4.8, we attribute the latter again to some few falsely classified observations. As mentioned before, these false classifications result in a rather large deviation of predicted from realized LGD. By improving the estimation accuracy using the multi-step approach, RF achieves better MAE and MSE values than direct OLS. Nevertheless, when we compare the results of both multi-step models, OLS remains throughout advantageous. Moreover, in line with the out-of-sample findings, the benefit of using the multi-step approach is again more distinctive for OLS than for RF.

The estimation of ALGD and MLGD conforms to our expectations. We still note that ALGD estimation is easier and, therefore, more accurate than that of MLGD. On closer inspection of Table 4.7, we see that RF again produces at least slightly lower out-of-time errors than OLS in the MLGD estimation. However, contrary to the in-sample and out-of-sample findings, OLS becomes somewhat advantageous in the ALGD estimation. We attribute these comparatively inaccurate estimates of RF to its general difficulties of forecasting future observations. These results might also contribute to the fact that the advantage of the multi-step approach over the direct model is smaller when we use RF instead of OLS.

4.3.4 Further estimation and classification

Calculating separate LGD ratios and classifying the contracts increases the complexity of the estimation process from a methodological perspective. Recent studies show that an increased complexity might influence the estimation accuracy negatively (compare e. g. Qi and Zhao (2011)). However, in previous sections we show that our multi-step approach is clearly advantageous compared to direct estimation. We attribute this to the fact that our approach bases on economic considerations. Nevertheless, the accuracy of the final LGD estimation of our multi-step approach crucially depends on the respective estimation and classification accuracy in each step.

We first analyze the estimations of ALGD and MLGD, outlined in Table 4.7. As expected, throughout all validation techniques and methods, we note that estimating ALGD is easier than estimating MLGD. Moreover, comparing the results of ALGD with those of direct LGD estimation, shown in Tables 4.4–4.6, we find that the estimates of ALGD are significantly more accurate than those of LGD.

Method	In-sample		Out-of-sample		Out-of-time	
	MAE	MSE	MAE	MSE	MAE	MSE
ALGD estimation						
OLS	0.2884	0.1385	0.2945	0.1455	0.3097	0.1583
RF	0.1256	0.0298	0.2739	0.1385	0.3101	0.1588
MLGD estimation						
OLS	0.3758	0.2227	0.3829	0.2310	0.3836	0.2211
RF	0.1629	0.0452	0.3557	0.2103	0.3780	0.2186

Table 4.7: Asset-related loss given default (ALGD) and miscellaneous LGD (MLGD) estimation results. The realized ALGD and MLGD are calculated according to Equation (4.3). The used methods are ordinary least squares (OLS) and random forest (RF) regression. The determined performance measurements are mean absolute error (MAE) and mean squared error (MSE). These are calculated according to Equations (4.9) and (4.10) respectively. The table summarizes the results of the three validation techniques: in-sample, out-of-sample, and out-of-time.

Additionally, in particular in-sample and out-of-sample, the results of MLGD are only slightly worse than their counterparts of the LGD. This little difference seems to contribute to the advantage of our multi-step approach over direct estimation. Conform to direct estimation, we further observe that RF outperforms OLS both in-sample and out-of-sample regarding ALGD and MLGD. This effect is particularly evident in-sample. Out-of-sample the advantage of RF over OLS is less pronounced because the level of the estimation error significantly increases for RF but remains stable for OLS. Consequently, RF benefits more from using the multi-step approach in-sample than OLS, whereas out-of-sample the opposite holds. Moreover, out-of-time, the mentioned difficulties of RF in forecasting unseen observations result in slightly more accurate MLGD estimations than OLS, but worse ALGD predictions.

The second crucial aspect of generating accurate LGD estimations with our multi-step approach is the classification. After we estimate ALGD and MLGD with OLS or RF in the first step, classification is performed by random forest classification. Because the estimates of ALGD and MLGD are used to classify the contracts, we report the classification results labeled OLS and RF in Table 4.8.

Method	In-sample	Out-of-sample	Out-of-time
OLS	0.0248	0.2191	0.2484
RF	0.0100	0.2215	0.2364

Table 4.8: Classification results of classifying according to Equation (4.7). The classification error is calculated according to Equation (4.13). We use random forest classification in each case. The incorporated estimates of asset-related loss given default (ALGD) and miscellaneous LGD (MLGD) from step one of our approach are estimated by ordinary least squares (OLS) and random forest (RF) regression. The table summarizes the results of the three validation techniques: in-sample, out-of-sample, and out-of-time.

We find that the classification error varies significantly according to the validation technique. As expected, the classification is very precise in-sample but at about 20 times this rate out-of-sample and out-of-time. Nevertheless, classification remains sufficiently accurate as the multi-step approach still yields more accurate LGD predictions than direct estimation. However, the advantage is not as pronounced as in-sample. In general, it should be noted that despite the advantage of our multi-step approach over direct estimation, the concrete accuracy of the final LGD estimation depends on the applied method. For instance, if OLS generates better direct estimates than RF, its multi-step approach also produces more accurate results than that of RF.

Classification accuracy is not only important from a methodological perspective but also from an economic point of view. Based on the outcome of the classification the lessor might decide to restrict the workout process to the disposal of the leased asset. However, a false restriction results in waiving additional payment collection during the workout process and affects the realized LGD negatively. While classification is almost perfect in-sample, we see in Table 4.8 that out-of-sample and out-of-time classification is less reliable. Here the classification error is about 22% and 24% respectively. False classification typically arises from classification probabilities near 50% indicating that classification is ambiguous. These are probably cases in which a lessor does not restrict the workout process to the asset's disposal, although our classification would suggest it. To address such am-

Method	In-sample	Out-of-sample	Out-of-time
OLS	0.0000	0.1336	0.1726
RF	0.0000	0.1745	0.2089

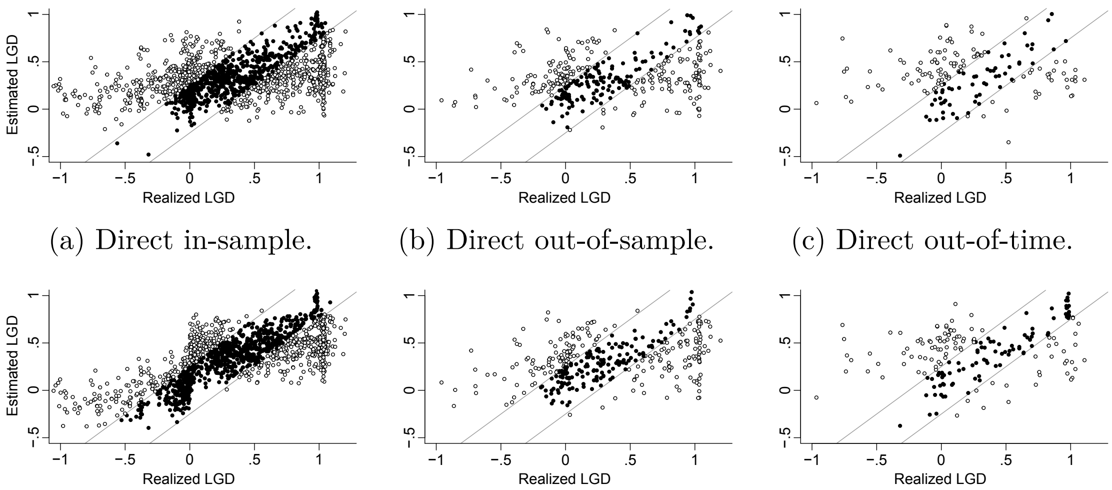
Table 4.9: Classification results of classifying according to Equation (4.7), considering exclusively classification probabilities below 25% or above 75%. The classification error is calculated according to Equation (4.13). We use random forest classification in each case. The incorporated estimates of asset-related loss given default (ALGD) and miscellaneous LGD (MLGD) from step one of our approach are estimated by ordinary least squares (OLS) and random forest (RF) regression. The table summarizes the results of the three validation techniques: in-sample, out-of-sample, and out-of-time.

biguous cases, we present the classification results for contracts with classification probability below 25% or above 75% in Table 4.9. As expected, we note that the classification error decreases consistently. In particular, regarding OLS we note significantly lower classification errors of about 13% out-of-sample and 17% out-of-time. Consequently, for these contracts the classification is clearly more reliable and seems to be suited for practical use.

4.3.5 Interpretation

The previously discussed results clearly show the benefit of our multi-step approach compared to direct LGD estimation in terms of the used performance measurements. While the chosen measurements are convenient for precise comparison of the models, the scatter plots in Figure 4.5 provide an additional visual proof of our findings. In particular, the figures allow for a more detailed analysis than the aggregated measures MAE, MSE, or NAREC. In these figures the perfect estimation of LGD would be located on the diagonal through the plot's origin. We draw two diagonal lines to frame a 0.5-wide interval around the perfect estimation. The interval contains all estimates that are close to the realized LGD. These are displayed as solid points.

The scatter plots in Figure 4.5 refer to OLS, but the outcomes are similar concerning RF. According to Figure 4.5a, in-sample direct estimation produces a



(a) Direct in-sample. (b) Direct out-of-sample. (c) Direct out-of-time.

(d) Multi-step in-sample. (e) Multi-step out-of-samp. (f) Multi-step out-of-time.

Figure 4.5: Visual comparison of realized and estimated loss given default (LGD). Figures a–c display direct OLS estimations in the in-sample, out-of-sample, and out-of-time validation respectively. Out-of-sample we randomly choose and display one run out of 1,000. Out-of-time we plot estimates and realizations of LGD of one year. Figures d–f show the counterparts of the multi-step approach. The simple diagonal lines frame a 0.5-wide interval to highlight estimates close to their realized value. Additionally, these points are solid, whereas points outside the interval and, thus, far from their realized LGD are hollow.

large number of accurate estimates. Nevertheless, the multi-step estimation in Figure 4.5d generates a larger number of accurate estimates on the whole range of realized LGD values. In particular, due to a downward shift in estimation, it is visibly better than direct estimation for realized LGDs smaller than 0. Concerning the out-of-sample estimates in Figures 4.5b and 4.5e, we also note that the multi-step approach is again more accurate than direct estimation. To be precise, we observe a significantly higher concentration of estimates within the drawn interval for the multi-step model compared to direct estimation. This observation is particularly true for realized LGDs larger than about 0.3. Moreover, again the multi-step estimates tend to converge toward their realized value in general. The outcomes of the out-of-time validation in Figures 4.5c and 4.5f show a similar picture to that of in-sample and out-of-sample. Compared to direct estimation the predictions of the multi-step approach move closer to the diagonal lines from outside the interval. The increased number of estimates within the drawn interval indicates notably precise predictions for realized LGDs larger than 0.5.

For all three validation techniques the scatter plots in Figure 4.5 confirm that the estimates of the multi-step approach tend to converge toward their realized value. Therefore, the results of our performance measurements should not be affected by outliers. In fact, the plots reflect the advantage of our multi-step approach over direct estimation.

Our results in Section 4.3 and the scatter plots in Figure 4.5 clearly show that the proposed multi-step approach outperforms direct estimation of the LGD. To evaluate the results in the context of related literature, we summarize the results of several studies in Table 4.10. Yao et al. (2015) argue that it is hard to compare empirical results when using different data and information sets. Nevertheless, they compare absolute values of R^2 from selected literature on LGD prediction performance. Instead of using absolute values we propose to examine the improvement of the estimation accuracy of a model compared to OLS. Almost all related studies use the latter as benchmark. For comparison, we focus on MAE, MSE, and root mean squared error (RMSE). More precisely, in Table 4.10 we present the improvement of a study's best model compared to OLS regarding the respective performance measurement for out-of-sample and out-of-time validation.

Across the performance measurements, the authors primarily achieve improvements in the range from 2% to 10%. One major exception is Bastos (2013) with improvements around 25%. This exceeding improvement of the estimation accuracy might be attributed to specific characteristics of the used data. The most frequently reported performance measurement in the literature is MAE. Out-of-sample our multi-step approach clearly achieves the highest MAE improvement of the considered studies except for Bastos (2013). Concerning MSE or RMSE particularly Bastos (2010) reports a promising increase in estimation accuracy

[4] Bastos (2010) uses the historical average as benchmark. Hence, the value reported is the performance increase of RT compared to the historical average.

Study	Data	Best technique	ΔMAE	ΔMSE	ΔRMSE
Out-of-sample validation					
Bastos (2010)	SME loans	RT			6.9[4]
Loterman et al. (2012)	Types of loans	SVM & NN	5.0		
Zhang and Thomas (2012)	Personal loans	OLS	3.7		
Bastos (2013)	Loans & bonds	RT ensemble	28.0	25.0	
Hartmann-Wendels et al. (2014)	Leases	Model tree	5.5		0.8
This study	**Leases**	**MS approach**	**7.8**	**6.5**	
Out-of-time validation					
Bastos (2010)	SME loans	RT			6.7[4]
This study	**Leases**	**MS approach**	**2.3**	**6.2**	

Table 4.10: Comparison of performance improvements in loss given default estimation literature. The table reports the percentage improvement of a study's best model compared to direct ordinary least squares (OLS) regression. The error measurements are: mean absolute error (MAE); mean squared error (MSE); and root MSE (RMSE). The techniques are: regression tree (RT); support vector machine (SVM); and neural network (NN); two-step (TS) approach; and multi-step (MS) approach.

using regression trees. However, instead of OLS he uses the historical average as benchmark. This outcome should be treated with caution. According to the results of Hartmann-Wendels et al. (2014), OLS performs at least 3% better than the historical average in terms of RMSE. Yashkir and Yashkir (2013) find a similar deviation. Therefore, comparing the regression tree results of Bastos (2013) with OLS, the improvement probably would not exceed 4%. Consequently, our multi-step approach obtains also competitive results in terms of MSE and RMSE.

Currently, several studies report out-of-sample errors, but out-of-time results are scarce. Hence, concerning the latter, we can hardly evaluate our multi-step approach. Nevertheless, when we consider the benchmark used by Bastos (2010), our multi-step approach seems to generate good out-of-time results. Moreover, we emphasize that our multi-step approach can indeed perform better than direct OLS out-of-time. As our results of the direct estimation with RF indicate, it is not common that complex models that perform well out-of-sample also produce stable and accurate out-of-time estimates.

4.4 Conclusion

The development of an appropriate and dynamic model for estimating LGD requires the consideration of mathematical aspects and economic factors. For defaulted leasing contracts, we argue that detailed consideration of the revenues during the workout process is a key driver to improve LGD forecasting accuracy. Instead of the traditional holistic consideration of LGD, we separate LGD into an asset-related and a miscellaneous share. This separation is economically reasonable because, typically, cash flows have different sources. To account for the different revenues at the time of contracts' default, we estimate an asset-related LGD (ALGD) and a miscellaneous LGD (MLGD).

Leasing companies are experts in evaluating and disposing of their leased assets. The estimation of the ALGD takes this expertise into account. Moreover, together with the estimated MLGD it provides decision support for actions to be taken during the workout process. We show that ALGD is theoretically an upper boundary to the LGD. Likewise, the estimation of MLGD yields economic value. Its value indicates whether the effort of collecting overdue payments during the workout process will be rewarding or rather unprofitable considering incurred workout costs. Consequently, we present a guideline to organize the workout process, i. e., in case workout costs are expected to exceed collected payments, the workout process should be restricted to the disposal of the leased asset.

This finding is particularly interesting because cash flows from the asset's disposal are positive in 99% of the cases, net of disposal costs. However, for 35% of the contracts, the collection of miscellaneous payments turns out to generate losses due to the incurred costs. We find that following our suggestion to restrict the workout process to the asset's disposal in certain cases would reduce the average LGD significantly. With our data, the reduction of the average LGD amounts to 10% or €2,250,000 in absolute losses.

Based on the sophisticated economic separation of the LGD we introduce a new multi-step LGD estimation approach. We apply our approach to a real-life dataset of a German leasing company and perform in-sample, out-of-sample, and out-of-time validation. While the approach supports the workout process, we find that the separation of LGD is very beneficial to its estimation accuracy. We apply OLS and RF regression to our approach. With both methods, we note a significant increase in estimation accuracy compared to the benchmarking results of the respective direct estimation. The proposed multi-step approach is more complex than direct estimation. However, the increase in complexity does not lead to overfitting, which is a common concern of advanced LGD estimation models. Nonetheless, the interpretability of the variables' influence might suffer slightly. However, to put it in Bastos (2013) words, it is often the case that simplicity has to be sacrificed in order to achieve a higher degree of precision.

5 Summary and conclusion

This dissertation investigates different approaches to estimate the loss given default (LGD) of financial assets. In the first study, we analyze fundamental differences in parametric, semiparametric, and nonparametric estimation techniques. The second work focuses on the recovery of defaulted lease contracts and its impact on LGD. The final essay introduces a new approach to step-wisely estimate LGD.

The key difference of parametric and nonparametric methods is their assumption about the distribution of the underlying data. While the latter do not assume any distribution, parametric methods are always based on a specific probability distribution, such as normal or beta distribution. Concerning the found bimodality of LGD density, using a parametric approach for LGD estimation is theoretically implausible. Finite mixture models focus on reproducing the distribution of LGD. However, this focus is rather inefficient out-of-sample. In contrast, nonparametric regression and model trees are not concerned with distribution approximation. Despite the unusual shape of the LGD distribution, these methods produce particularly robust estimates on a low level of error. In order to realize this superiority, we find that nonparametric methods, such as model trees, require a large data and information base. On small datasets, however, parametric linear regression turns out to be partially advantageous.

In a follow-up study, we examine contracts that defaulted at one time during contract term but recovered later. This development is possible if the initial default reason is resolved. The opportunity to recover depends to a large extent on the lender's workout policy. Indeed, there are lenders who liquidate the collateral and write-off contracts immediately after their default. However, we find empirical evidence that recovered contracts yield significantly lower LGDs. The reason is that after recovery, these contracts continue properly. Delayed or potentially deferred payments, as well as new payment plans, cause the lender only small losses if any. Moreover, when predicting the LGD, we account for this special event of defaulted contracts and its counterpart, the write-off. Consequently, we can explain a significantly higher portion of LGD's variance than without this information. Also, estimation of LGD becomes more accurate. Regarding these results, we conclude that lenders would benefit from developing expertise in dissolving the default of their debtors. They at least might want to assess these contract's continuation worthiness precisely. In any case, we find that the distinction between recovered and written off contracts leads to lower errors in LGD prediction.

Finally, we break with the traditional holistic consideration of LGD. By separating it into disjointed components, we account for the two major cash flow sources in the workout process of defaulted leases. These sources are asset proceeds net of the incurred disposal costs and incoming cash flows from the lessee and collateral net of collecting costs. Moreover, by calculating the asset-related LGD, we determine a theoretical upper limit to the LGD. When forecasted, this boundary is an important decision-making factor concerning the actions the lessor should take during the workout process. The developed multi-step approach estimates the LGD components, accounts for the determined limit, and it uses these factors to predict LGD. We find that this approach is beneficial to the estimation accuracy, independent of the concrete estimation models. This result even holds in the ultimate out-of-time validation.

Bibliography

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

Altman, E.I., Ramayanam, S., 2007. Default and Returns in the High-Yield Bond Market 2006 in Review and Outlook. Technical Report. Citigroup Global Markets.

Basel Committee on Banking Supervision, 2003. Overview of The New Basel Capital Accord. Bank for International Settlements.

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

Bastos, J.A., 2013. Ensemble Predictions of Recovery Rates. Journal of Financial Services Research 46, 177–193.

Bellotti, T., Crook, J., 2012. Loss given default models incorporating macroeconomic variables for credit cards. International Journal of Forecasting 28, 171–182.

Bi, J., Bennett, K.P., 2003. Regression Error Characteristic Curves, in: Twentieth International Conference on Machine Learning (ICML-2003), Washington D.C.. 43–50.

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32.

Calabrese, R., Zenga, M., 2010. Bank loan recovery rates: Measuring and non-parametric density estimation. Journal of Banking & Finance 34, 903–911.

Campbell, J.Y., Thompson, S.B., 2008. Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? Review of Financial Studies 21, 1509–1531.

Chalupka, R., Kopecsni, J., 2009. Modeling Bank Loan LGD of Corporate and SME Segments: A Case Study. Czech Journal of Economics and Finance 59, 360–382.

Committee of European Banking Supervisors, 2006. Guidelines on the implementation, validation and assessment of Advanced Measurement (AMA) and Internal Ratings Based (IRB) Approaches (GL10).

De Laurentis, G., Riani, M., 2005. Estimating LGD in the Leasing Industry: Empirical Evidence from a Multivariate Model, in: Altman, E., Resti, A., Sironi, A. (Eds.), Recovery Risk: The Next Challenge in Credit Risk Management. Risk Books. Chapter 8, 143–164.

Dermine, J., Neto de Carvalho, C., 2005. How to Measure Revcoveries and Provisions on Bank Lending: Methodology and Empirical Evidence, in: Altman, E., Resti, A. (Eds.), Recovery Risk: The Next Challenge in Credit Risk Management. Risk Books. Chapter 6, 101–119.

Eisfeldt, A.L., Rampini, A.A., 2009. Leasing, Ability to Repossess, and Debt Capacity. Review of Financial Studies 22, 1621–1657.

Elbracht, H.C., 2011. Statistische Methoden zur Quantifizierung und Schätzung des Loss Given Default. Ph.D. thesis. University of Cologne. 2011.

Freund, Y., Schapire, R.E., 1996. Experiments with a New Boosting Algorithm. Working paper.

Frühwirth-Schnatter, S., 2006. Finite Mixture and Markov Switching Models. Springer Series in Statistics, Springer, New York.

Friedman, C., Sandow, S., 2005. Estimating Conditional Probability Distributions of Recovery Rates: A Utility-Based Approach, in: Altman, E., Resti, A. (Eds.), Recovery Risk: The Next Challenge in Credit Risk Management. Risk Books. Chapter 19, 347–359.

Frontczak, R., Rostek, S., 2015. Modeling loss given default with stochastic collateral. Economic Modelling 44, 162–170.

Frye, J., 2005. A False Sense of Security, in: Altman, E., Resti, A., Sironi, A. (Eds.), Recovery Risk: The Next Challenge in Credit Risk Management. Risk Books. Chapter 10, 187–200.

Gadd, A., Wold, H., 1964. The janus coefficient: a measure for the accuracy of prediction. North-Holland, Amsterdam.

Gibilaro, L., Mattarocci, G., 2007. The Selection of the Discount Rate in Estimating Loss Given Default. Global Journal of Business Research 1, 15–33.

Gray, J.B., Fan, G., 2008. Classification tree analysis using TARGET. Computational Statistics & Data Analysis 52, 1362–1372.

Grippa, P., Iannotti, S., Leandri, F., 2005. Recovery Rates in the Banking Industry: Stylised Facts Emerging from the Italian Experience, in: Altman, E., Resti, A. (Eds.), Recovery Risk: The Next Challenge in Credit Risk Management. Risk Books. Chapter 7, 121–141.

Grün, B., Leisch, F., 2007. Fitting finite mixtures of generalized linear regressions in R. Computational Statistics & Data Analysis 51, 5247–5252.

Gürtler, M., Hibbeln, M., 2013. Improvements in loss given default forecasts for bank loans. Journal of Banking & Finance 37, 2354–2366.

Gupton, G.M., Stein, R.M., 2005. LossCalc v2: Dynamic Prediction of LGD. Technical Report. Moody's KMV Company.

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

Hartmann-Wendels, T., Honal, M., 2010. Do Economic Downturns Have an Impact on the Loss Given Default of Mobile Lease Contracts? An Empirical Study for the German Leasing Market. Kredit und Kapital 43, 65–96.

Hartmann-Wendels, T., Miller, P., Töws, E., 2014. Loss given default for leasing: Parametric and nonparametric estimations. Journal of Banking & Finance 40, 364–375.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning. Springer New York.

Hlawatsch, S., Ostrowski, S., 2011. Simulation and estimation of loss given default. The Journal of Credit Risk 7, 39–73.

Jankowitsch, R., Nagler, F., Subrahmanyam, M.G., 2014. The determinants of recovery rates in the US corporate bond market. Journal of Financial Economics 114, 155–177.

Kaastra, I., Boyd, M., 1996. Designing a neural network for forecasting financial and economic time series. Neurocomputing 10, 215–236.

Khieu, H.D., Mullineaux, D.J., Yi, H., 2011. The determinants of bank loan recovery rates. Journal of Banking & Finance 36, 923–933.

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Joint Conference on Artificial Intelligence (IJCAI), Morgan Kaufmann. 1137–1143.

Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer, New York.

Kuhn, M., Weston, S., Coulter, N., 2013. C5.0 Decision Trees and Rule-Based Models. Technical Report. Comprehensive R Archive Network.

Lasfer, M.A., Levis, M., 1998. The determinants of the leasing decision of small and large companies. European Financial Management 4, 159–184.

Laurent, M., Schmit, M., 2005. Estimating "Distressed" LGD on Defaulted Exposures: A Portfolio Model Applied to Leasing Contracts, in: Altman, E., Resti, A., Sironi, A. (Eds.), Recovery Risk: The Next Challenge in Credit Risk Management. Risk Books. Chapter 17, 307–322.

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

Leow, M., Mues, C., Thomas, L., 2014. The economy and loss given default: evidence from two UK retail lending data sets. Journal of the Operational Research Society 65, 363–375.

Li, P., Qi, M., Zhang, X., Zhao, X., 2014. Further Investigation of Parametric Loss Given Default Modeling. Working paper.

Liaw, A., Wiener, M., 2012. Breiman and Cutler's random forests for classification and regression. Technical Report. Comprehensive R Archive Network.

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

Miller, P., 2015. Does the economic situation affect the Loss Given Default of leases? Working paper.

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

Quinlan, J.R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann.

Quinlan, J.R., 1996. Bagging, Boosting, and C4.5, in: In Proceedings of the Thirteenth National Conference on Artificial Intelligence, AAAI Press. 725–730.

Renault, O., Scaillet, O., 2004. On the way to recovery: A nonparametric bias free estimation of recovery rate densities. Journal of Banking & Finance 28, 2915–2931.

Schmit, M., 2004. Credit risk in the leasing industry. Journal of Banking & Finance 28, 811–833.

Schmit, M., Degouys, C., Delzelle, D., Stuyck, J., Wautelet, F., 2003. Credit Risk in the Leasing Business – A case study of low probability of default. Working paper.

Schmit, M., Stuyck, J., 2002. Recovery Rates in the Leasing Industry. Working paper presented at Leaseurope's Annual Working Meeting, Salzburg.

Theil, H., 1967. Applied economic forecasting. North-Holland, Amsterdam.

Töws, E., 2014. The impact of debtor recovery on loss given default. Working paper.

Wang, Y., Witten, I.H., 1997. Inducing Model Trees for Continuous Classes. Working paper.

Witten, I.H., Frank, E., Hall, M.A., 2011. Data Mining: Practical Machine Learning Tools and Techniques. Elsevier Science.

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

Yashkir, O., Yashkir, Y., 2013. Loss given default modeling: a comparative analysis. Journal of Risk Model Validation 7, 25–59.

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, 204–215.