Predicting Loss Given Default in Leasing: A Closer Look at Models and Variable Selection*

Florian Kaposty[†]
University of Muenster

Johannes Kriebel[§]
University of Muenster

Matthias Löderbusch[‡]
University of Muenster

April 27, 2017

* We are grateful for valuable comments from participants of the GOR Workshop on Financial Management and Financial Institutions 2017 in Magdeburg.

† Corresponding author, Finance Center Muenster, University of Muenster, Universitaetsstr. 14-16, 48143 Muenster, Germany, phone +49-251-83-29924, fax +49-251-83-22882, florian.kaposty@wiwi.uni-muenster.de.

§ Finance Center Muenster, University of Muenster, Universitaetsstr. 14-16, 48143 Muenster, Germany, johannes.kriebel@wiwi.uni-muenster.de.

‡ Finance Center Muenster, University of Muenster, Universitaetsstr. 14-16, 48143 Muenster, Germany, matthias.loederbusch@wiwi.uni-muenster.de.

Abstract

Along with the introduction of the Basel II Accord and its huge implications for credit risk management, modeling, estimating, and forecasting the loss given default (LGD) have become increasingly important tasks. Whereas much is known about the LGD for loans and bonds, this is not the case for leasing contracts. Using a proprietary data set of 1,184 defaulted corporate leases in Germany, this study explores different parametric and non-parametric approaches attempting to forecast and explain the LGD. We compare the in-sample and out-of-sample prediction accuracies of the historical average, OLS and Tobit regressions; in terms of non-parametric methods, we consider regression trees (RTs), random forests (RFs), and artificial neural networks (ANNs). By conducting these analyses for different points in time, we study how prediction accuracy changes depending on the set of available information. Furthermore, using a variable importance measure, we identify the input variables with the greatest effect on LGD prediction accuracy for each method. We find that (1) more sophisticated non-parametric methods, especially RFs, generally outperform linear methods leading to a remarkable increase in the prediction accuracy, (2) default related information improves prediction accuracy considerable, and (3) the outstanding exposure at default, the work-out duration, and an internal rating turn out to be important drivers of accurate LGD predictions.

Keywords: Loss given default, forecasting, variable selection methods, leasing.
JEL Classification: C14 (semi-parametric and non-parametric models), C53 (forecasting and prediction methods), G17 (financial forecasting and simulation), G32 (financial risk and risk management).

I Introduction

Modeling loss given default (LGD) has become increasingly important for banks since the Basel II and III frameworks were introduced. Especially in the advanced internal ratings-based approach (A-IRBA) banks are required to use own estimates for LGD to quantify their credit risk exposure accurately. Along with the regulatory changes in credit risk measurement, a varied body of literature has more and more shifted its focus towards modeling, estimating, and forecasting LGD rather than the probability of default (PD).[1] In addition to the regulatory requirements, having accurate LGD forecasts and estimations is also important for financial institutions from an internal risk management perspective for adequate pricing and management of credit risk (Hartmann-Wendels et al. (2014)). Moreover, higher accuracy in LGD forecasts and estimations can generate competitive advantages, as precise knowledge about potential losses is indispensable for both an adequate pricing of leasing contracts and an efficient allocation of regulatory capital.

Accurate LGD predictions may be particularly important in the case of leases, since the legal title of the leased asset remains with the lessor for most leasing contracts, implying that the lessor can retain all utilization proceeds in excess of the exposure at default. Hence, the LGD of leases can take values below zero, which can be considered as a financial gain for the lessor. Moreover, in contrast to loans, it is often possible to repossess the asset promptly (Eisfeldt and Rampini (2009)). Maximizing recoveries of repossessed assets on secondary markets is therefore of major importance for the lessor and as a part of their day-to-day business not only restricted to defaulted leases. Likewise, lessors are usually supposed to have a good understanding of the liquid secondary markets, which facilitates disposal of the often standardized assets and lowers potential losses. Due to these characteristics of leasing contracts, it becomes obvious that accurate LGD predictions are indispensable for banks and leasing companies to increase revenues from asset disposal.

This paper contributes to the existing literature on LGD predictions by exploiting a unique dataset of 1,184 defaulted leasing contracts concluded with small and medium-sized enterprises (SMEs) in three ways. First, we apply different parametric and non-parametric methods to predict LGD. Interestingly, the existing literature does not present an unambiguous picture in terms of which method achieves the most accurate predictions. Whereas some work has been carried out predicting the LGD of defaulted loans using parametric

[1] Selected works on PD classification techniques include Davis et al. (1992), Hand and Henley (1997), Baesens et al. (2003), and Yeh and Lien (2009). An overview on recent contributions is given in Lessmann et al. (2015).

techniques (e.g., Yashkir and Yashkir (2013)), applying non-parametric techniques in order to estimate the LGD is a relatively new field of research (Loterman et al. (2012)).[2] Although non-parametric methods, such as random forests (RFs) and artificial neural networks (ANNs), are highly sophisticated and well-developed approaches that have attracted huge interest in statistics and machine learning, the broad application to credit risk modeling and especially LGD of some of these techniques is still lacking in the literature. In addition, extremely little is known about the goodness of fit of these modern non-parametric methods when applied in estimating the LGD of defaulted leases, which is surprising given the important role of proper risk management for leasing contracts. As LGDs are commonly known to be rather difficult to predict, particularly by parametric methods, non-parametric methods may help to detect additional patterns in the data, thereby improving the prediction performance. We contribute to this field of research by comparing the in-sample and out-of-sample accuracies of several parametric and non-parametric methods for predicting the LGD of defaulted leasing contracts.

In terms of the second area studied in this paper, almost all studies dealing with the prediction of LGD solely focus on the question of which method performs best in terms of generating the lowest prediction error rather than investigating which variables are indeed driving the results. However, simply working with the model output in order to assess the credit risk of a portfolio and looking at the methods as a black-box may induce the risk of having weak or even unstable methods. To provide first insights on these possibilities, we identify the importance of the input variables in each prediction method and show how they behave across the different methods. Having a clear view of the main determinants of LGD predictions is important for both banks and regulators to be able to evaluate the appropriateness of the applied methods. Moreover, knowing the parameters that improve the accuracy of LGD predictions may provide guidance on what information should be carefully collected for each contract.

As a third topic, the prediction accuracy of LGD may considerably depend on the available information (Hartmann-Wendels et al. (2014)). Methods only using variables that are known when the contract is concluded (hereafter referred to as LGD forecasting) are much more restricted in terms of the available set of information than those prediction models enriched with additional information gained after the work-out is completed (hereafter referred to as LGD estimation).[3] We examine whether using additional information

[2] A few nameable exceptions are discussed in Section II.

[3] To clarify our notation, we use the phrase "prediction" as a more general term comprising both forecasting and estimation. Whenever we do not unambiguously speak of either forecasting or estimation, we will refer to this as prediction.

obtained during the life of a lease improves prediction accuracy. As more information is usually available after the work-out process is complete, we expect that including this additional information will improve estimation accuracy. Updating the LGD estimates for defaulted leases by including additional information gained from the work-out is required by the regulator for banks applying the A-IRBA of the Basel framework (EU Regulation No 575/213, Article 179). Furthermore, forecasting LGDs is important for developing reliable expectations about the inherent credit risk of leasing portfolios and assessing the riskiness of new leases. Hence, both precise LGD forecasts and accurate LGD estimations are relevant for lessors.

Considering the three main aims of the study, we first show that more sophisticated prediction methods can improve LGD predictions, and identify RFs as generating the best results consistently. Second, we show that information obtained during the contract's lifetime or after default can improve predictions considerably. Third, we identify variables driving the results in the forecasting and the estimation cases.

Ultimately, we uncover a clear trend whereby RFs generate the lowest prediction errors in the forecasting and estimation case. The ANN outperforms the regression tree (RT) and the linear methods in-sample for both the forecasting and estimation cases. This resembles the fact that ANNs are more flexible than the other approaches. In the out-of-sample case, the results depend on whether we are using the methods to forecast or estimate the LGD. In most cases, the forecasting methods produce predictions that outperform the historical average in a statistically significant way. We also underline existing evidence that out-of-sample testing is crucial to obtain comparable and reliable results concerning the models' performance. By comparison, the out-of-sample estimation results for the ANN and in particular the RF do outperform the other three prediction methods in a meaningful way. To the best of our knowledge, this is the first study using RFs in LGD predictions for leases. Whereas the very good in-sample fit is not too surprising by construction, the high out-of-sample prediction accuracy clearly indicates that the RF identifies patterns in the data that the other methods will most likely not detect.

Our results clearly indicate that models built conditional on default generate considerably better predictions than unconditional models. Hence, our findings suggest that in addition to lessee and contractual related information, banks should also make use of information becoming available during the work-out process, or in some cases, even during the lifetime of the contract.

To analyze the importance of our explanatory variables, we compute the relative changes in the prediction errors for each method when permuting a single variable. By doing so, we examine the relative relevance of the permuted variable for LGD predictions. We find remarkable differences across all variables regarding their strongly different effects on the prediction accuracy. In the forecasting case, we show that the results are mainly driven by an internal rating and a distinction between industries and asset types. The RF further incorporates information about the exposure size, the distance between the lessee and the lessor, and characteristics of the business cycle. In the estimation case, the most relevant characteristic is the outstanding exposure as a fraction of the contract's original size. This is consistent with respect to all five methods. Our results identify the duration of the work-out process and the exposure at default as significantly influencing prediction accuracy. In addition, the initial internal rating and the distinction between industries and asset types are important.

The remainder of the paper is organized as follows: Section II reviews the related literature. Section III briefly describes our dataset and discusses the measurement of the LGD. Section IV introduces the applied methods and presents the methodical framework. Section V reports the empirical results. Section VI concludes.

II Literature review

A lot of work has been done in order to investigate the determinants of LGD by looking at the prices of defaulted bonds (e.g., Altman and Kishore (1996), Altman et al. (2005), Varma and Cantor (2005), and Acharya et al. (2007)) and loans (e.g., Araten et al. (2004), Frye (2005), Dermine and Neto de Carvalho (2006), Grunert and Weber (2009), Khieu et al. (2012), and Han and Jang (2013)). Although leasing has become increasingly important for (small and medium-sized) firms as an alternative form of financing, findings on the LGDs of leasing contracts are far less comprehensive in the empirical credit risk literature. Schmit and Stuyck (2002) and Laurent and Schmit (2005) show that the LGD depends on the type of contract as well as the contract's age at the time of default. Furthermore, De Laurentis and Riani (2005) point out that the LGD depends on the lessee's form of organization and whether the contract is protected with collateral. In terms of the influence of the macroeconomic environment on LGD, the existing evidence is mixed (Schmit and Stuyck (2002), Laurent and Schmit (2005), and Hartmann-Wendels and Honal (2010)).

The restricted interval of LGDs, which is typically between 0 and 1 for bonds and loans, and the frequently observed bimodality impose some challenges for LGD modeling as well as for regression analysis (Asarnow and Edwards (1995) and Grunert and Weber (2009)). More recently, the focus on LGDs in credit risk management has shifted towards proposing and benchmarking different parametric, non-parametric, and semi-parametric methods that – at least theoretically – can account for bounded and bimodally distributed LGDs. Furthermore, some studies have explicitly assessed the applicability of the proposed methods for the purpose of forecasting LGDs in an out-of-time and/or out-of-sample setup. The study by Gupton and Stein (2002) represents one of the first approaches to account for the specific LGD characteristics in a regression model by transforming raw LGDs into normally distributed LGDs using both the cumulative beta distribution function and the standard normal quantile function for bonds, loans, and preferred stock. By means of Italian bank loans, Calabrese and Zenga (2010) elucidate the drawbacks of adopting common (beta) kernel density estimates for LGD distribution modeling. As a remedy, they propose a beta inflated regression model for LGD, which is a mixture of a Bernoulli distribution and a continuous random variable estimated using a mixture of beta kernel estimators. Sigrist and Stahel (2011) extend the Tobit model using a censored gamma regression model and test it with LGDs from insurance data. They find their method to produce a better fit to the data than other tested parametric methods. Several parametric regression methods are calibrated and tested by Yashkir and Yashkir (2013) on the S&P

LossStats database, including Tobit regressions, inflated beta regressions, and censored gamma regressions. Interestingly, their study reveals that, with respect to the goodness-of-fit of their methods, the thorough choice of explanatory variables is of major importance, whereas the type of method only plays a minor role. We consider this important finding in our analyses when examining the importance of our set of explanatory variables across several prediction methods. Hlawatsch and Ostrowski (2011) explicitly take into account the bimodality of LGD distributions by using a mixture of two beta distributions. In a simulation study, they find that their approach is superior to the benchmark methods and more capable of imitating a bimodal distribution.

Turning toward non-parametric methods applied in more recent studies on the prediction of LGD, Qi and Zhao (2011) compare ANNs and RTs with several parametric methods, including fractional response regression and regression-transformation methods, for a sample based on Moody's Ultimate Recovery Database. In their study, they show that non-parametric methods usually produce more accurate results than their parametric counterparts. However, they point out that the methods' ability to generate a bimodal distribution for LGDs is only of minor importance since those methods that are able to produce such a pattern do not necessarily yield good predictions. Bastos (2010) evaluates the prediction accuracy of fractional response regression and a RT model to predict LGD out-of-sample and out-of-time for Portuguese bank loans. In several out-of-sample tests, he finds that the prediction accuracy of the analyzed methods depends on the chosen recovery horizon, clearly indicating that no method is generally superior to another.

A surprising result is obtained by Bellotti and Crook (2012) for a portfolio of UK credit cards. They benchmark a decision tree method with several other approaches, including fractional response regression, the Tobit model, and transformation regression methods for LGD and find that OLS regressions in combination with macroeconomic explanatory variables turn out to be the best forecasting approach. The prediction performance of 24 regression techniques is tested in a comprehensive benchmarking study for six bank loan portfolios by Loterman et al. (2012). Their methods comprise linear methods, non-linear methods such as RTs, support vector machines, and ANNs as well as two-stage models that basically combine logistic or OLS regressions with other linear or non-linear techniques. Although much of the LGD cannot be explained by the methods and variables under consideration, the authors observe that non-linear methods – especially ANNs and support vector machines – and the two-stage models clearly outperform the pure linear methods. To some extent, these results are confirmed by Tobback et al. (2014) for LGD data on home equity and corporate loans. They provide evidence that the best out-of-time

prediction is given by a two-stage model, whereas RTs perform best in the out-of-sample analysis. However, a pure support vector regression model is considered to produce poor results. Similar to Bellotti and Crook (2012), Tobback et al. (2014) show that the predictive power of the methods can be improved by including macroeconomic factors. Different improved support vector regression models are also considered in Yao et al. (2015) and compared to other linear and non-linear methods for corporate bonds from Moody's Ultimate Recovery Database. For the entire sample of all bonds, the proposed support vector regression techniques are found to yield the best prediction performance. However, when the sample is differentiated between the bond's seniority, a least squared support vector regression technique still provides a good model fit, while the prediction performances of the parametric fractional response regression and linear regression methods become significantly better for bonds of higher seniority classes.

In terms of leasing portfolios, prediction methods for LGDs of leases are a relatively new field of research and have only been studied in a rudimentary fashion. To the best of our knowledge, Hartmann-Wendels et al. (2014) is the only study dealing with prediction methods for leases. The authors conduct in-sample and out-of-sample analyses using model and decision trees, hybrid finite mixture distributions and OLS regressions. Their results suggest that finite mixture distributions, which try to reproduce the empirically observed LGD distribution, are only adequate for in-sample testing but fail in the out-of-sample analysis. In contrast, model trees show convincing performance in the out-of-sample tests and consistently outperform RTs. The richness of our dataset enables us to enhance the findings of Hartmann-Wendels et al. (2014) in two important ways. First, we also examine which explanatory variables are the main determinants of LGD predictions. Second, we propose new parametric and non-parametric prediction methods for leases that take into account the characteristics of the LGD of leases and/or have been found to produce noteworthy results for loans.

III Data

We use a proprietary dataset provided by a mid-sized German bank containing information about 26,735 active and terminated leasing contracts concluded with SMEs, including one-man businesses, between 1999 and 2014. In total, 1,189 contracts have defaulted between April 2009 and December 2014. The dataset provides us with a great variety of information on (1) lessee and firm characteristics, (2) contractual characteristics, and (3) additional data on defaulted leases. This enables us to split up the set of variables potentially influencing the LGD in information already known when the contract is concluded (groups (1) and (2)) and default-related information that becomes accessible when the work-out is completed (group (3)). Including information obtained from group (3) implies that our model is calibrated conditional on default whereas a model ignoring group (3) data would be unconditional. We expect that conditional models using additional information gathered during the work-out process and during the lifetime of the lease will show better prediction accuracies. Gürtler and Hibbeln (2013) raise several issues concerning the data selection in the context of LGD that may result in biased LGD estimates. Our sample selection is consistent with the improvements suggested by Gürtler and Hibbeln (2013), and hence, we do not expect similar concerns in our study. Most importantly, their results suggest that accounting for the different points in time during the life of a lease when information becomes available is essential to achieve unbiased LGD forecasts respective estimations. We carefully follow this view in our paper by distinguishing between models built conditionally on default and unconditional models. Moreover, the inputs used in these models may differ with the bank's purpose of evaluating data on LGD, and they are therefore of major economic relevance: On the one hand, if banks aim at forecasting LGDs for borrowers applying for new leases, only information for groups (1) and (2) is ascertainable. On the other, estimating LGDs using the entire amount of information (groups (1)–(3)) is often carried out in connection with backtesting and ex post model validation.

We apply the bank's definition of default which, is consistent with the regulatory framework; that is a contract is classified as defaulted if a lessee has become insolvent or is overdue with his payments on the underlying contract.[4] When validating our data, we

[4] Note that a default is not assigned to a lessee but to a contract. The reasoning behind this strategy is that lessees can sign multiple contracts with the bank and a default on one (or more) contract(s) does not necessarily imply a default on other contracts provided that the lessee still meets the payments for those contracts.

drop five contracts from the initial sample because of a negative exposure at default; thus, the final sample comprises 1,184 defaulted contracts.

For the defaulted contracts, we calculate the LGD on a contractual level as one minus the recovery rate using a simplified work-out approach. We define recovery rates as the cash inflows collected by the bank after a contract has defaulted in proportion to the exposure at default. Due to the availability of detailed information on the cash inflows, we distinguish between liquidation proceeds of the leasing item on secondary markets and other collected payments during the work-out process. Hence, the recovery rate and LGD, respectively, are defined as

$$RR_i := \frac{\text{Liquidation proceeds}_i + \text{Other payments}_i}{\text{Exposure at Default}_i} \quad \text{and} \quad LGD_i = 1 - RR_i. \tag{III.1}$$

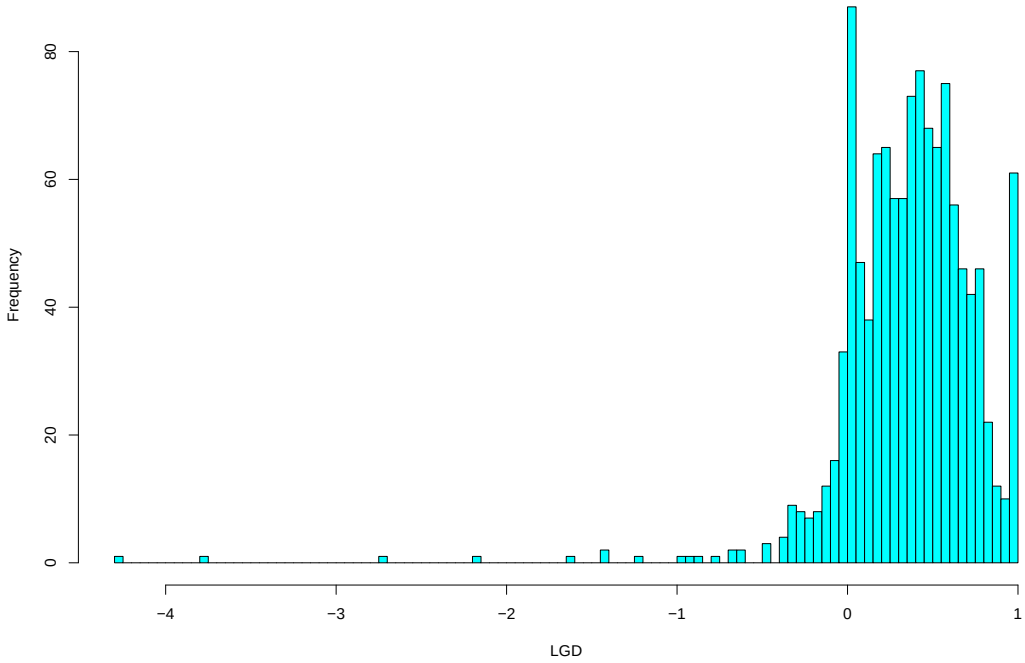
The LGD is calculated without considering the time value of liquidation proceeds and other payments, since we were not provided with detailed information on the distribution of the intermediate cash flows collected by the bank between a contract's default and the end of the work-out. However, using undiscounted LGDs instead, does not systematically bias our dataset, as deviations between time value-adjusted and undiscounted LGDs are found to be negligible (Franks et al. (2004) and De Laurentis and Riani (2005)).

Our dataset enables us to feed our methods with a great variety of variables that have been found to potentially influence LGD to analyze predictors for accurate LGD forecasts and estimations. Table 1 provides a comprehensive summary of all variables used in our analyses.

Consistent with other studies on LGDs, contrary to loans, LGDs are not restricted to the unit interval. As indicated by Figure 1, LGDs are not symmetrically distributed, but rather left skewed. In our sample, the lowest LGD equals -427.4%, whereas the mean LGD equals 36.4%. On average, LGDs of leases are lower than their counterparts usually reported for loans. Our dataset contains 111 contracts with a negative LGD, i.e., a recovery rate larger than, which can be considered as a financial gain for the lessor. Excluding those contracts however, would result in an average LGD of 43.8%, which comes fairly close to both the LGD values typically reported for loans and the LGD of 45% prescribed in the Basel Foundation IRBA for unsecured receivables. Moreover, the LGD peaks at values of around 0%, 40%, and 100%. This characteristic seems to be specific for leasing portfolios, and a similar pattern has also been found by Hartmann-Wendels et al. (2014) for two of three leasing portfolios under consideration. In contrast, the empirical literature on loans

has often found the LGD to be bimodally distributed, with modal values close to either 0 or 1 (e.g., Dermine and Neto de Carvalho (2006) and Caselli et al. (2008)).

Figure 1: LGD for 1,184 defaulted leasing contracts.



NB: LGDs of defaulted contracts are separated in 5% buckets.

Table 1: Input variables for LGD forecasts and estimations.

Type	Variable	Description	Info.	Source
Contract Characteristics	*SIZE_ORI*	Amount of the leasing contract when it was written (log(Euro))	Ex ante	Bank
	D_COLL	Dummy for provision of any type of collateral (1 = yes/0 = no)	Ex ante	Bank + CS
	DUR_CONTRACT	Contractually agreed duration of contract i (log(days))	Ex ante	Bank
	SPEC_PAYMENT	Advance payments (special lease payments) by lessee (log(Euro))	Ex ante	Bank
	D_OBJMULT	Dummy for contracts that comprise multiple objects	Ex ante	Bank
	D_FA	Dummy for full amortization contracts (1 = yes/0 = no)	Ex ante	Bank
	D_PA	Dummy for partial amortization contracts (1 = yes/0 = no)	Ex ante	Bank
	D_HP	Dummy for hire-purchase contracts (1 = yes/0 = no)	Ex ante	Bank
	D_EQUIP	Dummy for items classified as factory and office equipment (1 = yes/0 = no)	Ex ante	Bank + CS
	D_ELEC	Dummy for items classified as electronic devices (1 = yes/0 = no)	Ex ante	Bank + CS
	D_VEHIC	Dummy for items classified as vehicle or car accessories (1 = yes/0 = no)	Ex ante	Bank + CS
	D_MACH	Dummy for items classified as machinery (1 = yes/0 = no)	Ex ante	Bank + CS
Lessee Characteristics	*D_LIMITED*	Dummy for limited liability of the lessee (1 = yes/0 = no)	Ex ante	Bank + CS
	RAT_INT_BOND	IRR minus yield of the 10-year German government bond (%)	Ex ante	Bank + IMF
	DISTANCE	Linear distance between the lessee and the closest branch of the bank (log(km))	Ex ante	Bank + C
	D_SERV	Dummy for lessees operating in the services industry (1 = yes/0 = no)	Ex ante	Bank + CS
	D_CONST	Dummy for lessees operating in the construction industry (1 = yes/0 = no)	Ex ante	Bank + CS
	D_TRADE	Dummy for lessees operating in the trade industry (1 = yes/0 = no)	Ex ante	Bank + CS
	D_MANU	Dummy for lessees operating in the manufacturing industry (1 = yes/0 = no)	Ex ante	Bank + CS
	D_MOVED	Dummy for lessees that moved during the life of the lease (1 = yes/0 = no)	Ex post	Bank + C
Default Charac.	*DUR_WORK*	Duration of the work-out process (log(days))	Ex post	Bank + C
	SIZE_DEF	Size of the leasing contract at default (log(Euro))	Ex post	Bank + C
	PROP_PAID	Outstanding exposure at default in relation to the amount of the leasing contract (%)	Ex post	Bank + C
Macro	*GDP_INI*	Quarterly GDP growth rate when leasing contract was written (%)	Ex ante	DESTATIS
	INSOL_DEF	Firm's quarterly insolvency rate in Germany one quarter prior to default (%)	Ex post	DESTATIS

NB: The table specifies the variables used in our models to forecast and estimate LGDs. We logarithmize *SIZE_ORI*, *DUR_CONTRACT*, *SPEC_PAYMENT*, *DISTANCE*, *DUR_WORK*, and *SIZE_DEF* due to skewness. "Info." specifies whether a variable is observable when the contract is concluded ("ex ante") or when the work-out is completed ("ex post"). "C" or "CS" means that we derived the respective variables from data in the bank's leasing portfolio using own calculations or employing a classification scheme, respectively. "DESTATIS" refers to the Federal Statistical Office of Germany and IMF denotes the International Monetary Fund. IRR is the contract's internal rate of return and GDP denotes the Gross Domestic Product.

IV Forecasting and Estimation Methods

IV.1 Parametric methods

As outlined above, we compare forecasting and estimation results for different parametric and non-parametric methods. Specifically, we apply OLS and Tobit regressions as the parametric estimation techniques. We use the historical average of LGDs as a benchmark for all methods. To take into account that LGDs for leasing contracts are bounded at 1 in the higher domain (if work-out costs are not incorporated) but unbounded in the lower domain, we use the Tobit regression model. If β and x_i denote the vector of coefficients, respectively, the vector of explanatory variables for observation i, ϵ_i the error term, and assuming that the latent variable for LGD is given by

$$LGD_i^* = \beta' \cdot x_i + \epsilon_i, \quad \text{and} \quad \epsilon_i | x_i \sim N(0, \sigma^2), \tag{IV.1}$$

then

$$LGD_i = \begin{cases} LGD_i^* & -\infty < LGD_i^* < 1, \\ 1 & LGD_i^* \geq 1. \end{cases} \tag{IV.2}$$

The Tobit model is estimated using maximum likelihood methods.

Running linear regressions assumes that the distribution of the error term is (approximately) normal distributed. However, as can be inferred from the LGD distribution in Figure 1, LGDs in our dataset are not normally distributed; thus, linear regression techniques applied to predict LGD violate this assumption. Hence, predicting LGD by means of linear regression methods is not without problems from a statistical perspective (Zhang and Thomas (2012) and Hartmann-Wendels et al. (2014)). As LGD distributions of leasing portfolios are commonly multi-modal, often cannot be described with a single distribution, and vary remarkably depending on the dataset, applying methods that ex ante do not make any assumptions concerning the distribution of the underlying data or the functional relationship with the explanatory variables could be more suitable for obtaining reliable LGD predictions. Therefore, we implement the three non-parametric methods described below to forecast and estimate the LGD. A further advantage of non-parametric methods is that they are often better at coping with non-linear and more complex relationships.

IV.2 Regression Trees and Random Forests

The use of classification and regression trees (CART) is first suggested by Breiman et al. (1984). Regression Trees (RT) represent a non-linear, non-parametric method in which a tree structure comprising multiple logical if-then conditions is built. RTs are constructed by sequentially partitioning a dataset along values of the explanatory variables to produce partitions that are as little heterogeneous as possible in terms of the dependent variable. The heterogeneity of a partition $\mathcal{A}$, $\mathcal{A} \subseteq \{1, 2, ..., n\}$, where n is the number of observations, is given by an impurity measure $\mathcal{I}$ that maps a higher variation of dependent variable values to larger numbers. In the case of metric variables, the impurity of a partition $\mathcal{A}$ in this study is given by the sum of squared errors (SSE), defined as

$$\mathcal{I} := \text{SSE}(\mathcal{A}) = \sum_{i \in \mathcal{A}} \left(LGD_i - \overline{LGD_{\mathcal{A}}}\right)^2, \tag{IV.3}$$

where $\overline{LGD_{\mathcal{A}}}$ describes the mean of the dependent variable in partition $\mathcal{A}$. In each node of the tree, a split θ is chosen such that the two resulting partitions, i.e., the left (l) and the right (r) tree node $\mathcal{A}_l^\theta \subset \mathcal{A}$ and $\mathcal{A}_r^\theta \subset \mathcal{A}$, $\mathcal{A}_l^\theta \cap \mathcal{A}_r^\theta = \emptyset$, have a lower SSE combined compared to the original partition. The split can be chosen from all possible binary splits of the explanatory variables (e.g., $SIZE\_ORI < 10,000$ or $RAT\_INT\_BOND \leq 12.4$). When there is more than one possible split, the one yielding to the highest decrease in SSE is chosen; i.e., each split θ satisfies

$$\max_\theta \;\; \mathcal{I}(\mathcal{A}) - \left(\mathcal{I}(\mathcal{A}_l^\theta) + \mathcal{I}(\mathcal{A}_r^\theta)\right) = \max_\theta \;\; \text{SSE}(\mathcal{A}) - \left(\text{SSE}(\mathcal{A}_l^\theta) + \text{SSE}(\mathcal{A}_r^\theta)\right). \tag{IV.4}$$

The algorithm continues to split the dataset until a predefined stop criterion is fulfilled. In our case, this stop criterion is given by a minimum increase in R^2 by a factor of 0.01. The resulting tree structure consists of the split rules in the tree nodes and the partitions in the terminal nodes. When making predictions for new observations, the new observation is assigned to one terminal node using the split rules. The prediction is then given as the average dependent variable value in this tree node.

RTs are accessible for interpretation and produce results that are comparable to other more complex methods in many applications. However, a concern may be the tendency to overfit or underfit data depending on the chosen stop criterion. A fully developed tree may have a terminal node for each observation, leaving it unappropriated for out-of-sample prediction. In contrast, an extremely flat tree may have low explanatory power. However, the chosen stop criterion to construct the trees ensures that overfitting or underfitting

the data is mostly avoided. Moreover, given the discontinuity of the possible prediction values, small variations in an independent variable may cause an observation to end up in a leaf node with an extremely different predicted value. To tackle these concerns, we conduct comprehensive out-of-sample tests to ensure that the RTs are adequately fitted to the data, as well as to mitigate large prediction errors due to the discontinuity in the predictions.

As enhancements to RTs and to cope with the potential overfitting, Breiman (1996) suggests the use of ensemble methods. In particular, he proposes random forests (RF), which represent an ensemble method of building multiple trees and using the mean prediction aggregating over all trees as the prediction of the RF (Breiman (2001)). The rationale for this is that aggregating the prediction of many different trees mitigates the problem of overfitting. To produce a variety of dissimilar trees, a random sub-sample of the dataset is selected to build each tree. The trees are fully developed with no stop criterion. To produce even more diverse trees, Breiman (2001) introduces random feature selection, which reduces the variables available for choosing a split in each node to a randomly selected sub-sample of all explanatory variables. The RF is easy to calibrate but produces qualitatively good results in many situations.

IV.3 Artificial neural networks

An artificial neural network (ANN) is a non-linear, non-parametric modeling method that can be applied extremely flexibly. ANNs allow re-creation of almost every complex relationship without any assumptions about the true functional relationships or distribution of the dataset. Essentially, the idea of the neural network is to extract linear combinations of the explanatory variables and then model the target variable(s) as a non-linear function of these explanatory variables. ANNs consist of a certain number of nodes arranged in different (hidden) layers.[5] Each node in a layer is potentially connected to all the nodes in the subsequent layer. The strengths of the connections between the nodes may differ across the network, and these variations are represented by an adequate set of weights. The unknown weights reflect the importance of the single inputs for each node and are learned by the network during training.

[5] Usually, only the output of the final layer (node) is visible outside the network; consequently, this layer (node) is often called the output layer (node), whereas the layers (nodes) inside the network are referred to as hidden layers (nodes).

To construct the ANN, the input information enters the network and then passes forward through each subsequent layer until it reaches the output layer. ANNs passing on information in only one direction are referred to as feedforward networks.[6] When constructing feedforward ANNs from left to right, each node in each layer uses the weighted sum of all nodes from its left side as input and generates an outcome, which is used as the new input values in the next layer on its right side. The transformation from a node's aggregated input into the output is carried out via an activation function, which essentially ensures a smooth transition of changes in the input values.

As LGDs are commonly known to be difficult to predict, especially using parametric methods, the high flexibility of ANNs seems to be suitable for detecting additional patterns in the data and improving the prediction performance. We therefore fit a two-layer feedforward neural network including one hidden layer, as described by MacKay (1992), Forese and Hagan (1997), and Rodriguez and Gianola (2016). The model is given by

$$LGD_i = \sum_{k=1}^{s} w_k \cdot g_k \left(b_k + \sum_{j=1}^{p} x_{ij} \cdot \alpha_{kj} \right) + \epsilon_i, \quad i = 1, \ldots, n, \tag{IV.5}$$

where α_{kj} represents the weight of the j-th input to the network; x_{ij} denotes the included information of contract i; s and p are the number of the nodes and input variables, respectively; and b_k is the bias of the k-th node.[7] Moreover, w_k is the weight of the k-th node and determines the with w_k weighted influence of this node on the model outcome, i.e., the LGD.

IV.4 Error measures, out-of-sample testing, and model stability

To compare the quality-of-fit of the in-sample and out-of-sample LGD predictions, we apply two different error measures that have been widely discussed in the literature on

[6] In contrast to other types of ANNs, such as recurrent neural networks, there is no feedback between the layers. However, feedforward ANNs with only one hidden layer and a sufficient number of nodes in the layers can fit any finite mapping problem.

[7] The bias is an additional intercept added to each hidden node, and it is connected to all the nodes in the next layer and none in the previous one. It stores the constant input one, which allows the activation function to be shifted to the left or right. Hence, the bias can be interpreted as the intercept of the linear combinations of the inputs.

forecasting and estimation. These are the mean absolute error (MAE) and the root-mean-square error (RMSE).[8] The MAE is given by

$$\text{MAE} := \frac{1}{|\mathcal{F}|} \sum_{i \in \mathcal{F}} \left| LGD_i - \widehat{LGD}_i \right|, \tag{IV.6}$$

and the RMSE, which is conceptually related to the SSE used in the RT, is given by

$$\text{RMSE} := \sqrt{\frac{1}{|\mathcal{F}|} \sum_{i \in \mathcal{F}} \left(LGD_i - \widehat{LGD}_i \right)^2}, \tag{IV.7}$$

where $\widehat{LGD}_i$ denotes the predicted LGD for observation i, and $\mathcal{F} \subseteq \{1, 2, ..., n\}$ is the set of in-sample and out-of-sample observations, respectively.

For the interpretation of the two measures, a lower value indicates a better prediction accuracy of the method; hence, lower values of the error measures are desirable. With respect to the handling of outliers, the RMSE penalizes outliers more than the MAE does as it increases quadratically with the deviations between the actual observed LGD and its predicted value. Hence, we expect that methods predicting relatively constant differences between the observed and predicted LGD across all observations will lead to a lower RMSE but probably higher values for the MAE. In contrast, methods predicting a higher proportion of outliers should perform worse on RMSE.

Finally, we compute a method's R^2 as a common measure of the quality-of-fit of regression methods, which is given by

$$R^2 := 1 - \frac{\sum_{i \in \mathcal{F}} \left(LGD_i - \widehat{LGD}_i \right)^2}{\sum_{i \in \mathcal{F}} \left(LGD_i - \overline{LGD} \right)^2}, \qquad \overline{LGD} := \frac{1}{|\mathcal{F}|} \sum_{i \in \mathcal{F}} LGD_i. \tag{IV.8}$$

The R^2 quantifies the variance that can be explained by the respective method. Higher values indicate greater explanatory power. By construction, R^2 can lie outside the unit interval when applied to methods other than OLS regression. However, the traditional R^2 defined above is commonly used in research and practice and can be computed for all our methods.[9]

---

[8] See, for instance, Zhang and Thomas (2012) for an application of LGD prediction of unsecured consumer loans and credit cards.

[9] See Loterman et al. (2012) for a discussion of the applicability of our traditional R^2 to non-linear methods and possible generalizations.

In-sample predictions are derived from the whole dataset. To produce robust out-of-sample results, we implement a 500-times bootstrap cross-validation as follows: In the first step, the dataset is randomly split into a training dataset consisting of 75% of all observations and a test dataset containing the remaining 25% of observations 500 times. In the second step, all methods are first calibrated on the training datasets. Subsequently, the predictions are conducted on the test datasets. The out-of-sample errors for all measures are computed as the respective mean error measures over all 500 iterations.

In-sample predictions are not necessarily a good indicator of reliable out-of-sample predictions. This might be particularly true for unstable methods, i.e., methods where the in-sample and out-of-sample prediction accuracies deviates significantly. To evaluate the stability of our methods, as well as the validity of our in-sample and out-of-sample predictions, we compute the Janus quotient, which is given by

$$\text{Janus quotient} := \sqrt{\frac{\frac{1}{m}\sum_{i=1}^{m}\left(LGD_i - \widehat{LGD}_{i,\,\text{out}}\right)^2}{\frac{1}{\ell}\sum_{i=1}^{\ell}\left(LGD_i - \widehat{LGD}_{i,\,\text{in}}\right)^2}}, \tag{IV.9}$$

where $\widehat{LGD}_{i,\,\text{in}}$ is the in-sample prediction and $\widehat{LGD}_{i,\,\text{out}}$ represents the out-of-sample prediction of the LGD of contract i, while m and ℓ are the number of observations in the in-sample and out-of-sample datasets, respectively (Gadd and Wold (1964)). Along these lines, stable methods should have virtually the same prediction errors in-sample and out-of-sample. According to Equation (IV.9), stable methods are therefore characterized by a Janus quotient close to one. A Janus quotient greater (less) than one indicates that the out-of-sample error is greater (less) than the in-sample error.

IV.5 Model calibration

In this subsection, we provide detailed insights into how we tuned the models' parametrization to ensure the comparability of our results.[10] We test several combinations of each methods' main parameters and choose the parametrization with the overall best out-of-sample performance on MAE and RMSE in both the forecasting and estimation cases for all analyses. By doing so, we aim at analyzing only methods throughout this pa-

[10] All unreported calibration results are of course available from the authors upon request.

per calibrated with comparable parameter settings.[11] Thus, our results are almost surely not boosted by comparing unreasonable parameter settings or other discretionary optimizations. Even though the results differ in magnitude for other parameter settings, the overall order of the methods in terms of their prediction accuracy remains unchanged for other comparable parameter values. In the same vain, the methods virtually identify the same variables as important with respect to their influence on the prediction accuracy irrespective of the used parametrization.

For the non-parametric regression methods, namely OLS and Tobit regressions, optimization possibilities are naturally limited and, more importantly, there is no need to optimize any parameter choices. Nevertheless, we re-estimate both methods using standard errors clustered on a contractual level instead of using unclustered standard errors. Our results are robust to both options.

In the case of RTs, the two most important parameters are the minimum number of observations in any terminal leaf node and the choice of the stop criterion defining when the algorithm stops splitting the tree. As regards the latter, the stop criterion is given by a minimum increase in R^2 by a predefined value in this paper. We examine the in- and out-of-sample model performance for the values 0.5, 0.1, 0.05, 0.01, 0.005, 0.001, 0.0005, and 0.0001. As for the former parameter, we start with a minimum number of one observation in any leaf node and replace it gradually with 3, 5, 7, 10, 15, and 20. Defining lower limits for the number of observations in the leaf nodes is crucial to cope with overfitting. Otherwise, a RT based on n observations and containing $n-1$ splits will lead to perfect in-sample predictions with each observation being its own prediction but rather poor out-of-sample results. To make this explicit, the worst out-of-sample performance in the forecasting as well as estimation case in terms of MAE and RMSE is indeed obtained using a minimum increase of R^2 of 0.001 with one observation only in the terminal leaf node. Furthermore, the resulting parameter combinations show remarkable variations in terms of their prediction accuracy making a careful parameter selection indispensable for RTs. Our analyses are based on a minimum increase in R^2 of 0.01 and the minimum number of observations in the leaf nodes is set to 10.

[11] Although we mainly consider the out-of-sample results for setting the final parametrization, we monitor the methods' in-sample performance as well. Comparing the in- and out-of-sample results clearly indicates that overfitting is a serious concern for non-parametric methods since the non-parametric methods performing best in-sample show commonly the worst out-of-sample performance. Hence, mainly focusing on the out-of-sample errors while still monitoring the in-sample results ascertain that working with methods overfitting the data in-sample is most likely avoided.

With respect to the RFs, we vary the number of variables randomly sampled as candidates at each split from 1 to 8 in steps of 1 and the number of trees in the forest from 50 to 250 in steps of 50. A thorough choice of the number of trees to grow in the forest is important in order to balance prediction accuracy and computational complexity. On the one hand, if the forest encompasses too few trees while the number of observations is rather large, some observations may not (or just seldom) be considered for predictions because the RF selects a subset of all observations and variables (which is also varied during the model calibration) to grow a tree which in turn will result in a significant loss of predictive power. On the other, simply assuming a sufficient large number of trees does not necessarily improve the prediction performance but increases computational costs rapidly. Interestingly, for the forecasting and estimation case, the out-of-sample accuracy in terms of MAE and RMSE is rather robust to the choice of the number of trees and the amount of randomly sampled candidates at each split. However, the R^2 exhibits a substantial variation across the considered parameter combinations. Thus, we try to identify a threshold beyond which no significant improve in the predictive power can be expected, whereas computational complexity would soar. Ultimately, we end up with 150 trees to compose the forest, each tree randomly sampling 3 variables at each split.

Finally, for the ANNs, the determination of the initial weights is carried out as described in Nguyen and Widrow (1990). The activation function is specified by the sigmoid function $g_k(x) = \frac{\exp(2x)-1}{\exp(2x)+1}$. The error ϵ_i in Equation (IV.5) depends on the number of nodes and can be made sufficiently small for in-sample predictions by adding hidden nodes. In this case however, the method will overfit the data and the error in the out-of-sample predictions increases quickly. As the choice of the optimal number of hidden nodes is data dependent and there is no common rule for determining it, we follow an approach similar to that of Qi and Zhao (2011) and optimize our network with several numbers of hidden nodes to obtain the optimal out-of-sample prediction; we use this configuration in all in-sample analyses to ensure the comparability of our results.

We test a different number of nodes, ranging from one to 150.[12] We keep the number of layers fixed, since one hidden layer is generally sufficient to reflect the underlying patterns in the data moderately well. For the forecasting and estimation cases, the in-sample and out-of-sample accuracies suggest a contrasting picture for an increasing number of neurons in the hidden layer. Starting with a small number of nodes in the out-of-sample tests, a small increase of the number of nodes causes the MAE and RMSE to deteriorate

[12] In detail, we construct several networks with different numbers of nodes ranging from 1 to 2, 3, 4, 5, 10, 15, 20, 30, and so on, up to 150.

significantly. Moreover, the neural network is very sensitive to the parametrization and starts to memorize the patterns in the data already for a small number of nodes. We find that the optimal number of nodes in our dataset is most likely three.

We use a Bayesian regularization algorithm to set the optimal weights of the network. The algorithm minimizes the weighted sum of both the sum of the squared validation errors and the sum of squares of the weights and biases. The algorithm randomly divides the training dataset into calibration and validation sets. The calibration set is used to compute and update the network weights and biases. These parameters are then validated on the validation set, and the corresponding validation error is monitored. This process is repeated until the algorithm terminates. The validation error generally decreases during the initial phase of training and typically increases once the network starts to overfit the data. The algorithm stops training before overfitting the data, and the weights and biases at the minimum of the validation error are returned. Moreover, many model parameters are penalized since the sum of the squared model parameters clearly increases with respect to the number of nodes.

V Analysis

V.1 In-sample results

In the first step, the analysis of the empirical findings of our paper concentrates on the in-sample forecasting performance of the parametric and non-parametric methods. Thus, all methods are only trained with the available information when the contract was concluded. The results are shown in columns (1) to (3) of Table 2. In terms of the error between the realized LGD and its forecast, the distinctly smaller in-sample errors reported for the MAE and RMSE clearly indicate that the RF outperforms the other methods for both error measures. This superior forecasting performance of RFs can be derived from their construction. As some of the trees included in the RF already contain the observations that will be used for the in-sample forecasts, these forecasts will be predicted nearly perfectly, thereby improving the in-sample forecasting accuracy.[13] Likewise, we observe accurate in-sample forecasts for the ANN yielding to in-sample errors which are well ahead of the other methods. While the forecasting accuracy of the RTs and the parametric regression methods differs only slightly for the MAE, the RTs have remarkably lower forecasting errors as measured by the RMSE. This finding indicates that OLS and Tobit regressions are more prone to predicting outliers, which are more strongly penalized by the RMSE.

Table 2: In-sample quality-of-fit measures.

	(1)	(2)	(3)	(4)	(5)	(6)
Method	MAE_{fore}	$\text{RMSE}_{\text{fore}}$	R^2_{fore}	MAE_{est}	RMSE_{est}	R^2_{est}
Hist. avg.	0.2739	0.3984	0.0000	0.2739	0.3984	0.0000
OLS	0.2672	0.3903	0.0403	0.2519	0.3644	0.1632
Tobit	0.2674	0.3904	0.0395	0.2525	0.3645	0.1627
RT	0.2673	0.3818	0.0816	0.2461	0.3372	0.2837
RF	0.1484	0.2248	0.6816	0.1158	0.1711	0.8155
ANN	0.2507	0.3779	0.1000	0.2254	0.3308	0.3103

NB: The error measures, namely the MAE, RMSE, and R^2, are subscribed with "fore" and "est" to indicate forecasting and estimation results, respectively.

[13] It should be noted that the non-parametric methods can reproduce the in-sample data almost perfectly. However, this substantially enhances overfitting the data, which in turn increases the out-of-sample forecasting error. Thus, a high in-sample prediction accuracy may be misleading in terms of the out-of-sample prediction performance. As this paper is strongly devoted to the assessment of the out-of-sample results of different prediction methods, we refrain from boosting the in-sample prediction accuracy.

Concerning the variance that can be explained by the in-sample forecasts as measured by the R^2, we find a superior explanatory performance of the non-parametric methods. The rather high R^2 values for RTs, ANNs, and particularly RFs, underline their ability to recreate the structure of the dataset to a significant extent.[14] Overall, the results of the in-sample forecasting analysis provide initial evidence that explicitly modeling the LGD by means of parametric and non-parametric methods is more advantageous than simply relying on historical averages, which show the worst performance across all quality-of-fit measures.

In the estimation part of our analysis, we model the LGD with all information known after the work-out process is completed. This dataset contains most of the information considered in the forecasting analyses and the ex-post variables presented in Table 1.[15]

Analyzing the in-sample estimation results presented in columns (4) to (6) of Table 2, we note that the prediction errors are predominantly smaller than in the forecasting case. Hence, the information becoming visible after the work-out process is completed adds valuable input for all methods, leading to considerably better in-sample LGD predictions. Concerning the quality-of-fit of our methods, the RFs again give the most accurate LGD estimations clearly away from the other methods. The ANN exhibits the second-best performance on all error measures. The RT performs slightly worse than the ANN, but it shows a superior estimation performance to that of the parametric regression techniques. More specifically, RTs yield lower RMSE values suggesting that RTs give less volatile estimations. We attribute this to the implemented splitting method used to construct the RT, which mostly prevents the prediction of outliers, since those values are isolated by the observed exogenous variables.

Concerning the estimation performance of the parametric regression methods, we observe extremely similar results for OLS and Tobit regressions for all error measures in both the forecasts and estimations. Thus, the theoretical advantage of Tobit regressions in terms of explicitly considering the restricted interval of leasing LGDs does not significantly affect the in-sample prediction accuracy. Even if OLS estimations may be inconsistent in the case of LGDs, this tends to affect their prediction performance only marginally. Thus, OLS regression seems to be an easily implementable, understandable benchmark method for (parametric) LGD predictions.

---

[14] Note that the R^2 of the historical average for in-sample predictions equals zero by definition.

[15] The initial leasing size ($SIZE\_ORI$) is replaced by the exposure at default ($SIZE\_DEF$) and quarterly GDP growth (GDP_INI) is substituted with quarterly insolvency rates ($INSOL\_DEF$).

In summary, and as in the forecasting case, the non-parametric RF, ANN, and to some extent, RT, outperform the parametric methods, suggesting that their ability to recreate the structure of the dataset is preserved for an extended set of information. This is emphasized by the comparably high values of R^2 for the non-parametric methods. Thus, we conclude that the flexibly applicable non-parametric methods are generally more suited to achieve better in-sample estimations of leasing LGDs.

V.2 Out-of-sample results

The out-of-sample forecasting results are presented in columns (1) to (3) of Table 3. It is remarkable that RFs again provide the most accurate forecasts for both error measures and still exhibits a notable distance from the other methods. Among these approaches, the RT obtains the worst performance, and it is even outperformed by the historical average. The forecasting accuracy of OLS and Tobit regressions is better than that of the historical average for the MAE and RMSE. These findings are confirmed by the t-values from the pairwise t-tests of differences in the mean values of the MAE and RMSE for each pair of methods reported in Panel A of Table 5 in the Appendix. ANNs perform similarly to the parametric regression methods for the MAE and RMSE. For the MAE, all methods but the RT show significantly superior forecasting performance compared to the historical average, while for the RMSE, this only remains valid for the RFs. The variations in the results across the different error measures suggest that a sound evaluation of the methods' forecasting performance should be based on diverse error measures.

Table 3: Out-of-sample quality-of-fit measures.

	(1)	(2)	(3)	(4)	(5)	(6)
Method	MAE_{fore}	$\text{RMSE}_{\text{fore}}$	R^2_{fore}	MAE_{est}	RMSE_{est}	R^2_{est}
Hist. avg.	0.2742	0.3933	-0.0046	0.2738	0.3942	-0.0051
OLS	0.2724	0.3920	0.0018	0.2574	0.3698	0.1116
Tobit	0.2678	0.3922	0.0022	0.2502	0.3712	0.1090
RT	0.2755	0.3986	-0.0381	0.2540	0.3642	0.1187
RF	0.2538	0.3788	0.0680	0.2246	0.3270	0.2962
ANN	0.2698	0.3935	-0.0049	0.2483	0.3638	0.1409

NB: The error measures, namely the MAE, RMSE, and R^2, are subscribed with "fore" and "est" to indicate forecasting and estimation results, respectively.

Overall, we conclude that forecasting leasing LGDs is rather difficult if the valuation is tested for unknown contracts, and equally importantly, if the methods can only rely on

a restricted set of information known when a contract was concluded. In addition, the non-parametric methods seem to be slightly superior to the parametric ones, especially the extremely flexible RF.

Including additional information that became visible after the work-out was concluded leads to the estimation results displayed in columns (4) to (6) of Table 3. As expected, all methods obtain remarkably better out-of-sample estimations than in the forecasting case, underlining that the new information becoming available after the work-out improves prediction accuracy for parametric and non-parametric methods. As in the forecasting case, the non-parametric RFs obtain the best estimation accuracy. More specifically, the RF outperforms the other methods in terms of the MAE, RMSE, and R^2, while the ANN shows the next-best results. Thus, our findings provide strong evidence that the non-parametric RFs and ANNs yield the most accurate out-of-sample predictions of leasing LGDs.

In contrast to the forecasting case, the pairwise t-tests of differences in the mean values of the MAE and RMSE in Panel B of Table 5 clearly indicate that the improvements in estimation accuracy of all methods compared to the historical average are statistically significant (see Appendix). Again, RFs give significantly better LGD estimations than any other method. The significant differences in the prediction accuracy across our methods underline that predicting leasing LGD requires a thorough assessment of the methods by banks and supervisors to obtain reliable results.

We also compare the out-of-sample forecasting and estimation results presented in Table 3 with the in-sample results discussed in the previous section. We generally observe an increase in all error measures; this is much more pronounced for the best-performing in-sample RFs and ANNs than for the parametric regression methods and the RTs. This supports the widespread finding that out-of-sample predictions are usually less precise than their in-sample counterparts. The explanatory power of the methods as measured by the R^2 decreases remarkably across all methods and even turns negative for the RT, the ANN, and the historical average in the forecasting case and for the historical average in the estimation case.

In contrast to the in-sample estimation results, Tobit regressions show significantly better forecasting and estimation performance compared to OLS regressions for the MAE. However, they perform worse – albeit not significantly – for the RMSE, as can be seen from the t-tests in Table 5 in the Appendix. This may suggest that the theoretically more

adequate modeling of leasing LGDs by means of Tobit regressions could be particularly important for out-of-sample predictions. However, the deviations in the RMSE between Tobit and OLS regressions are rather small, suggesting that the latter are still a reasonable benchmark method.

The overall prediction errors of the LGD estimations are predominantly smaller than in the forecasting case for both the in- and out-of-sample findings. Therefore, the added ex post variables most likely contain useful information to improve the prediction accuracy. The performance increase should serve as an incentive for banks to explicitly model the LGDs of their leasing portfolios and collect comprehensive information about their portfolios including defaulted and/or recovered leases. Our results reveal that non-parametric methods, except the RTs, show superior performance to parametric approaches because they are rather able to exploit the multifarious inputs of our dataset.

In the next step, we assess the variation of the forecasting and estimation errors across the 500 out-of-sample cross-validation iterations by computing the standard deviations of the MAE and RMSE for each method across all out-of-sample iterations. In the same vein, this approach is helpful in generating reliable results concerning the robustness of the methods across the iterations. Table 4 reports the results. Except for the historical average, the standard deviations for the estimation error measures are smaller than for the forecasting errors, suggesting that including additional variables improves the methods' stability in matters of outliers in the estimation of LGDs. Comparing the standard deviations across the methods, it turns out that the parametric regression methods tend to show larger variations in their prediction errors for the RMSE. This finding underlines that OLS and particularly Tobit regressions are slightly less robust in predicting leasing LGDs than the other methods are. In terms of the RMSE, the tree-based methods yield the lowest variation across the iterations, suggesting that these methods are rather robust since it is unlikely that their average prediction accuracies are biased by extreme outliers.

Finally, comparing the in- and out-of-sample quality-of-fit measures for LGD predictions enables us to investigate whether in-sample predictions are a reliable indication for the out-of-sample performance. As can be seen from the Janus quotients given in Table 4, the out-of-sample errors are generally greater than the in-sample errors, as indicated by values of the Janus quotient greater than one. Likewise, the Janus quotient calculated for LGD estimations increases for all but the parametric regression methods compared to LGD forecasting since the deviations between in- and out-of-sample errors rise more for LGD estimations than for forecasts. Evaluating the overall stability of the methods, values

Table 4: Standard deviations of the out-of-sample forecasting and estimation errors and Janus quotients.

	Forecasting		**Estimation**		**Janus quotient**	
Method	MAE	RMSE	MAE	RMSE	Forecasting	Estimation
Hist. avg.	0.0137	0.0543	0.0143	0.0553	0.9960	0.9995
OLS	0.0134	0.0540	0.0123	0.0486	1.0159	1.0281
Tobit	0.0143	0.0562	0.0134	0.0535	1.0101	1.0176
RT	0.0145	0.0504	0.0130	0.0339	1.0713	1.1084
RF	0.0138	0.0536	0.0119	0.0349	1.7193	1.9199
ANN	0.0143	0.0554	0.0127	0.0486	1.0337	1.0653

NB: Standard deviations for the out-of-sample forecasting and estimation errors are calculated with 500 out-of-sample cross-validation iterations.

around one for the OLS and Tobit regressions, as well as the ANNs and RTs, indicate that most of our methods are stable. Thus, they are suitable methods for LGD forecasting and estimation. For these methods, the deviations between the in- and out-of-sample prediction accuracies do not weaken the stability of the methods; consequently, in-sample predictions can be considered as a first indication of the out-of-sample quality-of-fit for stable methods.

Concerning the RFs, the corresponding Janus quotients are clearly above one, which is not too surprising given the considerably low in-sample performance errors. However, the RFs obtains the best forecasting and estimation performance for the MAE and RMSE both in- and out-of-sample. Thus, accurate leasing LGD estimations should rely on RFs, but the in-sample prediction accuracy may overstate the out-of-sample quality-of-fit. This may be particularly true if the flexible RF algorithm has recreated the data well, leading to comparably good in-sample results but a loss in the prediction power in the out-of-sample analysis. Overall, we find a slight tendency for non-parametric methods to be more prone to instability than parametric methods are.

V.3 Variable importance

In this section, we target the question of what features drive the results of the prediction techniques. As the complex non-linear methods outperform the linear methods in many cases, we raise the question of what information is processed by the non-parametric methods that is not processed by the linear methods. To generate a visual comparison, we construct a measure built on the idea of the permutation importance, which is com-

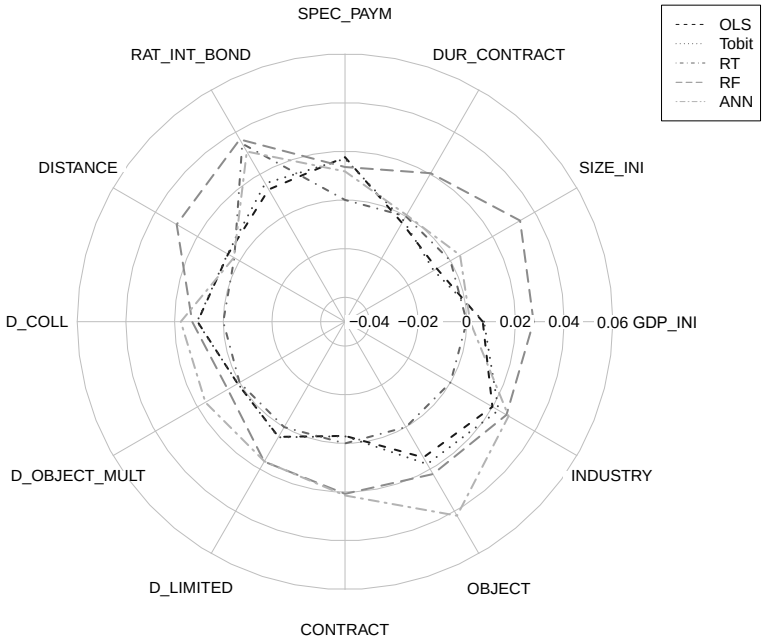
monly used in RF applications. The measure, as we use it here, is constructed by first calculating an error measure, such as the MAE, for each prediction method using the entire dataset. Second, we permute the values of each independent variable used in the prediction and calculate the new error measure for each method. We then calculate the increase or decrease in the error measure relative to the non-permuted case. For the out-of-sample prediction, we calculate the relative change in error measures after permutation on the same cross-validation training and test datasets as used in the 500 cross-validation iterations above.

The values for the variable importance as measured by changes in the MAE when permuting a single variable are plotted in Figure 2 and Figure 3.[16] The exact numbers are presented in Table 6 and Table 7 in the Appendix. There is an axis for each permuted variable. The deterioration in the error measure after the permutation is given by the position on these axes. Points from the same method are connected by vertical lines to illustrate the differences.

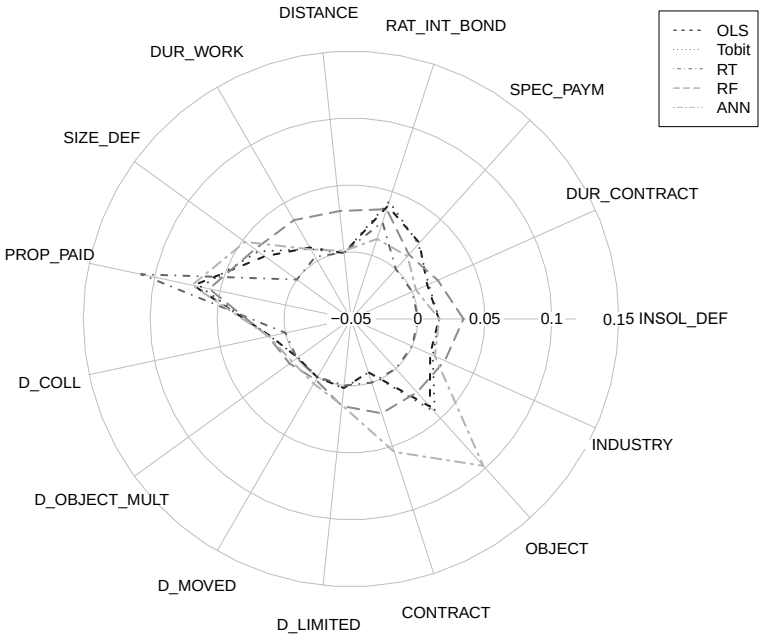
The results for the in-sample forecasts are plotted in Figure 2a. Except for the RF, the increase in error terms is generally low. This may indicate a weakness of the available characteristics in making predictions for the LGD. The highest importance measure values for the RT and the ANN are around 7% and 8%, respectively. The highest values for the regression methods are even lower, at around 2%. Starting with the values for the RT, this method almost solely relies on the internal rating (RAT_INT_BOND), where it has its only spike. All other measure values are close to or on the zero line. For the linear methods, while all values are generally low, special lease payment ($SPEC\_PAYMENT$), industry ($INDUSTRY$), asset type ($ASSET$), and collateral (D_COLL) seem to stand out slightly. In addition, the internal rating (RAT_INT_BOND), which was not significant in the regressions, stands out slightly in size. The ANN uses similar variables to the linear methods but also incorporates the contract type ($CONTRACT$), dummy for lessees with limited liability ($D\_LIMITED$), and dummy for multiple objects ($D\_OBJMULT$). Especially, the object type ($ASSET$) and internal rating (RAT_INT_BOND) have a comparably stronger effect on the importance measure. The RF generally has more importance peaks. It has particularly high values for the GDP growth (GDP_INI), initial size ($SIZE\_ORI$), internal rating (RAT_INT_BOND), distance ($DISTANCE$), and three categorical variables ($ASSET$, $INDUSTRY$, and $CONTRACT$). We stress the fact that it has some of its highest importance measure values for characteristics that

[16] The values for the RMSE are qualitatively the same and hence not reported. All unreported results can of course be made available on request.

Figure 2: Forecasting (upper panel) and estimation (lower panel) changes in the MAE for permuting the respective variables in the in-sample prediction.



(a) Forecasting.



(b) Estimation.

NB: The permutation changes in the error measures are divided by 10 for the RF and by 2 for the ANN and RT. For $ASSET$, $INDUSTRY$, and $CONTRACT$, we permute the corresponding dummy variables.

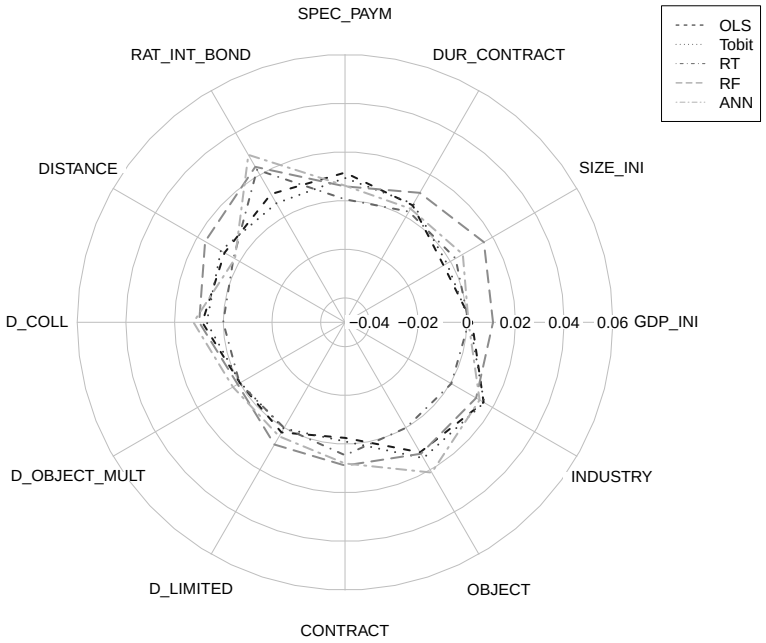
were almost omitted by the four other methods. The ANN also incorporates some more information than the regression methods.

The results for the in-sample estimation are presented in Figure 2b. It is immediately apparent that the overall level of importance measure values is considerably higher. Given the very high measure values, this is substantially driven by the newly acquired information about the proportion outstanding ($PROP\_PAID$), exposure at default ($SIZE\_DEF$), and in the case of the RF, the work-out duration (DUR_WORK). The results for the earlier available characteristics remain similar. For the case of the RT, only the outstanding proportion ($PROP\_PAID$) and internal rating (RAT_INT_BOND) are important. For the linear methods, special lease payment ($SPEC\_PAYMENT$), internal rating (RAT_INT_BOND), object type ($ASSET$), and industry type ($INDUSTRY$) contain important information. However, given the new information, the insolvency rate ($INSOL\_DEF$) and contract duration ($DUR\_CONTRACT$) become more important as well. The ANN generally resembles the linear methods on many points, but it makes less use of the special lease payment ($SPEC\_PAYMENT$) and internal rating (RAT_INT_BOND) variables and more use of contract type ($CONTRACT$). In addition to the work-out duration (DUR_WORK), the RF incorporates information about distance ($DISTANCE$) that is not relevant in the other model curves.

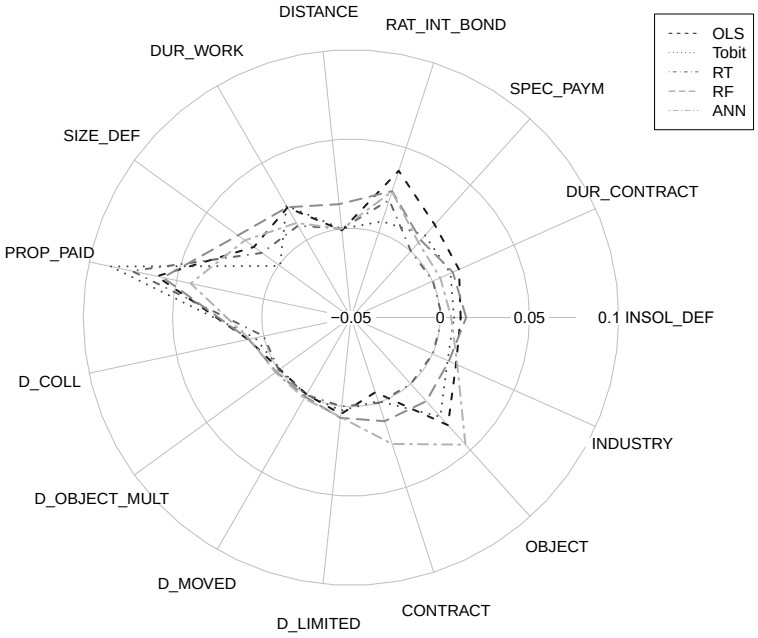
The out-of-sample forecasting results are presented in Figure 3a. When compared to the in-sample plots, the values are generally closer to the zero line for many characteristics. In the case of the ANN, some of the information that was relevant in the in-sample case ($CONTRACT$, $D\_LIMITED$, and $D\_OBJMULT$) no longer seems important. The ANN results generally resemble those of the regression methods. The RT mainly relies on the internal rating (RAT_INT_BOND). This characteristic is also especially important for the RF and ANN. Apart from the RT, all four methods use information from the asset type ($ASSET$) and industry type ($INDUSTRY$). The RF can use some of the characteristics in out-of-sample prediction that are almost neglected by the other methods ($DISTANCE$, $DUR\_CONTRACT$, $SIZE\_DEF$, and GDP_INI). This is interesting since, unlike the ANN, which could not continue to use some of its in-sample information out-of-sample, the RF can generalize patterns to new contracts.

The out-of-sample estimation results in Figure 3b appear similar to the in-sample ones. There is an accentuated spike for the proportion of the contract value that is still due at default ($PROP\_PAID$) for all five methods. The two other newly added characteristics ($SIZE\_DEF$ and DUR_WORK), internal rating (RAT_INT_BOND), and object type

Figure 3: Forecasting (upper panel) and estimation (lower panel) changes in the MAE for permuting the respective variable in the out-of-sample prediction.



(a) Forecasting.



(b) Estimation.

NB: The permutation changes in the error measures are divided by 2 for all non-parametric methods. For $ASSET$, $INDUSTRY$, and $CONTRACT$, we permute the corresponding dummy variables.

($ASSET$), are highly important across the five methods. The RT generally uses fewer characteristics. The RF and ANN use the contract type ($CONTRACT$), which is nearly neglected by the other methods. The RF further makes use of the distance ($DISTANCE$). The additional information employed by these methods is less apparent in this case. The results tend to be largely driven by the characteristics that become available after default.

To summarize the results of this section, given the especially low levels of our significance measure, prediction again appears difficult in the forecasting case, although there are some characteristics that seem relevant and carry over to new lessees. This supports previous findings that forecasting methods generally perform weakly. The internal rating (RAT_INT_BOND), object type ($ASSET$), and industry ($INDUSTRY$) may convey some relevant information here. The RF processes some information that may be omitted by the other methods, which results in better performance in the prediction accuracy. To a lesser extent, this is also true for the ANN. When comparing the forecasting and estimation plots, it is apparent that the default characteristics add much valuable information. The estimation results are mainly driven by these new characteristics. This is especially the case for the outstanding relative exposure ($PROP\_PAID$), which has far higher importance measure values than any of the other characteristics in either the in-sample or out-of-sample estimations. Concerning adequate credit risk management, as the outstanding relative exposure can be calculated at any time during the life of the lease, banks should continuously update this information to assess the (expected) loss rates based on reliable predictions consistent with the regulatory requirements for internal LGD estimates (EU Regulation No 575/213, Article 179). In addition to the three default characteristics, relevant characteristics that extend to new observations seem to be the internal rating (RAT_INT_BOND) and object type ($ASSET$). The more complex methods incorporate some additional information, but the differences from the linear methods are less apparent from the plots. Given that the estimation results are mainly driven by default characteristics, methods that do not consider the time dimension may not be adequate to some extent.

VI Conclusions

Although leasing has become increasingly important for firms in the last decade, little is known about the predictability of the LGD of leases and, even less is understood about which variables indeed improve the prediction accuracy. Using a unique dataset of 1,184 defaulted leases over the period from 2009 to 2014, we analyze the in-sample and out-of-sample performances of parametric and non-parametric methods to forecast and estimate LGDs. We compare the prediction accuracy of the parametric and non-parametric approaches by drawing on several performance measures. To the best of our knowledge, this is the first paper to study RFs in the context of leasing LGD predictions. Moreover, we contribute to addressing the question of how prediction accuracy depends on the set of available information and hence on the point in time when the LGD is predicted. We differentiate between information already known when the contract is concluded and additional information becoming available during the work-out process or the lifetime of the lease. In the past, some non-parametric methods have been frequently criticized for being impenetrable ("black-box"). We explicitly address these concerns and employ a measure of variable importance to examine which explanatory variables achieve the greatest improvements in LGD prediction accuracy in each of the methods.

Our analyses reveal that the LGD predictions of non-parametric methods – especially RFs and in most cases ANNs – are statistically significantly more accurate than those of traditional linear methods. This holds for in-sample and out-of-sample tests, even when properly controlling for overfitting. By conducting a comprehensive number of 500 bootstrapping iterations, we further support existing evidence that out-of-sample testing is crucial for obtaining comparable and reliable results regarding model performance. Against this background, banks using RFs, and to a slightly lesser extent ANNs, could improve their LGD predictions remarkably, thereby generating competitive advantages over other financial institutions. However, we acknowledge that implementing more complex methods may induce higher efforts for banks and increase model risk.

Concerning the set of information included in the predictions, we find that conditional models built on information becoming available at the time of default have much more explanatory power and generate considerably better predictions than unconditional models. In particular, the historical average of LGDs, which serves as a lower benchmark, is significantly outperformed by all considered linear and non-linear methods on all performance measures. Here, the contract's volume at the time of default and the stressed industry conditions seem to significantly influence LGDs. Hence, our findings indicate that in ad-

dition to lessee- and contract-related information, banks should also include information becoming available during the recovery process.

To analyze the importance of our explanatory variables, we compute the relative changes in the prediction errors for each method when permuting a single variable. We find remarkable differences across the variables and methods, suggesting that strong differences exist regarding their power to improve LGD forecasts and estimations. The exposure at default in relation to lease's original size, the age of the contract at the time of default, and to some extent, the lessee's internal rating seem to be the main determinants of accurate LGD predictions. We conclude that banks should carefully analyze which variables are driving their predictions to assess the credit risk of their leasing portfolios adequately and avoid working with weak or even unstable methods.

References

Acharya, V. V., S. T. Bharath, and A. Srinivasan (2007). Does industry-wide distress affect defaulted firms? Evidence from creditor recoveries. *Journal of Financial Economics 85*(3), 787–821.

Altman, E. I., B. Brady, A. Resti, and A. Sironi (2005). The Link between Default and Recovery Rates: Theory, Empirical Evidence, and Implications. *The Journal of Business 78*(6), 2203–2228.

Altman, E. I. and V. M. Kishore (1996). Almost everything you wanted to know about recoveries on defaulted bonds. *Financial Analysts Journal 52*(6), 57–64.

Araten, M., M. Jacobs Jr., and P. Varshney (2004). Measuring LGD on Commercial Loans: An 18-Year Internal Study. *The RMA Journal 86*(8), 96–103.

Asarnow, E. and D. Edwards (1995). Measuring Loss on Defaulted Bank Loans: A 24-Year Study. *Journal of Commercial Lending 77*(7), 11–23.

Baesens, B., T. Van Gestel, S. Viaene, S. Stepanova, J. Suykens, and J. Vanthienen (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society 54*(6), 627–635.

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

Bellotti, T. and J. Crook (2012). Loss given default models incorporating macroeconomic variables for credit cards. *International Journal of Forecasting 28*(1), 171–182.

Breiman, L. (1996). Bagging pedictors. *Machine Learning 24*(2), 123–140.

Breiman, L. (2001). Random forests. *Machine Learning 45*(1), 5–32.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone (1984). *Classification and Regression Trees*. Statistics/Probability Series. Belmont: Wadsworth Publishing Company.

Calabrese, R. and M. Zenga (2010). Bank loan recovery rates: Measuring and nonparametric density estimation. *Journal of Banking & Finance 34*(5), 903–911.

Caselli, S., S. Gatti, and F. Querci (2008). The Sensitivity of the Loss Given Default Rate to Systematic Risk: New Empirical Evidence on Bank Loans. *Journal of Financial Services Research 34*(1), 1–34.

Davis, R. H., D. B. Edelman, and A. J. Gammerman (1992). Machine-learning algorithms for credit-card applications. *IMA Journal of Management Mathematics 4*(1), 43–51.

De Laurentis, G. and M. Riani (2005). Estimating LGD in the Leasing Indutry: Empirical Evidence from a Multivariate Model. In E. I. Altman, A. Resti, and A. Sironi (Eds.), *Recovery Risk*, pp. 143–164. London: Risk Books.

Dermine, J. M. and C. Neto de Carvalho (2006). Bank loan losses-given-default: A case study. *Journal of Banking & Finance 30*(4), 1219–1243.

Eisfeldt, A. L. and A. A. Rampini (2009). Leasing, Ability to Repossess, and Debt Capacity. *Review of Financial Studies 22*(4), 1621–1657.

Forese, F. D. and M. T. Hagan (1997). Gauss-Newton approximation to Bayesian regularization. *Proceedings of the 1997 International Joint Conference on Neural Networks 3*, 1930–1935.

Franks, J., A. de Servigny, and S. Davydenko (2004). A comparative analysis of the recovery process and recovery rates for private companies in the UK, France and Germany. Working Paper, Standard & Poor's Risk Solutions.

Frye, J. (2005). The effects of systematic credit risk: a false sense of security. In E. I. Altman, A. Resti, and A. Sironi (Eds.), *Recovery Risk: The Next Challenge in Credit Risk Management*, pp. 187–200. London: Risk Books.

Gadd, A. and H. Wold (1964). The Janus quotient: a measure for the accuracy of prediction. In H. Wold (Ed.), *Econometric model building: essays on the causal chain approach*, pp. 229–235. Amsterdam: North-Holland Publishing Company.

Grunert, J. and M. Weber (2009). Recovery rates of commercial lending: Empirical evidence for German companies. *Journal of Banking & Finance 33*(3), 505–513.

Gupton, G. M. and R. M. Stein (2002). Losscalc: Moody's Model for Predicting Loss Given Default (LGD). Working Paper, Moody's Investors Service.

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

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

Hand, D. J. and W. E. Henley (1997). Statistical classification methods in consumer credit scoring: A review. *Journal of the Royal Statistical Society. Series A (Statistics in Society) 160*(3), 523–541.

Hartmann-Wendels, T. and M. Honal (2010). Do Economic Downturns Have an Impact on the Loss Given Default of Mobile Lease Contracts? – An Empirical Study for the German Leasing Market. *Credit and Capital Markets 43*(1), 65–96.

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

Hlawatsch, S. and S. Ostrowski (2011). Simulation and estimation of loss given default. *Journal of Credit Risk 7*(3), 39–73.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics 6*(2), 65–70.

Khieu, H. D., D. J. Mullineaux, and H.-C. Yi (2012). The determinants of bank loan recovery rates. *Journal of Banking & Finance 36*(4), 923–933.

Laurent, M.-P. and M. Schmit (2005). Estimating Distressed LGD on Defaulted Exposures: A Portfolio Model Applied to Leasing Contracts. In E. I. Altman, A. Resti, and A. Sironi (Eds.), *Recovery Risk*, pp. 307–322. London: Risk Books.

Lessmann, S., B. Baesens, H.-V. Seow, and L. Thomas (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research 247*(1), 124–136.

Loterman, G., I. Brown, D. Martens, C. Mues, and B. Baesens (2012). Benchmarking regression algorithms for loss given default modeling. *International Journal of Forecasting 28*(1), 161–170.

MacKay, D. J. C. (1992). Bayesian interpolation. *Neural Computation 4*(3), 415–447.

Nguyen, D. and B. Widrow (1990). Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. *Proceedings of the IJCNN 3*, 21–26.

Qi, M. and X. Zhao (2011). Comparison of modeling methods for loss given default. *Journal of Banking & Finance 35*(11), 2842–2855.

Rodriguez, P. P. and D. Gianola (2016). Bayesian Regularization for Feed-Forward Neural Networks. Working Paper.

Schmit, M. and J. Stuyck (2002). Recovery Rates in the Leasing Industry. Working Paper, Leaseurope.

Sigrist, F. and W. A. Stahel (2011). Using the censored gamma distribution for modeling fractional response variables with an application to loss given default. *ASTIN Bulletin 41*(2), 673–710.

Tobback, E., D. Martens, T. Van Gestel, and B. Baesens (2014). Forecasting loss given default models: impact of account characteristics and the macroeconomic state. *Journal of the Operational Research Society 65*(3), 376–392.

Varma, P. and R. Cantor (2005). Determinants of Recovery Rates on Defaulted Bonds and Loans for North American Corporate Issuers. *The Journal of Fixed Income 14*(4), 29–44.

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

Yashkir, O. and Y. Yashkir (2013). Loss given default modeling: a comparative analysis. *The Journal of Risk Model Validation 7*(1), 25–59.

Yeh, I. and C. Lien (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications 36*(2, Part 1), 2473–2480.

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

VII Appendix

Table 5: Results of pairwise out-of-sample t-tests of differences in the mean values of MAE and RSME for each pair of methods.

	Method	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Forecasting	*MAE*						
	(1) Hist. avg.	-					
	(2) OLS	2.1818*	-				
	(3) Tobit	7.2522***	5.1931***	-			
	(4) RT	-1.3722	-3.5047***	-8.3994***	-		
	(5) RF	23.5192***	21.5863***	15.7572***	4.2307***	-	
	(6) ANN	5.0544***	2.9355**	-2.2783*	6.2793***	-18.3672***	-
	RMSE						
	(1) Hist. avg.	-					
	(2) OLS	0.3976	-				
	(3) Tobit	0.3163	-0.0735	-			
	(4) RT	-1.5827	-2.0009	-1.8814	-		
	(5) RF	4.2609***	3.8748***	3.8686***	6.0175***	-	
	(6) ANN	-0.0327	-0.4267	-0.3456	1.5329	-4.2534***	-
Panel B. Estimation	*MAE*						
	(1) Hist. avg.	-					
	(2) OLS	19.5377***	-				
	(3) Tobit	26.9733***	2.2125	-			
	(4) RT	22.9942***	4.2294***	-4.5396***	-		
	(5) RF	59.3041***	42.8420***	31.9002***	37.3140***	-	
	(6) ANN	29.8562***	11.4081***	2.2125	6.9285***	-30.6003***	-
	RMSE						
	(1) Hist. avg.	-					
	(2) OLS	7.3891***	-				
	(3) Tobit	6.6765***	-0.4206	-			
	(4) RT	10.3194***	2.1096	2.4546*	-		
	(5) RF	23.0023***	16.0260***	15.4932***	17.1503***	-	
	(6) ANN	9.2278***	1.9674	2.2921	0.1729	-13.7659***	-

NB: The table displays pairwise t-statistics for differences in mean for the forecasting (Panel A) and estimation (Panel B) cases with p-values adjusted according to Holm (1979). Positive values indicate that the prediction accuracy for the method on the vertical axis is better than that for the method on the horizontal axis and vice versa. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

Table 6: In-sample forecasting and estimation changes in MAE for permuting the variables in each method.

		Forecasting					**Estimation**				
	Variables	OLS	Tobit	RT	RF	ANN	OLS	Tobit	RT	RF	ANN
Contract Characteristics	*SIZE_ORI*	-0.0064	-0.0079	0.0000	0.3303	0.0092	-	-	-	-	-
	D_COLL	0.0106	0.0108	0.0000	0.1278	0.0349	0.0115	0.0119	0.0000	0.1311	0.0255
	DUR_CONTRACT	-0.0032	-0.0020	0.0000	0.2041	0.0011	0.0123	0.0134	0.0000	0.2096	0.0062
	SPEC_PAYMENT	0.0176	0.0169	0.0000	0.1352	0.0234	0.0263	0.0253	0.0000	0.1462	0.0262
	D_OBJMULT	0.0009	0.0008	0.0000	0.0561	0.0325	-0.0008	-0.0011	0.0000	0.0671	0.0087
	CONTRACT	-0.0031	-0.0029	0.0000	0.2070	0.0428	-0.0081	-0.0080	0.0000	0.2411	0.1088
	ASSET	0.0144	0.0175	0.0000	0.2226	0.0840	0.0381	0.0434	0.0000	0.2367	0.1951
Lessee Charac.	*D_LIMITED*	0.0046	0.0049	0.0000	0.1636	0.0324	0.0019	0.0022	0.0000	0.1538	0.0296
	RAT_INT_BOND	0.0127	0.0156	0.0688	0.3644	0.0614	0.0387	0.0422	0.0521	0.3617	0.0256
	DISTANCE	0.0050	0.2983	0.0055	0.0057	0.0053	0.0000	0.3103	-0.0005	-0.0005	0.0021
	INDUSTRY	0.0199	0.0228	0.0000	0.2656	0.0551	0.0147	0.0168	0.0000	0.2674	0.0372
	D_MOVED	-	-	-	-	-	0.0003	0.0003	0.0000	0.0253	0.0116
Def. Char.	*DUR_WORK*	-	-	-	-	-	0.0116	0.0103	0.0070	0.3518	0.0200
	SIZE_DEF	-	-	-	-	-	0.0296	0.0358	0.0000	0.3865	0.0958
	PROP_PAID	-	-	-	-	-	0.0703	0.0649	0.2206	0.5828	0.1428
Macro	*GDP_INI*	0.0066	0.0071	0.0000	0.2744	0.0027	-	-	-	-	-
	INSOL_DEF	-	-	-	-	-	0.0156	0.0159	0.0000	0.3399	0.0317

NB: For *ASSET*, *INDUSTRY*, and *CONTRACT*, we permute the corresponding dummy variables.

Table 7: Out-of-sample forecasting and estimation changes in MAE for permuting the variables in each method.

		Forecasting					Estimation				
	Variables	OLS	Tobit	RT	RF	ANN	OLS	Tobit	RT	RF	ANN
Contract Characteristics	*SIZE_ORI*	0.0049	0.0317	-0.0033	-0.0017	0.0117	-	-	-	-	-
	D_COLL	0.0086	0.0068	0.0001	0.0198	0.0245	0.0089	0.0049	0.0000	0.0172	0.0161
	DUR_CONTRACT	0.0055	0.0062	0.0050	0.0228	0.0077	0.0162	0.0110	0.0004	0.0249	0.0095
	SPEC_PAYMENT	0.0115	0.0093	0.0010	0.0116	0.0122	0.0200	0.0119	0.0002	0.0167	0.0130
	D_OBJMULT	-0.0002	-0.0003	-0.0000	0.0018	0.0063	-0.0006	-0.0009	0.0000	0.0034	0.0045
	CONTRACT	-0.0025	-0.0011	0.0092	0.0176	0.0160	-0.0060	-0.0009	0.0006	0.0220	0.0489
	ASSET	0.0117	0.0144	0.0001	0.0247	0.0425	0.0311	0.0247	0.0000	0.0260	0.0915
Lessee Charac.	*D_LIMITED*	0.0022	0.0003	0.0000	0.0160	0.0082	0.0043	0.0024	0.0000	0.0130	0.0122
	RAT_INT_BOND	0.0108	0.0064	0.0446	0.0475	0.0588	0.0364	0.0066	0.0384	0.0489	0.0484
	DISTANCE	0.0071	0.0089	0.0057	0.0327	0.0044	-0.0009	0.0000	0.0000	0.0277	-0.0001
	INDUSTRY	0.0157	0.0160	0.0008	0.0243	0.0277	0.0144	0.0092	0.0002	0.0200	0.0303
	D_MOVED	-	-	-	-	-	-0.0003	0.0001	0.0000	0.0025	0.0046
Def. Char.	*DUR_WORK*	-	-	-	-	-	0.0212	0.0223	0.0188	0.0426	0.0223
	SIZE_DEF	-	-	-	-	-	0.0172	-0.0008	0.0232	0.0566	0.0470
	PROP_PAID	-	-	-	-	-	0.0612	0.0876	0.1511	0.1154	0.0836
Macro	*GDP_INI*	0.0022	0.0013	0.0016	0.0216	0.0014	-	-	-	-	-
	INSOL_DEF	-	-	-	-	-	0.0114	0.0077	0.0022	0.0290	0.0123

NB: For *ASSET*, *INDUSTRY*, and *CONTRACT*, we permute the corresponding dummy variables.