Journal of Banking and Finance 134 (2022) 106334

Contents lists available at ScienceDirect

Journal of Banking and Finance

journal homepage: www.elsevier.com/locate/jbf

Opening the black box – Quantile neural networks for loss given default prediction

Ralf Kellner[a], Maximilian Nagl[b,*], Daniel Rösch[b]

[a] University of Passau, Chair of Financial Data Analytics, 94032 Passau, Germany
[b] University of Regensburg, Chair of Statistics and Risk Management, 93040 Regensburg, Germany

ARTICLE INFO

Article history:
Received 25 February 2021
Accepted 10 October 2021
Available online 12 October 2021

JEL classification:
C21
G21
G33

Keywords:
Quantile regression
Black box
Neural networks
Explainable machine learning
Global credit data

ABSTRACT

We extend the linear quantile regression with a neural network structure to enable more flexibility in every quantile of the bank loan loss given default distribution. This allows us to model interactions and non-linear impacts of any kind without the need of specifying the exact form beforehand. The precision of the quantile forecasts increases up to 30% compared to the benchmark, especially for higher quantiles which are most important in credit risk. By using a novel feature importance measure, we calculate the strength, direction, interactions and other non-linear impacts for every conditional quantile and every variable. This enables us to explain why our extension exhibits superior performance over the benchmark. Moreover, we find that the macroeconomy is up to two times more important in USA than in Europe and has large joint impacts in both regions. The macroeconomy is most important in the US, whereas in Europe collateralization is essential.

1. Introduction

Estimation and prediction of Loss Given Default (LGD) is an important and challenging task for financial institutions. The LGD is the fraction of loss from the exposure at default and depending on the instrument, it can be divided into market-based and workout LGDs. The former is relevant for publicly traded instruments such as bonds and defined as one minus the ratio of the market price 30 days after default over the outstanding amount. The latter is commonly used for loan contracts and is determined by subsuming discounted payments from debtors during the process of default resolution. LGD distributions exhibit extreme and versatile shapes, typically with high probability masses centered around zero and one. Furthermore, predicting LGDs is a challenging task due to long-lasting and complex resolution processes. In this paper, we use access to a unique database of workout LGDs provided by Global Credit Data (GCD), which is a non profit initiative that supports banks by collecting and analyzing historical loss data from a multitude of member banks worldwide, encompassing several systematically relevant institutions (www.globalcreditdata.org). Moreover, our analysis is separated into a data set of American and European loans, as previous studies detected profound differences between those regions which can likely be ascribed to differences in the legal and regulatory environment.

After years of developing models for the probability of default (PD), LGD modeling has attracted more and more attention. In general, common drivers for LGDs are identified in different studies which also compare a variety of LGD models (see, e.g., Bastos, 2010; Grunert and Weber, 2009; Loterman et al., 2012; Qi and Yang, 2009; Qi and Zhao, 2011). Khieu et al. (2012) identified loan characteristics as more important for recovery rates of bonds, but macroeconomic variables also play an important role. Hence, the discussion whether and which economic variables are important is still active, (see, e.g., Betz et al., 2018; Krüger and Rösch, 2017; Leow et al., 2014; Nazemi et al., 2017, 2021; Nazemi and Fabozzi, 2018). The latter two use principal components gathered from many different macroeconomic variables, which we also follow in this study. Furthermore, Krüger and Rösch (2017) find a varying (linear) impact over the entire conditional LGD distribution, which points to a complex and potentially non-linear relationship between LGDs and the economy. This is in line with findings of Sopitpongstorn et al. (2021), who find non-linearities between covariates and market-based recovery rates by using a local logit regression.

However, a high amount of publications is dedicated to market-based or expected LGDs which exhibit significant differences to

* Corresponding author.
E-mail address: maximilian.nagl@ur.de (M. Nagl).

https://doi.org/10.1016/j.jbankfin.2021.106334
0378-4266/

workout LGDs.[1] In contrast to the latter, market-based LGDs are bound in the interval $[0, 1]$ and are characterized by short resolution processes. Especially for workout LGDs, the distributional form is extreme due to values below zero, above one, along with high probability masses at zero and one. This is probably the reason why mixture distributions and other models with flexible distributional forms best capture the workout LGD distribution. Altman and Kalotay (2014) develop a Bayesian finite mixture model of normal distributions with an underlying ordered logit model which links debtor features to mixture component affiliation. A frequentistic version of this model is presented by Kalotay and Altman (2017) and a mixture of beta distributions is applied by Calabrese (2014). Variants of mixture models are also used in Betz et al. (2018) and Tomarchio and Punzo (2019). An alternative approach is shown by Krüger and Rösch (2017) who use linear quantile regression to predict different parts of the LGD distribution. Even though this approach is able to capture a varying impact of predictors over the distribution, it is restricted to a linear relationship between predictors and the variable of interest, and the evaluation of (non-linear) interactions would be computationally burdensome.[2]

Neural networks have previously been applied to the estimation of LGDs (see, Qi and Zhao, 2011; Loterman et al., 2012). However, in comparison to our approach, the network is calibrated to predict the mean value. In contrast, we calibrate the network for a discrete set of quantiles of the LGD distribution using the quantile specific loss function and control for strict monotony of quantile estimates. As shown by Krüger and Rösch (2017), it is important to account for varying impact for different quantile levels. This may explain why the application of neural networks to LGD modeling in previous studies has often led to worse results in comparison to other flexible approaches like regression trees and support vector regression. However, neural networks already exhibit promising results in comparison to less flexible approaches like transformed regression type models. Yet, a disadvantage is the alleged incapability of identifying (relevant) predictors for LGDs (see, Qi and Zhao, 2011). However, this disadvantage vanishes in the light of recent techniques. In this paper, we use a feature importance measure based on gradient information as shown in Horel et al. (2018) and Nagl (2021). It enables us to decompose the prediction of neural networks into their relative feature importance and interactions with all other features. This gives us a broad and detailed description of the underlying relations. Furthermore, the computational burden to model and test joint effects and interactions between independent variables can be considerably reduced. As an example, assume that we want to estimate 100 quantiles using 26 different predictor variables. To test and model every pairwise interaction, e.g., $x_1 \cdot x_2$, one has to fit $26 \cdot \frac{26-1}{2} \cdot 100 = 32,500$ different models. If we think about non-linear interactions as well, e.g., $\exp(x_1 \cdot x_2)$ or $x_1^2 \cdot x_2$, this number rises fast and results in a computationally expensive and tedious task. By using our approach we have to fit only one model to capture all these possible forms at once. Hence, it is not necessary to assume a structural relationship beforehand, but we can quantify the strength of joint effects and non-linearities for every quantile afterwards.

Quantiles are important not only for the overall distribution, but also to differentiate between loans and their inherent risk profile. Consider the following stylized example of different quantiles of the Loss Given Default:

	$\hat{Q}_{0.05}$	$\hat{Q}_{0.25}$	$\hat{Q}_{0.50}$	$\hat{Q}_{0.75}$	$\hat{Q}_{0.95}$
loan 1	0.03	0.08	0.20	0.59	0.70
loan 2	0.03	0.08	0.20	0.67	0.90
loan 3	0.02	0.05	0.20	0.59	0.70

If we focus only on median estimates ($\hat{Q}_{0.50}$), one might come to the conclusion that every loan exposes the bank to the same risk in terms of LGD. But focusing on higher quantiles, e.g. unfavourable scenarios for the bank, we observe that loan 2 entails a considerably higher risk. On the contrary, loan 3 contains the least risk as the lower quantiles are smaller. Moreover, our empirical analysis finds that higher quantiles are driven by a higher sensitivity to the macroeconomy and overall higher non-linearity and interactions compared to the median, which underlines the importance of quantiles. In summary, mean-related methods may not be representative for bimodal distributions, which are characterized rather by their tails than their expectation and that quantiles can enhance the bank's ability to differentiate between risk profiles of obligors.

Estimating conditional quantiles has emerged as a powerful method and widespread potential application. The most common method is to minimize the expectation of the so-called check-function leading to the linear quantile regression introduced by Koenker and Bassett (1978) and for example used by Krüger and Rösch (2017). For a timely and comprehensive overview of various extensions and applications we refer to Koenker et al. (2017). In general, non-linearity can be allowed by additive quantile regressions using splines, see e.g. Koenker et al. (1994), Horowitz and Lee (2005) and Hoshino (2014). However, one has to choose which variables should expressed as splines or tensors if non-linear interactions should be allowed. Fully non-parametric quantile regressions were introduced by Koenker (2005), Li and Racine (2008) and Li et al. (2013). Along with the rise of quantile regressions, also the discussion of non-monotone quantile estimates appeared. For example, Takeuchi et al. (2006) shows that monotonicity can be included directly in the estimation procedure. A more general post-hoc method to ensure monotone quantile function is introduced by Chernozhukov et al. (2010) who argue to simply rearrange quantile estimates. This approach is for example used by Wu and Yan (2019).

We contribute to the literature by developing a quantile neural network regression model with a sound estimation procedure. The calibration to each quantile is subsumed in a single optimization step. This considerably reduces the computational burden, especially with respect to possible forms of interactions between variables. Furthermore, we find a superior in- and out-of-time performance in both out-of-time periods (Great Financial Crisis (GFC) and post GFC) compared to the linear counterpart. This can be traced back to non-linear relationships between predictors and LGDs especially in higher quantile levels. By using the feature importance measure, we can attribute the superior performance of the QRNN approach to non-linearity in specific quantiles and the variables driving it. This offers insights into risk drivers that have not been detected in earlier literature. Furthermore, explainability is of great importance for financial institutions, as neural networks are often falsely accused of being black boxes in the financial community and regulators strictly allow LGD predictions if their derivation is transparent to them. This has prohibited the use of superior models for regulatory and internal risk management purposes. From an economic perspective, we find that the impact of the macroeconomy is up to two times more important in the US than in Europe. The results suggest that in Europe the economic surrounding interacts most with variables describing the different forms of collateralization. On the contrary, in the United States the level of seniority has a large joint impact with the macroeconomy on the LGD prediction, especially for higher quantiles.

[1] A comprehensive study on expected LGDs based on Credit Default Swaps can be found in Doshi et al. (2018).

[2] Similar approaches focusing on the estimation of the distribution of LGD can be found in Hwang and Chu (2018) and Hwang et al. (2020). Both rely on inverse-probability-transformations of the true LGD values which is not feasible for workout LGDs as they are not bounded between zero and one.

The remainder of this paper is structured as follows. In Section 2, we give a short summary of relevant literature with the use of machine learning for credit risk modeling. Data is presented in Section 3, while the methodology is described in Section 4. Our empirical results are discussed in Section 5 and Section 6 concludes.

2. Literature review

Over the last decade, there is an increasing attention of machine and deep learning algorithms in credit risk. We can name two possible reasons. First, computational power increased massively in recent years and open-source solutions have been widely developed. This makes highly complex algorithms available to a very broad audience, which covers academics, but also practitioners and regulators. Second, superior performance of these algorithms is well documented. The following section provides a brief review of recent studies using machine or deep learning in credit risk.

Overall, the literature concerning PD is much wider than for LGD. There are many studies which compare supervised machine learning algorithms with respect to their predictive power (see, e.g. Cowden et al., 2019; Li and Chen, 2019; Chen et al., 2020; Petropoulos et al., 2020; Luo et al., 2020; Dumitrescu et al., 2021). A general consensus exists that more flexible models outperform the linear logit regression. Bakoben et al. (2020) employ unsupervised learning approaches to detect different credit card account behaviours, increasing predictive power. A promising part of machine learning are deep learning models. Kvamme et al. (2018) utilize deep convolutional networks to predict probability of default of mortgage loans, showing a superior performance. Mai et al. (2019) also use deep convolutional networks and ensemble techniques to predict corporate defaults. They further use textual disclosures into to enhance the discrimination power of their models. Another perspective of deep learning is shown by Sariev and Germano (2019). They utilize Bayesian regularized neural networks to automatically determine the regularization in the network. Furthermore, the combination of machine learning and classical statistical models seems to be promising as well. Li and Chen (2019) combine logistic regression and neural networks to enhance discrimination power. Sigrist and Hirnschall (2019) unite Tobit regression with regression trees. Our paper also follows this line, as we combine statistical models with machine learning methods. For a detailed overview of machine learning in PD modelling, we refer to Mai et al. (2019). Jing et al. (2020) are among the first to incorporate the evolution of PDs over time in a long short-term memory network.

The application of machine learning for Loss Given Defaults has become very popular in the last decade. Early studies were conducted by Matuszyk et al. (2010) and Bastos (2010) using tree based methods. Comprehensive benchmark studies were conducted by Qi and Zhao (2011), Bellotti and Crook (2012) and Loterman et al. (2012). The latter one may be the most comprehensive by testing 24 different regression algorithms in total, based on six real world datasets and finding evidence that non-linear techniques perform best. Some other studies put more emphasis on the comparison of two stage and single models, such as for example Tobback et al. (2014), Sun and Jin (2016) or Tanoue and Yamashita (2019). Yao et al. (2017) propose two-stage approaches involving support vector machines for LGD prediction. Contrary, Nazemi et al. (2017) find that fuzzy decision methods perform best. Nazemi et al. (2018) show that using principal components, an unsupervised learning approach, derived from a wide range of macroeconomic variables enhances the prediction performance. Most recent studies focus on the comparison of a very wide range of models, macroeconomic variables and LGD types. Bellotti et al. (2021) conduct an exhaustive analysis with various different models and find tree based methods to be superior. Kaposty et al. (2020) conduct a horse race of different models, in which random forests turn out to be the best ones. Gambetti et al. (2020) also use a vast selection of machine learning models and introduce meta-learning strategies, providing evidence that the macroeconomic surrounding is important for all methods if it is incorporated via uncertainty measures. This complements findings of Gambetti et al. (2019) that uncertainty is the most important macroeconomic variable for (market) LGD prediction.

In recent years, the body of literature focusing on explanation methods has grown fast. One can divide this body into local explanations, i.e. explain the individual prediction of an observation, and global explanations, i.e. explain the learned relation of the black-box model. For an excellent and detailed review regarding these methods, we refer to Horel and Giesecke (2020). Partial Dependence Plots (PDP) are introduced by Friedman (2001). They plot the importance of a feature by varying over its marginal distribution and calculate the (global) effect on the resulting prediction. Goldstein et al. (2015) extend this idea to individual predictions by introducing the Individual Conditional Expectation (ICE) Plots. Apley and Zhu (2020) introduce Accumulated Local Effects (ALE) Plots and focus on the conditional distribution of features instead, solving problems of PDP and ICE plots. One of the most prominent method is Local Interpretable Model-agnostic Explanations (LIME) introduced by Ribeiro et al. (2016). To explain any black-box model LIME perturbs the data for a given observation and get the black box predictions for these new points. Afterwards, a white-box model, such as a linear regression, is fitted to the permuted data and predictions. Ribeiro et al. (2018) introduced anchors, which decompose black-box predictions into highly interpretable if-else rules, e.g. if x_{i1} is greater than threshold z_{i1}, predict y_i. Recently, many applications of SHapley Additive exPlanations (SHAP), introduced by Lundberg and Lee (2017), can be found in the literature. This feature importance measure is the only one backed by an economic theory called coalitional game theory. SHAP values can be calculated for individual predictions and for global explanations. Horel and Giesecke (2020) continues the work by Horel et al. (2018) and derive a way to statistically test the impact of the feature importances. Their test statistics are valid for single-layer networks using mean-squared error loss and sigmoid activation functions. Nagl (2021) builds on the work by Horel et al. (2018) and addresses the quantification of non-linearity and interactions entailed in black-box predictions.

Summarizing, recent studies focusing on machine learning applied LGD are typically conducted in the spirit of “horse races” in which various methods are compared in their performance. We rather show how to extend the economically useful and meaningful method of quantile regression with non-linearity, feature importance of independent variables and with the identification of their interactions. Moreover, other studies commonly focus on mean predictions and not on the entire distribution (or quantiles). The approach that we present explicitly focusses on quantiles and, hence, delivers a much broader picture of LGDs. Additionally, the measure for feature importance can especially and easily be applied to the Quantile Regression Neural Network and sheds light on risk drivers and joint effects that have not been documented in the literature before. While the general idea to extend quantile regression with stacked layers of neural network architectures has been applied to fields other than credit[3] – to the best of our knowledge – this paper is the first using this approach in credit risk. Hereby, domain specific adjustments such as a unified esti-

[3] See, e.g., applications for time series models in Xu et al. (2016), Salinas et al. (2019) and Wu and Yan (2019).

mation procedure with monotonicity regularization are developed as a new unique contribution to the credit risk literature.

3. Data

We use access to Global Credit Data (GCD), one of the world's largest loss data bases. This consortium consists of 55 globally acting banks, encompassing several systemically relevant institutions. The data offer an unique and broad perspective of the banking universe.[4] The information is based on transactions, providing a detailed view on occurred losses and their determinants. We focus on workout Recovery Rates, including post-default cash flows. Recovery Rates are the difference between discounted positive cash flows (CF^+) and discounted direct as well as indirect costs (CF^-), divided by the exposure at default (EAD). The LGD is defined as one minus the Recovery Rate[5]:

$$LGD = 1 - \frac{\sum_{i=1}^{n} CF^+ - \sum_{i=1}^{n} CF^-}{EAD}. \quad (1)$$

To check for appropriateness of the calculated LGD values, we use the same procedure as Betz et al. (2018). We impose a materiality restriction of 500 USD and only use resolved loans to avoid the well-known default resolution bias, see e.g. Betz et al. (2018) and Betz et al. (2020). Considering this, we restrict our data sample to the time period of 2000–2016, as the resolution of loan contracts can last several years. Workout LGDs are not compulsorily restricted to the interval of 0% and 100%. Hence, we cut values outside of the range [-25%, 125%].[6] Actually, this restriction is only of minor concern using quantile-based methods, because they are generally less sensitive to outliers. Nevertheless, it insures a rather homogeneous sample. To compare different economic and geographic regions, we use a European and US American sample. Table 1 shows descriptive statistics for both considered regions. For metric dependent variables, selected quantiles, the mean and standard deviation are presented. For categorical variables, quantiles of the LGD distribution in these categorical subgroups are shown. The first level of each categorical variable is used as the reference level.

For example, facility types are divided into two subcategories, namely term loans and credit lines. Regarding the quantiles in the US American sample, we observe that the distribution is very similar for quantiles lower than the median, whereas credit lines have considerably more mass at the 75% quantile. The difference between credit line and term loan is even more pronounced in the European sample. This may indicate that high quantiles are quite different regarding the facility type. Considerably deviating effects can also be observed in other categories like asset class, industry and collateral. This may imply that one has to account for different impacts on different parts of the LGD distribution, underlining our quantile based approach. To account for systematic effects, we conduct a Principal Component Analysis (PCA) of the most common macroeconomic variables regarding workout LGD estimation, following Nazemi et al. (2018).[7] To account for roughly 95% of their variance, we derive eight components in Europe and eleven components in the United States.[8] This reduces the problem of selecting the most suitable macroeconomic variables, which is especially difficult for workout LGDs due to their very long resolution time, see for example Betz et al. (2020). The considered variables can be found in Table A.1 in Appendix A. In general, we include variables to capture the economic uncertainty, following Gambetti et al. (2019), the general economic situation and monetary and inflation related measures. Furthermore, our analysis includes several survey based variables.

For the rest of the paper, we divide the subsamples further, namely into a training sample starting in the year 2000, another one that contains LGDs during the GFC and the final one consisting of LGDs after the GFC until 2016. We follow the OECD which specifies the crisis period in the US from 2007 Q4 until 2009 Q2 and from 2008 Q1 until 2009 Q3 in Europe. We do not additionally split the training sample into in- and out-of-sample on a cross-sectional basis, as predicting the future is of major concern in credit risk. For example, Kalotay and Altman (2017) put emphasis on the fact that it is crucial to predict future and not only contemporary LGDs. Fig. 1 illustrates the different behaviour of average LGDs over time. In the training sample, the mean values follow a very similar path over time, whereas in early years the LGDs in USA are slightly higher. Furthermore, in both regions, the GFC is observable and characterized by considerably higher LGDs. The main difference between both regions can be deduced after the GFC. In the European sample, the mean LGD values deteriorate towards low levels. On the contrary, the LGD values in the US American sample remain high and even increase in the time period around 2013–2014. Both different time patterns may be challenging for any model, due to signals of non-stationary behaviour requiring great flexibility. The different behaviour also points towards different systematic effects in both regions.

Fig. 2 shows the density estimates for both regions and different subsamples. The bandwidth for the density plot is held equal to allow a comparison. The LGD values used in this paper show a bimodal shape with large masses around zero and one. With respect to the US American sample, we can see that the mass on extreme high losses (LGD $\geq$1) is lowest in the training sample. Furthermore, the probability mass of high losses is even higher in the sample period after the GFC than in the crisis period itself. This may mainly be driven by the peak around 2014. Overall, we can see that the probability mass is shifted from low to high losses along the time line. This is contrary to Europe, where we observe the largest mass on high losses during the GFC and very equal masses in the other periods. This analysis gives another empirical indication why modeling quantiles is important for Loss Given Defaults. Fig. 2 shows that for the 25% to 75% quantiles, i.e. the middle of the distribution, we have only very little observations as the very large part of all LGD realizations lie on the left and right tail. Hence, focusing on the expectation of this distribution would imply that one would focus on regions of the distribution which are rarely observed in practice.

With respect to the descriptive analysis of both regions, some differences and modelling implications are revealed. Both regions may be shaped by different systematic behaviours, which implies different models for both regions. Furthermore, the very different out-of-time behaviour requires a very flexible modelling approach. Additionally, we see deviating impacts on different parts of the LGD distribution, which suggests the need for increasing the model flexibility in this direction as well.

4. Methods

Quantile Regression (QR)

Krüger and Rösch (2017) use a linear quantile regression approach to tackle the challenging distributional form. Following Koenker and Bassett (1978), the τ-th conditional quantile of the

[4] For recent information about GCD we refer to Brumma et al. (2020a,b).

[5] For a detailed overview of positive and negative cash flows, see Betz et al. (2020).

[6] In total, we cut off 3.48% of the available resolved loans due to all restrictions.

[7] For a detailed methodical overview of the PCA, we refer to Nazemi et al. (2018).

[8] We also conducted all analyses with 90% and 99,5% accounted variance as a robustness test. All conclusions and results remain the same as outlined in the following sections. The evaluations with the alternative settings are available upon request.

Table 1
Descriptive statistics.

(a) US sample

Variable	Level	Quantiles					Mean	STD	Obs.
		0.05	0.25	0.50	0.75	0.95			
LGD		0.00	0.37	4.51	51.34	100.00	27.89	37.29	9649
log(EAD)		9.99	12.10	13.56	15.24	17.00	13.59	2.18	9649
PC 1		37.97	59.46	78.39	94.04	128.34	78.89	27.40	9649
PC 2		-194.10	-165.38	-127.17	-107.14	-74.34	-134.68	38.89	9649
PC 3		98.01	126.82	154.96	202.31	260.96	164.30	47.94	9649
PC 4		10.52	28.96	38.94	58.29	82.48	43.04	21.07	9649
PC 5		-44.24	-27.30	-17.93	-5.76	38.45	-14.72	21.75	9649
PC 6		-49.21	-14.80	-4.72	3.29	8.89	-7.39	15.44	9649
PC 7		-5.56	7.41	21.53	34.49	112.00	27.79	31.62	9649
PC 8		-27.16	-15.26	-1.80	13.52	35.46	0.34	18.47	9649
PC 9		23.48	30.01	37.84	55.61	75.20	42.58	15.31	9649
PC 10		-30.02	-4.65	7.90	14.70	30.74	5.58	15.96	9649
PC 11		-61.04	-20.00	-15.41	-12.76	-2.12	-19.52	16.52	9649
Facility type	Term loan	0.00	0.49	5.00	44.47	100.00	25.56	34.79	5331
	Credit line	-0.01	0.25	3.77	63.80	100.00	30.77	39.99	4318
Seniority	Pari passu	-0.14	0.17	6.62	100.00	100.00	38.33	44.35	3859
	Senior	0.00	0.54	4.06	33.87	93.43	20.89	29.61	5623
	Non senior	0.00	0.06	1.01	38.60	100.00	22.33	34.51	167
Industry	FIRE	-0.03	0.42	5.56	65.55	100.00	31.64	40.51	1181
	Agriculture	0.00	0.03	2.39	40.18	100.00	23.99	34.82	663
	Mining	-0.08	0.00	1.09	29.15	96.85	18.21	28.98	167
	Construction	0.00	0.34	4.38	42.36	100.00	24.63	34.04	1853
	Manufacturing	0.00	1.60	7.51	26.42	79.96	19.49	25.96	279
	Transport	0.00	0.69	6.59	66.63	100.00	32.24	39.94	1021
	Wholesale	0.00	0.42	4.91	52.96	100.00	28.06	36.86	2189
	Services	0.00	0.33	3.85	59.24	100.00	29.34	39.05	2296
Asset Class	SME	0.00	0.26	3.54	68.79	100.00	31.19	40.82	5870
	Large corporates	0.00	0.63	6.11	37.96	93.26	22.76	30.33	3779
Collateral	No	-0.05	0.54	6.28	69.27	100.00	31.73	40.29	1919
	Real Estate	-0.02	0.42	5.19	97.47	100.00	33.92	42.36	1160
	Yes	0.00	0.33	4.04	47.00	100.00	25.70	35.16	6570
Guaranteee	No	0.00	0.51	5.02	53.47	100.00	28.42	37.51	6864
	Yes	0.00	0.19	3.26	48.34	100.00	26.58	36.72	2785

(b) European sample

Variable	Level	Quantiles					Mean	STD	Obs.
		0.05	0.25	0.50	0.75	0.95			
LGD		0.00	0.00	1.18	26.25	100.00	21.18	34.89	44480
log(EAD)		8.52	10.52	11.83	13.15	15.26	11.85	2.04	44480
PC 1		-380.13	-253.38	-81.99	280.37	879.00	66.46	409.94	44480
PC 2		-5288.11	-1994.81	-91.59	831.00	1938.78	-721.75	2308.33	44480
PC 3		-1419.16	-570.62	141.55	1601.34	4117.37	614.83	1778.93	44480
PC 4		-5057.54	-1946.21	-189.13	710.03	1928.93	-699.37	2232.38	44480
PC 5		-899.02	-363.31	97.23	1032.14	2646.32	403.21	1145.24	44480
PC 6		-765.30	-280.80	124.80	826.23	2103.84	322.12	916.72	44480
PC 7		-399.50	-103.94	160.36	576.27	1457.96	292.55	588.54	44480
PC 8		-3109.94	-1169.66	-37.36	525.33	1250.03	-375.62	1399.83	44480
Facility type	Term loan	0.00	0.00	0.91	15.92	100.00	17.26	31.24	26705
	Credit line	0.00	0.00	1.94	57.00	100.00	27.08	39.02	17775
Seniority	Pari passu	0.00	0.00	1.66	31.94	100.00	21.72	34.50	6553
	Senior	0.00	0.00	1.18	26.38	100.00	21.30	35.09	36264
	Non senior	0.00	0.00	0.28	10.52	98.43	16.57	31.53	1663
Industry	FIRE	0.00	0.00	0.69	14.07	100.00	17.76	32.65	7039
	Agriclulture	0.00	0.00	0.94	27.00	100.00	21.01	35.19	1382
	Mining	0.00	0.00	1.72	14.29	94.86	15.80	28.70	239
	Construction	0.00	0.02	1.72	24.90	100.00	20.92	34.28	7841
	Manufacturing	0.00	0.00	1.62	18.08	99.66	19.59	33.66	262
	Transport	0.00	0.00	1.49	26.36	100.00	21.42	34.88	5901
	Wholesale	0.00	0.00	1.01	31.84	100.00	22.40	36.08	10592
	Services	0.00	0.00	1.37	31.93	100.00	22.41	35.50	11224
Asset Class	SME	0.00	0.00	1.03	26.77	100.00	21.37	35.28	39563
	Large corporates	0.00	0.02	3.76	22.68	100.00	19.71	31.56	4917
Collateral	No	0.00	0.03	1.95	50.43	100.00	26.28	39.15	15964
	Real Estate	0.00	0.00	0.47	7.83	94.06	13.81	27.88	19420
	Yes	0.00	0.03	3.69	55.96	100.00	27.98	37.37	9096
Guarantee	No	0.00	0.00	1.51	38.19	100.00	23.58	36.57	34936
	Yes	0.00	0.00	0.44	7.24	88.64	12.40	26.08	9544

Note: The table shows means, standard deviations and quantiles for the LGD and the metric variables. For categorical variables, means, standard deviations and quantiles of the LGD for each level are displayed.The PCs capture roughly 95% of the macroeconomic variable's variance outlined in Table A.1. The first level of each category is used as the reference level.

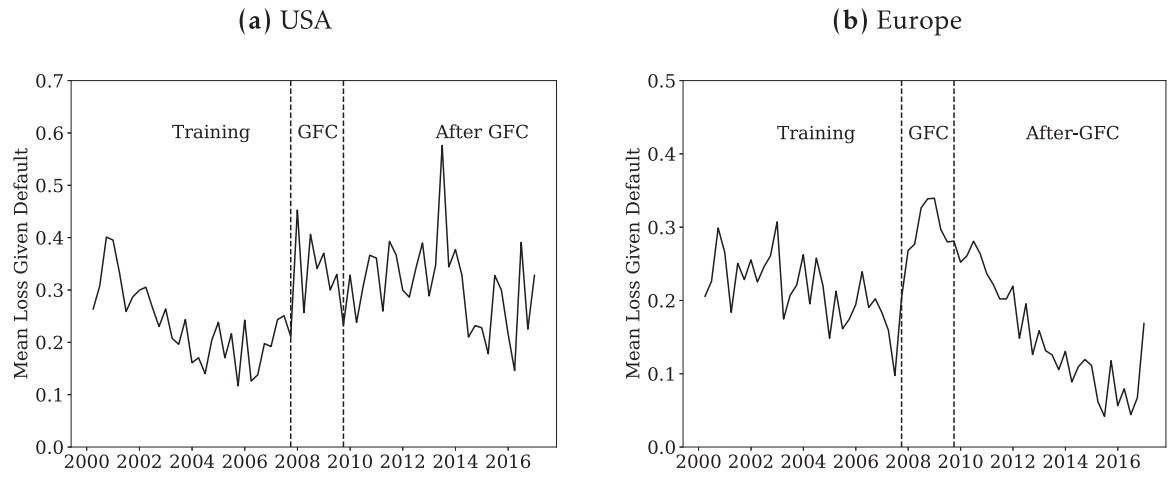

Fig. 1. Time variation of average LGDs. Note: This figure shows the time variation of the LGD means over time. The left panel illustrates mean values for the US American sample, whereas the right panel refers to the European sample. Furthermore, the time span of different samples is indicated via vertical dashed lines. We divide our sample in training, GFC and the time after the GFC.

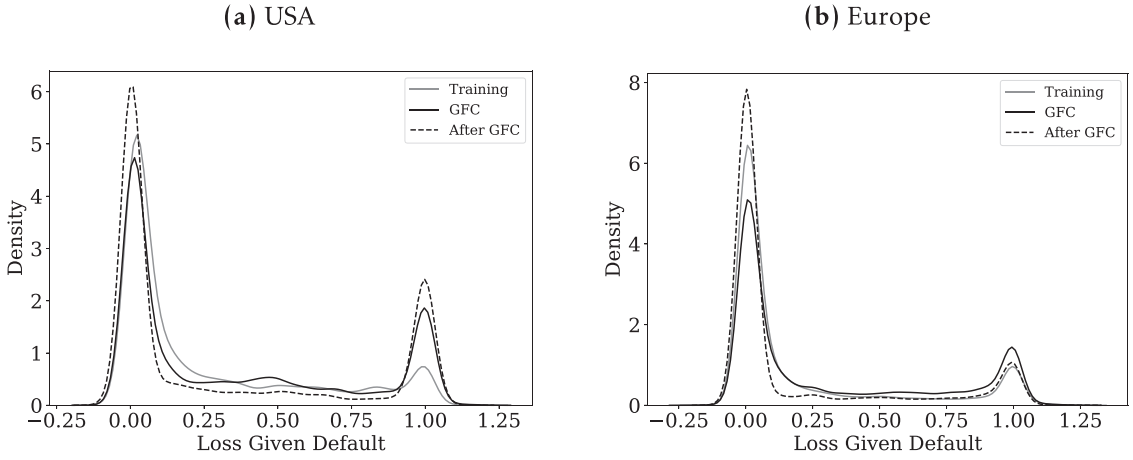

Fig. 2. Estimated densities of LGD distributions. Note: This figure shows kernel density estimates of different time periods in different regions. The left panel illustrates the shape of the LGD distributions in the US American sample, whereas the right panel refers to the European sample. We used the same bandwidth for all density estimates to allow for comparison between regions as well as time periods.

response y_i given $\boldsymbol{x}_i$ is:

$$Q_\tau(y_i|\boldsymbol{x}_i) = \beta_{0,\tau} + \boldsymbol{\beta}_\tau^T \boldsymbol{x}_i, \tag{2}$$

with $\boldsymbol{x}_i \in \mathbb{R}^p$ as a vector of p covariates for any observation $i = 1, \ldots, N$ and $\beta_{0,\tau} \in \mathbb{R}$, $\boldsymbol{\beta}_\tau \in \mathbb{R}^p$ as model parameters. The so-called check function $\rho_\tau(\omega)$ is defined as:

$$\rho_\tau(\omega) = \begin{cases} \tau\omega, & \text{if } \omega \geq 0, \\ (1-\tau)|\omega|, & \text{else.} \end{cases} \tag{3}$$

Parameter estimates are derived by minimizing the sum over all data points:

$$\sum_{i=1}^{N} \rho_\tau\left(y_i - \beta_{0,\tau} - \boldsymbol{\beta}_\tau^T \boldsymbol{x}_i\right). \tag{4}$$

This approach has at least two shortcomings: First, the quantile functions are estimated separately, leading to non-monotone quantiles. Hence, the estimated values of the quantile function may not increase with τ. Second, the impact of covariates is linear in the quantiles which may not be flexible enough, especially for extreme quantiles.

Quantile Regression Neural Networks (QRNN)

We extend this approach by using an Artificial Neural Network (ANN) to approximate the LGD distribution. Fig. B.1 in Appendix B shows a graphical comparison of both approaches for the interested reader. To generate predictions for a set of τ-th LGD quantiles, the neural network starts with covariate matrix $\boldsymbol{X} \in \mathbb{R}^{N \times p}$ as inputs in the input neurons. The network consists of stacked hidden layers $l = 1, \ldots, L$ whereby each layer consists of K_l neurons $\boldsymbol{h}_l \in \mathbb{R}^{K_l}$ that are determined by an affine combination of neurons in the previous layer which is composed of an arbitrary (non-linear) activation function σ.

$$\boldsymbol{h}_l = \sigma\left(\boldsymbol{W}_l \boldsymbol{h}_{(l-1)} + \boldsymbol{b}_l\right)$$

with $\boldsymbol{W}_l \in \mathbb{R}^{K_l \times K_{l-1}}$, $\boldsymbol{b}_l \in \mathbb{R}^{K_l}$ as parameters which are usually called weights and biases. Quantiles are derived from the last layer, the so-called output layer $L+1$ and are given by choosing the identity function for σ, resulting in:

$$Q_\tau(\boldsymbol{y}|\boldsymbol{X}) = \boldsymbol{W}_{\tau,L+1} \boldsymbol{h}_L + \boldsymbol{b}_{\tau,L+1}$$

with $\boldsymbol{y} \in \mathbb{R}^N$ is a vector of all response realizations y_i. It should be noted that weights and biases are shared among different levels of τ, except in the output layer, which is highlighted by the subscript τ for weights and biases in the output layer, only. That is different to traditional linear quantile regression and motivated to keep the model as parsimonious as possible. The weights and biases are estimated via a backpropagation algorithm based on Rumelhart et al. (1986). This requires a loss function to be dif-

ferentiable at any point. However, $\rho_\tau(\omega)$ cannot be differentiated at the origin. Therefore, we approximate this region following Huber (1964). This approach approximates this region quadratically, ensuring differentiation at any point. Moreover, as we estimate several quantile functions simultaneously, we have to ensure monotonicity using a penalty similar to Takeuchi et al. (2006). The penalty increases if there are non-monotone quantiles in any estimated quantile function for any different LGD observation.

Formally, the new quantile loss $\rho_\tau^{QRNN}(\omega)$, the Huber loss $h(\omega)$ and the monotonicity penalty $m(\boldsymbol{X})$ are defined as:

$$\rho_\tau^{QRNN}(\omega) = \begin{cases} \tau h(\omega), & \text{if } \omega \geq 0, \\ (1-\tau)h(\omega), & \text{else.} \end{cases} \quad \text{(Quantile Loss)}$$

$$h(\omega) = \begin{cases} \frac{1}{2}\omega^2, & \text{if } -\varepsilon \leq \omega \leq \varepsilon, \\ \varepsilon\left(|\omega| - \frac{1}{2}\varepsilon\right), & \text{else.} \end{cases} \quad \text{(Huber Loss)}$$

$$m(\boldsymbol{X}) = \sum_{i=1}^{N}\sum_{t=1}^{\theta-2} \max\left(0; Q_{\frac{t}{\theta}}(y_i|\boldsymbol{x}_i) - Q_{\frac{t+1}{\theta}}(y_i|\boldsymbol{x}_i)\right). \quad \text{(Monotonicity Penalty)} \tag{5}$$

where $\theta - 1$ is the number of quantiles which are estimated and $\frac{1}{\theta}, \frac{2}{\theta}, \ldots, \frac{\theta-1}{\theta}$ are corresponding quantile levels. The target function to minimize via the QRNN is defined as the sum of quantile and loss for all data points $i = 1, \ldots, N$ plus the monotonicity penalty which punishes non-(strict) monotonic behaviour for every data point:

$$L = \sum_{t=1}^{\theta-1}\sum_{i=1}^{N} \rho_{\frac{t}{\theta}}^{QRNN}(y_i - Q_{\frac{t}{\theta}}(y_i|\boldsymbol{x}_i)) + m(\boldsymbol{X}) \tag{6}$$

For Eq. (6) we use $\varepsilon = 0.0001$. Overall, we find that the fit does not depend on ε.

Feature Importance

Neural networks have become widespread in finance over the past years. However, one main issue is still the lack of interpretability. We use approaches based on Horel et al. (2018) and the extensions in Nagl (2021) to open up these black boxes. The approaches focus on the "learned" relations of the neural network and are therefore estimated using the training data. Overall, we use three different measures to explain the QRNN. The first order feature importance $FI_\tau^{First}(x_r)$ quantifies the overall importance of an input variable $r = 1, \ldots, p$. For our purpose, the first order feature importance is given by:

$$FI_\tau^{First}(x_r) = \frac{1}{C}\,\text{sgn}\left(\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial \hat{Q}_\tau(y_i|\boldsymbol{x}_i)}{\partial x_{ir}}\right)\right)\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial \hat{Q}_\tau(y_i|\boldsymbol{x}_i)}{\partial x_{ir}}\right)^2} \tag{7}$$

$FI_\tau^{First}(x_r)$ is the feature importance of covariate x_r at quantile level τ, $\hat{Q}_\tau(y_i|\boldsymbol{x}_i)$ is the conditional quantile estimate and C is a normalizing constant that ensures $\sum_{r=1}^{p}|FI_\tau(x_r)| = 1$. The $\text{sgn}(\cdot)$ operator defines the direction in which the feature drives the prediction. This feature importance employs the gradient for every covariate x_r in relation to $\hat{Q}_\tau(y_i|\boldsymbol{x}_i)$. All variables must be standardized, e.g., mean-scaling, to allow for comparison. The gradients are squared to avoid cancellations of positive and negative values. Furthermore, it sums up to 1, allowing an easy interpretation of "relative" importance. The extension in Nagl (2021) also quantifies the direction of the feature importance. This is achieved by taking the mean values of each gradient.

Additionally, we may argue that some input features have a joint impact, i.e., interacting with each other. For example, the importance of a collateral may also depend on the state of the business cycle, as in downturns bankruptcies may become more widespread. The feature importance can be extended to quantify joint impacts of features, see Nagl (2021). Additionally, we calculate the second partial derivative with respect to the same input feature to find the quantity of (single) non-linear impact. The second order feature importance $FI_\tau^{Second}(x_r)$ measures the extent of non-linear relationships of an input variable r and $FI_\tau^{Joint}(x_{rs})$ quantifies the strength of joint effects of two variables r and $s = 1, \ldots, p$ (interactions). We do not calculate the direction of impact, as the direction of joint impacts are tedious to disentangle. Rather, we are more interested in the question whether there is a joint-impact and it's potential strength.[9]

$$FI_\tau^{Second}(x_r) = \sqrt{\frac{1}{N}\cdot\sum_{i=1}^{N}\left(\frac{\partial^2 \hat{Q}_\tau(y_i|\boldsymbol{x}_i)}{\partial x_{ir}\partial x_{ir}}\right)^2}, \tag{8}$$

$$FI_\tau^{Joint}(x_{rs}) = \sqrt{\frac{1}{N}\cdot\sum_{i=1}^{N}\left(\frac{\partial^2 \hat{Q}_\tau(y_i|\boldsymbol{x}_i)}{\partial x_{ir}\partial x_{is}}\right)^2}. \tag{9}$$

If $FI_\tau^{Joint}(x_{rs})$ and $FI_\tau^{Second}(x_r)$ are close to zero in a quantile, we can negate single non-linear and joint impacts of the input variables. This leaves only a linear impact, which corresponds to the linear quantile regression.[10] Hence, we expect the quantile loss of QRNN and QR to be very similar. This allows us to explain differences in performance, as we can trace them back to $FI_\tau^{Joint}(x_{rs})$ or $FI_\tau^{Second}(x_r)$ and answer why this approach is superior and which effects (joint or second) and variables are responsible.

At the end of this section, we briefly discuss why we choose the explainability methods in Horel et al. (2018) and Nagl (2021). Regarding PDP, ICE and ALE, it is very tedious for interpret the QRNN using plots. For example, as our output consists of 99 quantiles and we use 26 variables, we would have to interpret $99 \cdot 26 = 2,574$ plots for the main importances and $\frac{26\cdot(26-1)}{2}\cdot 99 = 32,175$ plots for interactions. The LIME and anchor approaches cannot be aggregated to global explanations and, thus, we cannot identify the (overall) main drivers of workout LGDs. The SHAP approach is from a computational perspective unfeasible in our application. Lastly, the approach in Horel and Giesecke (2020) is not valid for our quantile based definition of a neural network.

5. Empirical results

This paper compares the ability of forecasting the conditional distribution of LGDs using the QRNN. The main focus is a comparison with the linear quantile regression, for example used by Krüger and Rösch (2017). However, to give a very broad picture we include additional models. We compare our approach to beta regression, e.g., used by Yashkir and Yashkir (2013) and Gambetti et al. (2019), and fractional logit regression, e.g., employed in Bastos (2010) and Qi and Zhao (2011). Furthermore,

[9] The standardization with the constant C is also neglected for illustration purposes.

[10] Of course it is possible that LGD realizations are driven by higher order impacts (third or fourth order) or joint effects of more than two variables. However, our empirical results in Section 5 confirm that these effects are negligible.

we use regression trees used by Altman and Kalotay (2014) and Bellotti et al. (2021). Mixture models are known for their flexibility, especially for bimodal distributions. They are frequently used for LGD prediction, see Altman and Kalotay (2014) or Betz et al. (2018). Hence, they are solid challengers for the QRNN.[11]

One of the main challenges using neural networks is to find a suitable architecture. In various applications, a cross-validation strategy is used to find the optimal hyper parameters. This is very common in most machine learning applications in credit risk, see e.g. Bastos (2010); Hartmann-Wendels et al. (2014); Gambetti et al. (2020). However, in credit risk not only the out-of-sample (cross-sectional), but rather the out-of-time prediction is of major concern, as emphasized by Kalotay and Altman (2017). Hence, we also address this issue in the model calibration by splitting our training data into $k = 5$ time buckets. We call this procedure k-fold Time Validation.[12] Furthermore, we use so-called Dropout Layers based on Srivastava et al. (2014) and L1-Regularization of parameter weights to avoid overfitting and increase the robustness of our model.

Another issue which arises for most machine learning methods is the suspicion that the hyperparameters are tuned until the method beats the benchmark model. This optimization is analogous to the p-hacking problem in classical statistics. To tackle this issue, we calculate the Spearman correlation coefficient between the estimated in-sample target function, defined in Eq. (6) and for the two out-of-time samples in each k-fold Time Validation. If this coefficient is statistically different from zero and positive, it means that a reduction of in-sample target function implies a reduction of the out-of-time target function as well. We choose $\tau = 0.01, 0.02, \ldots, 0.98, 0.99$ leading to 99 quantile estimates for each observation in our dataset.[13]

The QRNN network is special as the architecture has more output neurons (quantiles) than input neurons (features). Therefore, we choose a so-called "baseline" structure, which ensures that the number of nodes in the hidden layers increases from one to the other, adopting the strategy in Gu et al. (2020). The baseline of the first hidden layer contains eight neurons, whereas we use 16 neurons for the second hidden layer. The largest configuration with a multiple of eight would result in 64 neurons in the first and 128 neurons in the second hidden layer. We opt against using more than two layers to avoid the vanishing gradient problem, see Glorot and Bengio (2010). It is well known that from the universal approximation theorem, following Cybenko (1989) or Hornik (1991) among others, that a single layer neural network can approximate any continuous function. However, Rolnick and Tegmark (2018) show exemplarily that adding more layers is more efficient with the same approximation property. We use sigmoid and tanh as activation functions.

Table B.1 in Appendix B provides the selected ranges for the hyperparameters of the QRNN and a more detailed description of the hyperparameter search.

Table 2 shows the resulting architecture in both regions which are very similar. For example, in both subsamples a *learning rate* of 0.001 provides the best alternative. A more interesting determinant is the *multiple*, which controls the shallowness and therefore to some extent, the complexity of the neural network. In both regions, a rather small value is selected. This may indicate two things. First, the 5-fold Time Validation ensures that the complexity does not go off the rails and second, that there is no need for extremely broad networks.

In the US American sample, a network with two *hidden layers* seems to be best, whereas a less deep and more shallow network fits better to the European sample. Comparing the in-sample with the out-of-sample target functions, we recover the discrepancies in the distributions. The value in the US American sample increases along the time line. On the contrary, the highest value in the European sample appears in the GFC, whereas the in-sample and After-GFC numbers are very similar. To investigate whether our validation strategy is robust, we provide the Spearman's ρ between in-sample and the two corresponding out-of-sample target functions for the 100 best specifications, see Table 3. For each 5-fold Time Validation subsample, a model is fitted based on four folds and the target function of the remaining fold is calculated. Subsequently, we calculate the target function values for the two out-of-time periods. All three values are stored for each parameter combination and averaged over the five repetitions. A positive value indicates that a very good model based on our validation strategy probably performs well in the out-of-time sample. For both out-of-time periods, we obtain ρ statistically different from zero with a positive sign.

The majority of the literature focuses on mean predictions. For example, the standard coefficient of determination R^2 and the Pearson ρ are common measures to evaluate the fit of a model. They only consider the mean, not capturing the distributional characteristics of workout LGDs. However, to be in line with the literature,

Table 2
5-fold Time Validation | Final values.

Parameter	USA	Europe
Learning Rate	0.001	0.001
Dropout	0.20	0.20
Multiple	1	2
L1 Loss	0.005	0.005
Hidden Layer	2	1
Activation	tanh	sigmoid
Epochs	150	100
Loss based on Eq. (6)| In Sample	7.2451	6.9115
Loss based on Eq. (6)| GFC	10.8921	10.3597
Loss based on Eq. (6)| After GFC	11.3331	7.2016

Note: The table shows the final values of the hyperparameter search. For each sample, an independent gird-search is employed. The results are comparable, although we observe differences in the number of hidden layers and the multiple.

Table 3
Spearman correlation coefficient of the top 100 target functions.

(a) USA

	IS	GFC	After GFC
IS	1	0.28**	0.34***
GFC		1	0.33**
After GFC			1

(b) Europe

IS	1	0.34***	0.40***
GFC		1	0.24***
After GFC			1

Note: The table shows the estimated Spearman's ρ. We test the hypothesis $H_0 : \rho = 0$ against the alternative $H_1 : \rho \neq 0$. $***, **, *$ means statistically significant at the 1%, 5%, and 10% levels, respectively. We use a rank-based correlation metric as we are more interested in the similarity of the hierarchical structure of the validation task than the correlation in losses.

[11] We would like to thank the two anonymous referees for suggesting this comparison, which has substantially improved our paper.

[12] We also evaluate another validating strategy by subsequently filling up the training set and validating on the next time bucket. For example in the first run, time bucket 1 is used for training and the model is validated on time bucket 2. Next, time bucket 1 and 2 are used for training and the model is evaluated on time bucket 3 and so on. The final model and all evaluation metrics are very similar.

[13] We also tried finder splits such as $0.001, 0.002, \ldots, 0.998, 0.999$, but did not find any substantial differences to the outcome of this analysis. If readers are interested in these results, please contact the corresponding author.

Table 4
Goodness of fit based on mean predictions.

Method	R^2			$\rho_{Pearson}$		
	In Sample	GFC	After GFC	In Sample	GFC	After GFC
(a) European Sample						
Quantile Regression	0.109	0.036	0.048	0.282	0.096	0.164
Quantile Regression Neural Network	**0.145**	**0.067**	0.061	**0.363**	0.151	0.227
Gaussian Mixture Model	0.139	0.029	0.041	0.238	0.124	0.233
Regression Tree	0.114	0.051	0.048	0.227	**0.164**	0.262
Beta Regression	0.099	0.052	**0.096**	0.293	0.081	**0.278**
Fractional Logit Regression	0.050	0.019	0.022	0.297	0.089	0.272
(b) US Sample						
	In Sample	GFC	After GFC	In Sample	GFC	After GFC
Quantile Regression	0.068	0.032	0.024	0.236	0.045	0.081
Quantile Regression Neural Network	0.070	0.038	0.043	0.245	**0.096**	0.086
Gaussian Mixture Model	**0.079**	0.037	**0.080**	**0.285**	0.073	**0.091**
Regression Tree	0.040	0.040	0.031	0.181	0.012	0.075
Beta Regression	0.077	**0.041**	0.030	0.233	0.066	0.057
Fractional Logit Regression	0.022	0.015	0.017	0.234	0.041	0.065

Note: This table shows on the left hand side the R^2 of the mean predictions for each method. Furthermore, on the right hand side the Pearson correlation coefficient between mean estimates and the true realizations of the LGD is displayed. The best value for each metric in each sample is indicated in bold. The mean predictions of the quantile regression and Quantile Regression Neural Network are calculated by taking the expectation over the estimated quantiles for every obligor. The number of components for the Gaussian Mixture Model is chosen according to the lowest AIC on the training data, following Altman and Kalotay (2014). For both regions, three components fit best, which is in line with the literature, see, e.g., Krüger and Rösch (2017). We optimize the hyperparamter (maximum depth, minimum samples required for the split and the minimum number of samples in a leaf node) of the regression tree using our Time Validation approach. To apply the beta regression and factorial logit regression, we transform the LGD values outside [0,1] with $LGD_{[0,1]} = \frac{LGD - min(LGD)}{max(LGD) - min(LGD)}$, following Krüger and Rösch (2017) and Altman and Kalotay (2014). In general, the QRNN shows competitive results in Europe, but the Gaussian Mixture Model seems to be more appropriate in the US if one focuses only on mean estimates, neglecting the fit over the full conditional distribution.

we report R^2 and $\rho_{Pearson}$ only for the sake of completeness and evaluate the goodness of fit for the entire distribution using more appropriate measures.

From Table 4, we can argue that the QRNN, although not constructed for mean predictions, provides competitive results in both regions. In the European dataset, the QRNN achieves the highest R^2 in sample and in the GFC, while the second best value after the GFC can be obtained. With respect to the $\rho_{Pearson}$ values, we achieve the highest correlation only in sample. In the US, the QRNN provides competitive results, but not the overall best performance. The results in the US American sample are in line with Krüger and Rösch (2017), that the Gaussian Mixture Model is a solid challenger and for some samples even outperforms the linear quantile regression. Overall, we can summarize that the QRNN provides also reasonable mean estimates. However, we would like to stress the point that a proper evaluation should be based on the whole distributional fit and not on one location parameter only.

To compare the distributional fit, we evaluate values from the quantile loss function in Eq. (4) to determine which quantile predictions fit best. We propose this as the best way to compare model performance, as the lower the loss is, the better the conditional quantile estimates are. To quantify the improvement compared to the linear quantile regression, we focus on the relative values of Eq. (4):

$$1 - \frac{\sum_{i=1}^{N} \rho_\tau \left(y_i - \hat{y}_{\tau,i}^{Alternative}\right)}{\sum_{i=1}^{N} \rho_\tau \left(y_i - \hat{y}_{\tau,i}^{QR}\right)}. \tag{10}$$

We estimate Eq. (10) for every quantile, illustrated in Fig. 3. The performance of the alternative model is better, if the value is greater than 0, e.g. a value of 0.3 means that precision of the quantile forecasts is 30% higher compared to the linear quantile regression. Focusing on the comparison of the QRNN and QR, values greater than zero can directly be ascribed to existing non-linear relationships and joint impacts between LGDs and covariates at this quantile level. This interpretation seems reasonable, as non-linearity and interactions are the only differences between both models. To exclude the chance that better results stem from the monotonicity constraint, we also estimated a linear QRNN model using no hidden layers, obtaining similar results.

The solid line represents the in-sample comparison, whereas the dashed line refers to the GFC and the dotted line to the After GFC period. In the US American sample, the precision improvement increases almost monotonically up to the 80% quantile. Hence, for higher LGD realizations more non-linearity and joint effects are present. The difference between QR an QRNN is up to 15%. If we focus on the out-of-time periods, we observe two interesting things. First, we see steady values up to the 80% quantile, but afterwards the superiority of the QRNN yields over 30% improvement. This means, that more extreme realizations in the out-of-time perspective can be better predicted by the QRNN. For the lowest quantile we have a superior performance of the QRNN as well. On the contrary, for lower quantiles from 20% to 40% in the After GFC sample, the QR is slightly better with relative values of about -2.0%.

In the European sample, we observe slightly different results. Both methods perform similar in-sample up to the median. Afterwards, the improvement of the QRNN increases sharply to values of around 25%. From the out-of-time perspective, the improvements start from the 60% quantile with values around 12%. Similar to the US sample, in the middle quantile in the After GFC sample the linear quantile regression has a small edge of -1.5% over the QRNN. In summary, Fig. 3 shows that the QRNN clearly outperforms in-sample the linear quantile regression, especially for high LGD realizations. In both out-of-time samples, we observe the superiority for high LGD realizations as well. Yet, there are some quantiles in the After GFC sample, in which we do not outperform the QR. Nevertheless, for the vast majority of quantiles and most importantly for high realizations, the QRNN substantially outperforms the QR. To provide evidence, that the QRNN is indeed a reasonable extension of the linear quantile regression, we compare our results to other common methods in the literature.

Table 5 shows the average improvement over the full conditional distribution compared to the linear quantile regression.

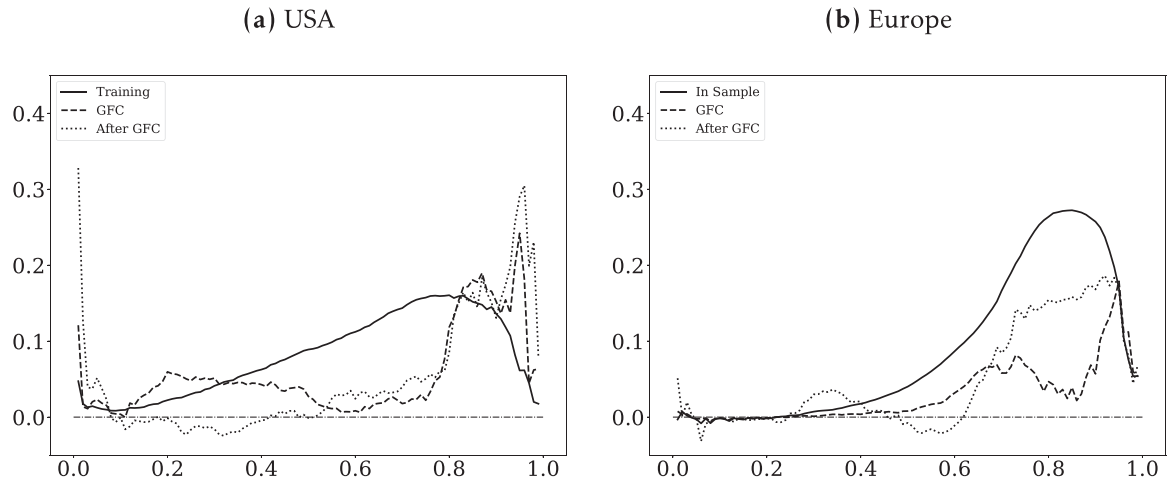

Fig. 3. Loss over quantiles. Note: These figures show the relative quantile-specific loss based on Koenker and Bassett (1978) for each quantile. The solid line refers to the training sample, the dashed line GFC and the dotted line to the after GFC sample. A value greater than zero indicates a better distributional fit of the QRNN.

Table 5
Loss over quantiles in comparison to other methods.

(a) European Sample

Method	Full range			Right tail		
	In Sample	GFC	After GFC	In Sample	GFC	After GFC
Quantile Regression Neural Network	**0.087**	**0.029**	**0.049**	**0.221**	**0.070**	**0.146**
Gaussian Mixture Model	0.038	-0.008	0.036	0.138	-0.016	0.092
Regression Tree	-0.279	-0.182	-0.296	0.058	0.067	0.070
Beta Regression	-0.122	-0.092	-0.151	0.085	-0.009	0.088
Fractional Logit Regression	-0.325	-0.526	-0.261	0.010	-1.393	-0.500

(b) US Sample

Method	Full range			Right tail		
	In Sample	GFC	After GFC	In Sample	GFC	After GFC
Quantile Regression Neural Network	**0.080**	**0.056**	**0.049**	**0.123**	**0.134**	**0.156**
Gaussian Mixture Model	0.051	-0.010	-0.051	0.076	-0.068	-0.249
Regression Tree	-0.157	-0.125	-0.158	0.008	0.012	0.075
Beta Regression	-0.025	-0.082	-0.094	0.044	0.050	-0.177
Fractional Logit Regression	-0.208	-0.407	-0.377	-0.014	-0.526	-1.024

Note: This table shows the average values of Eq. (10) for each method. A value larger than 0 indicates a better distributional fit compared to the linear quantile regression. The best value in each sample is indicated in bold. To put more emphasis on larger LGD realization, the quantile fit is also calculated for $\tau \in (0.75, 0.99)$, labelled as "right tail". The result shows that the QRNN outperforms all benchmark models, especially for the right tail. The quantiles of mean-focussed methods are calculated following Krüger and Rösch (2017).

Overall, we can summarize that the QRNN provides the largest improvement, which may be attributed to non-linear and joint effects. More interestingly, we recover the evidence of Fig. 3 that especially on the right tail, i.e., for $\tau \in (0.75, 0.99)$, the QRNN performs very well. We also recover findings of previous studies, that the Gaussian Mixture Model performs reasonably well and in some cases outperforms the linear quantile regression (see e.g. Krüger and Rösch, 2017).

Next, we analyse the monotonicity among predicted quantiles. In general, increasing quantiles could be guaranteed by construction, if they are extracted from a parametric model, such as Beta Regression or Gaussian Mixture Models. However, this entails model risk, if the model is not flexible enough to capture all characteristics of the target distribution. With respect to Table 5, we can see that the (non-parametric) QRNN provides the overall best fit. We include a monotonicity penalty to combine the advantages of non-parametric estimation of conditional distributions while ensuring (almost) monotonic increasing quantiles. Fig. 4 illustrates a random sample of 100 different LGD distributions of the training sample. The left-hand side shows the QRNN and on the right-hand side the QR. Comparing both methods in the US, we see the impact of the monotonicity penalty in Eq. (5). The QRNN approach yields monotonic quantiles, whereas the linear model shows heavily deviating quantiles. This is also true for the European dataset. However, for the outer right tail, e.g. from 95% to 98%, the monotonicity is not necessarily met in the QRNN approach.[14] An easy solution to the small range of non-monotonic quantiles could be the approach by Chernozhukov et al. (2010), i.e., simply rearrange the quantile to ensure monotonicity.

Fig. 5 displays the estimated values for $FI_\tau^{First}(x_r)$. To the best of our knowledge, this paper is the first to disentangle the great flexibility of neural networks in single and joint impacts of LGD drivers. We estimate the feature importance $FI_\tau^{First}(x_r)$ (vertical axis) for each quantile τ (horizontal axis) to discover the impact on each quantile. The sum of all absolute feature importances $\sum_{r=1}^{p} |FI_\tau^{First}(x_r)|$ sums up to 1 in each quantile. For both regions, all feature importances are aggregated in a stack plot to give a comprehensive overview. The ordering coincides with the sequence of the displayed plots. We find that the impact of all feature im-

[14] We tested various other specifications of the penalty and increased the weight of the monotonicity loss in the optimizing problem and increased the number of estimated quantiles. Nevertheless, compared to our benchmark model, the estimated distributions look much more reasonable.

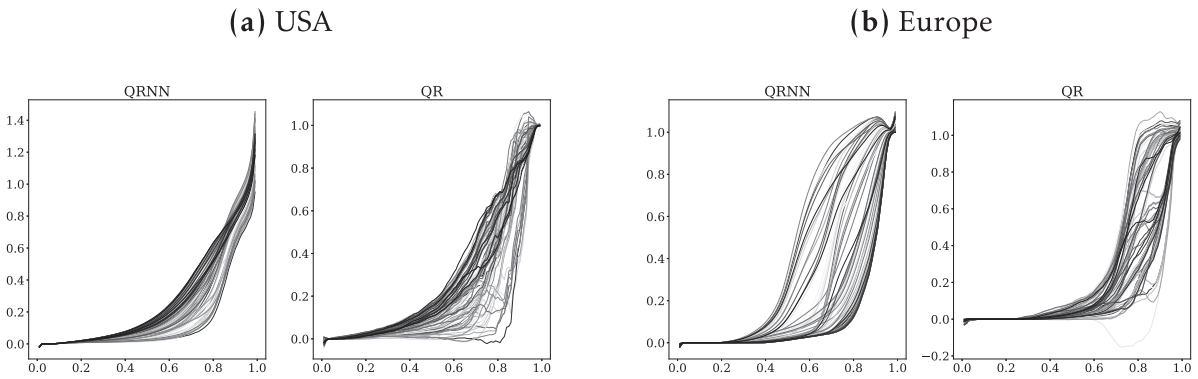

Fig. 4. CFD-Plot. Note: These figures show a sample of 100 different estimated LGD cumulative distribution functions. The left panels show the results of the QRNN approach, whereas the right panel the linear method. Within every region, we select the same loans to allow for comparison. By definition, a cumulative distribution function must be monotonic in quantiles, which is guaranteed by the QRNN to a large extent. On the contrary, the linear quantile regression shows a non-monotonic behaviour for a wide range.

portances is stable, but for some we find differences especially in the tails of the conditional distribution.

Concerning the US, a positive value for log(EAD) means that the larger the log(EAD) of the loan, the higher the resulting quantile forecasts. The values around 0.07 indicate that the log (EAD) has a share of 7% of the total importance. In general, the higher the absolute value, the more important the feature is for quantile forecasts. Economically, the positive values of $FI_{\tau}^{Frist}(x_r)$ for log(EAD) indicate that banks potentially give credit to companies who cannot finance on capital markets and, hence, need larger loans from banks. As these companies may have a higher tendency to default, we observe higher LGDs. This coincides with the positive feature importance of large corporates.

Another interesting feature is Collateral- RE with a negative effect. This means if a loan is protected by a real estate collateral, the LGD is lower. The single most important loan specific feature is the dummy variable Senior. The reference group is pari-passu, hence, we can infer that if a bank is senior in the resolution process, the

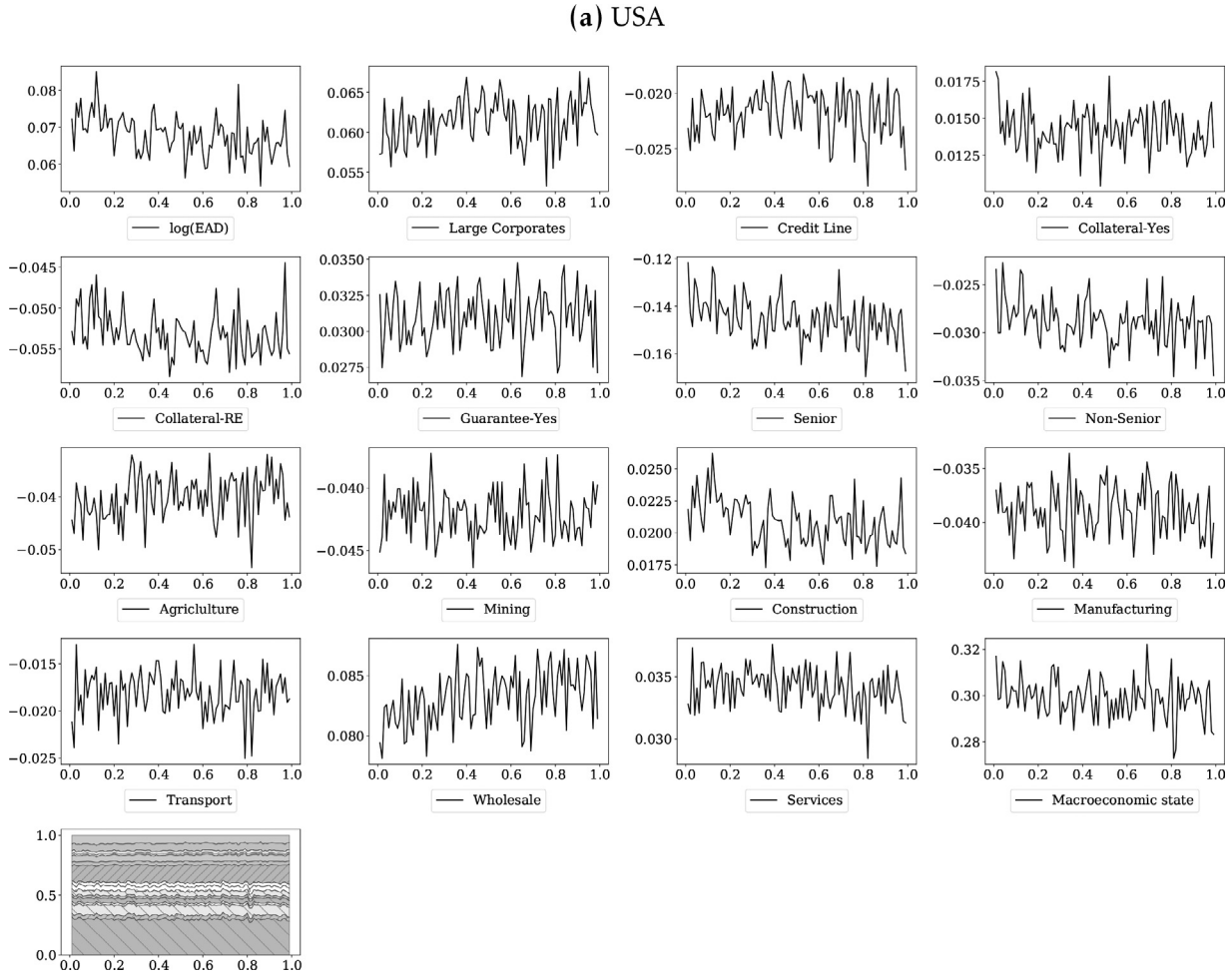

Fig. 5. Feature Importance. Note: These plots show the estimated feature importance of every variable for every quantile. Please note that the sum of all feature importances for each quantile sums up to 1. The last plot illustrates the importance of all variables in a stacked fashion. This allows us to evaluate the overall importance for LGD prediction at a glance.

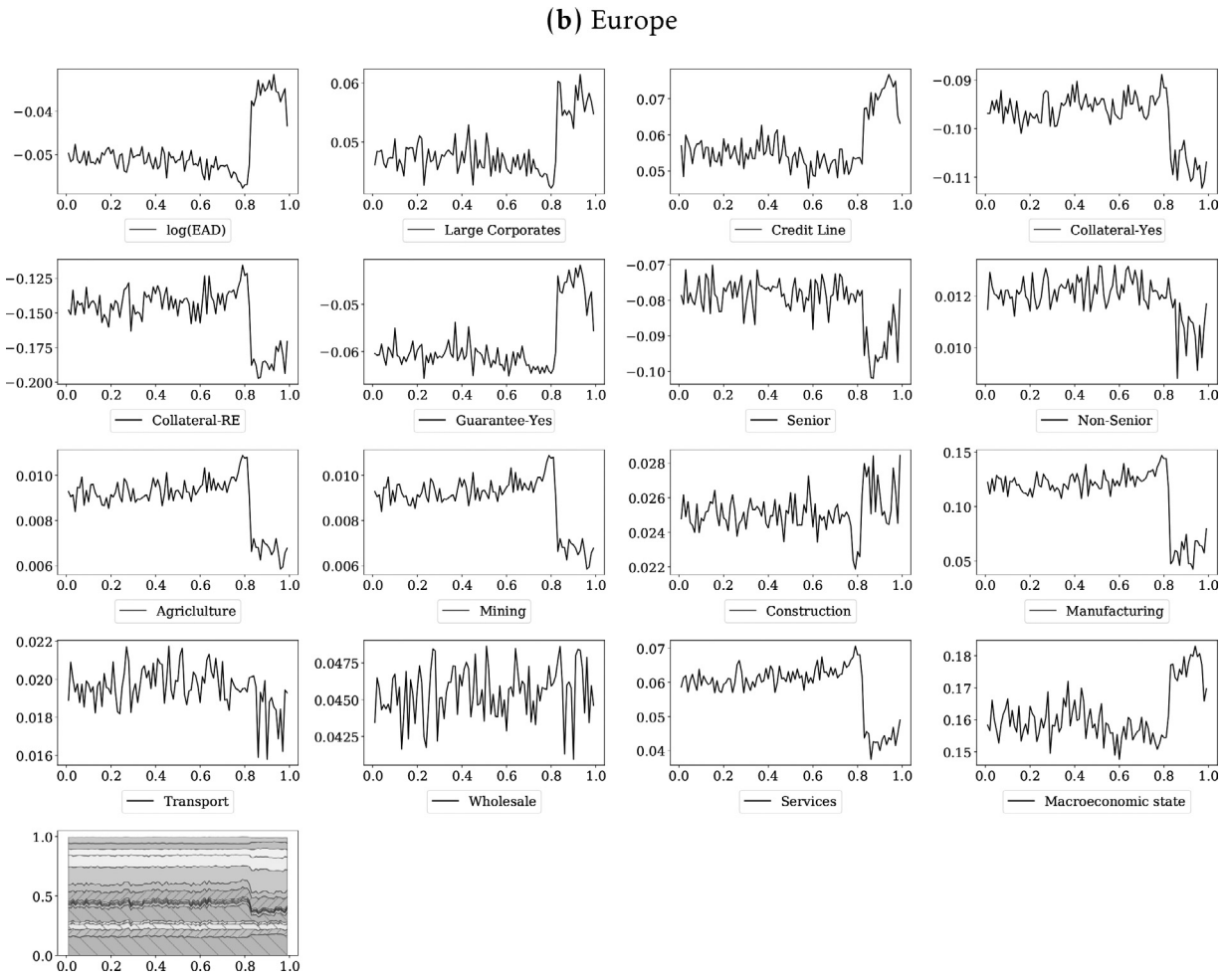

Fig. 5. Continued

LGD realizations are lower. Abstracting from the estimated value, the dummy variable Senior has a share of 14% of the total importance.

As mentioned in Section 3, we capture the systematic variation via principal components. As the direction and the exact meaning of every single component are hard to interpret, we focus on the overall impact. Hence, the plot labelled as "macroeconomic state" displays the absolute sum of all principal components. From the absolute values of around 30% we can see that the macroeconomy is the most important determinant of workout LGDs in the United States. From these values we can infer that the macroeconomy has a quadruple impact on the LGD distribution compared to the log(EAD) (30% vs. 7%). Hence, the economic surrounding is four times more important than the loan value itself and twice as important as a senior rank in the resolution process (30% vs. 14%). As robustness, we also used a PCA with 90% and 99,5% of variance, but the overall effect is stable around 30% in any setting.[15] This result coincides with evidence that US American LGDs are very cyclical and the macroeconomy is an important driver, see Tobback et al. (2014); Betz et al. (2018); Nazemi and Fabozzi (2018).

The results are somewhat different in the European sample, in which the single most important determinant is the protection of the loan, i.e., the combined impact of the collaterals of around 30%. Both feature importances have a negative sign, indicating that a protection decreases the losses of banks. If the loan is protected by real estate, the resulting LGD is decreased considerably, especially in high quantiles. Regarding the impact of the macroeconomy we find a considerable impact on the LGD prediction. Interestingly, the impact increases in quantiles, which indicates that especially for higher LGD values, the economy is more important. We find that the collateralization has a importance twice as high compared to the macroeconomy (30% vs 15%) and roughly four times higher than the senior rank in the resolution process (30% vs 7%). Moreover, we observe a shift in importance for quantiles above 80% to the macroeconomy, facility asset class, collaterals, high seniority and credit lines. This can be seen as evidence that high LGD realizations are driven by these five key drivers. Their relative importance sums up to roughly 65% over these quantiles.

Besides the first order importance, the second order and joint impacts are also interesting. The following analysis helps us to understand why the QRNN is superior in some quantiles and which variables drive this superiority. Figs. 6 and 7 show the absolute sum of all pairwise interactions with all other variables for every input variable, $FI_{\tau}^{Joint}(x_{rs})$, illustrated by the dashed line and the second order effect, $FI_{\tau}^{Second}(x_r)$, via the solid black line. The last subplots in both figures show the relative importance of both, second order and joint impact, for every variable. The left y-axis

[15] The conclusions about the importance of the macroeconomy in Europe and USA are similar when using less (more) principal components. Hence, the high importance comes not from the number of components used in the analysis. This may also seen as a robustness of our employed feature importance measure, as the overall importance does not seem to be sensitive to the number of components. The results with less (more) components are available upon request.

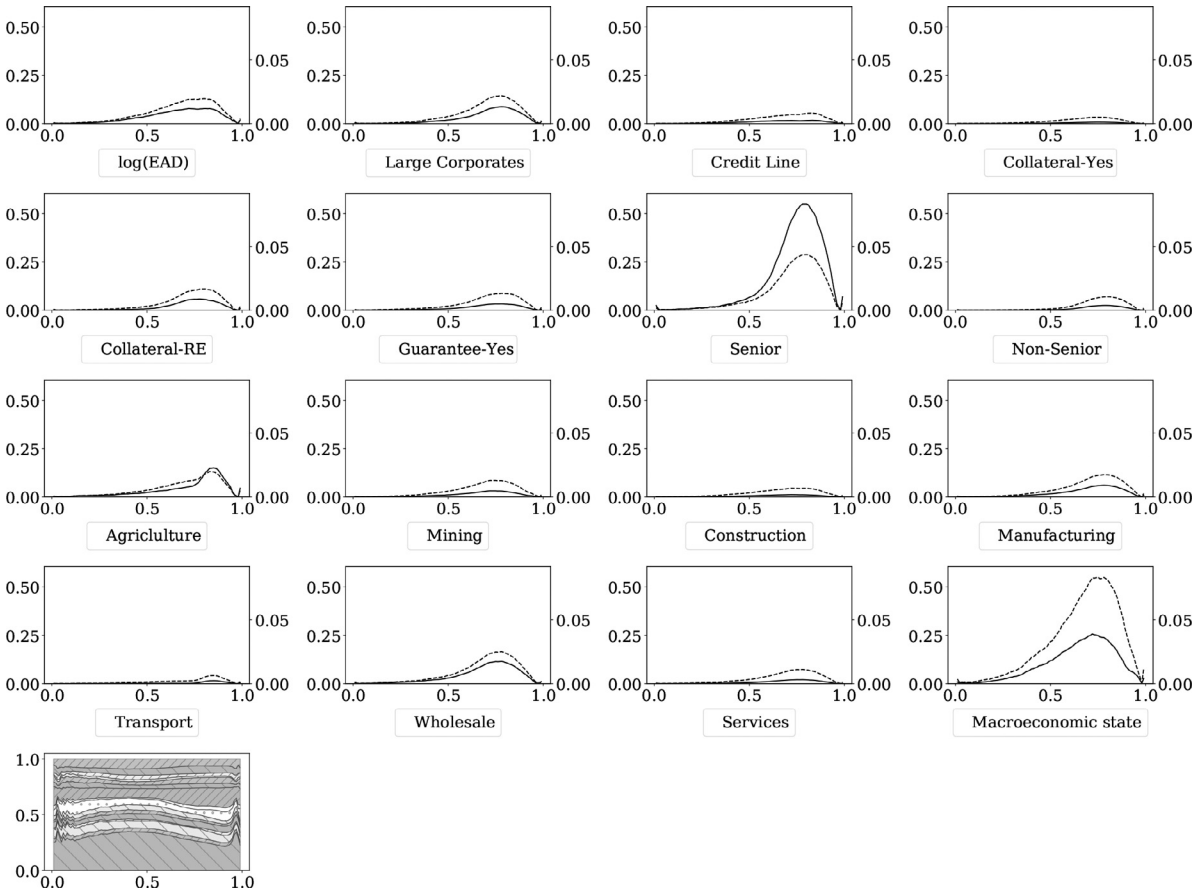

Fig. 6. $FI_{\tau}^{Joint}(x_{rs})$ and $FI_{\tau}^{Second}(x_r)$ | US sample. Note: These plots show the estimated importance of every variable for every quantile. The horizontal line shows the different quantiles, whereas the vertical line illustrates the importances. The left y-axis and the dashed line refers to $FI_{\tau}^{Joint}(x_{rs})$, whereas the right y-axis and the solid line refers to $FI_{\tau}^{Second}(x_r)$. This allows us the evaluation of the overall joint effects for LGD prediction at a glance.

corresponds to the values of $FI_{\tau}^{Joint}(x_{rs})$ and the right y-axis to $FI_{\tau}^{Second}(x_r)$.

From the last plot in Fig. 6 for the United States, we can see that the principal components account for roughly 35% of these effects. The dummy variable Senior also shows considerable joint effects with all other variables. However, the macroeconomic state has a value roughly twice as high for joint effects and seems to be the overall most interacting determinant.[16] Regarding the second order effects, we observe the highest value with the variable Senior. This indicates that the rank in the resolution process has a strong non-linear relation to higher quantiles. The remaining variables show no substantial effects. Overall, we can see that the joint effects are clustered into two important drivers. Furthermore, important information can be obtained from quantiles, where these effects are close to zero. If we compare these ranges with the relative target functions in Fig. 3, we can see that these coincide with the quantiles where QR and QRNN have a similar target function. Hence, we can see why the QRNN does not outperform the QR in these quantiles. There are simply no joint or second order effects for any input variable. Hence, we conjecture that there is probably only a linear impact, very well described by the linear quantile regression.

A similar picture can be seen in Europe in Fig. 7, where we have almost zero values for joint and second order effects for quantiles up to the median. This coincides with the similar performance regarding the target function, outlined in Fig. 3. Regarding joint effects, we find two important things. First, the overall magnitude is higher in Europe, illustrated by the lager range of the vertical axes. This means that there are overall higher joint effects in Europe than in the United States. Second, the joint effects are far more distributed across the different input variables. Hence, we observe more than two main drivers of joint effects. Similar to the United States, the macroeconomy shows the highest value of joint effects, but in comparison to other variables the relative magnitude is not as big as in the United States. Especially, if we sum up the values for the collateralization, we get very similar numbers. With respect to second order effects, the collateralization shows the largest non-linear impact on the LGD distribution for higher quantiles. Concluding, we find that the macroeconomy has the highest joint effects with other variables, closely followed by the collateralization. Finally, we analyse which variables have the largest joint effects with the macroeconomic state. Therefore, Figs. C.1 and C.2 in Appendix C show the absolute sum of interactions for all eight, respectively eleven, components for every remaining variable. Thus, we can disentangle the dashed line of the macroeconomic state in Figs. 6 and 7 into its elements. For the sake of clarity, we moved these figures to the Appendix. In the United States, the variable Senior has the largest joint effects with the macroeconomy, especially for higher quantiles. This may be plausible as in economic downturns the rank in the resolution process is very important as the overall proceeds from bankruptcies may decrease and, thus, there is nothing left for lower ranks in the resolution process. In the European dataset, we observe that the largest joint effects are between the macroeconomy and the collateral of the loan. From an economic perspective, we can argue in the same way as for the seniority in the United States. In downturns it is probably more

[16] Again, the drawn conclusions for Europe and USA about second order and joint effects are similar when using less (more) components in the macroeconomic state. The results are available upon request.

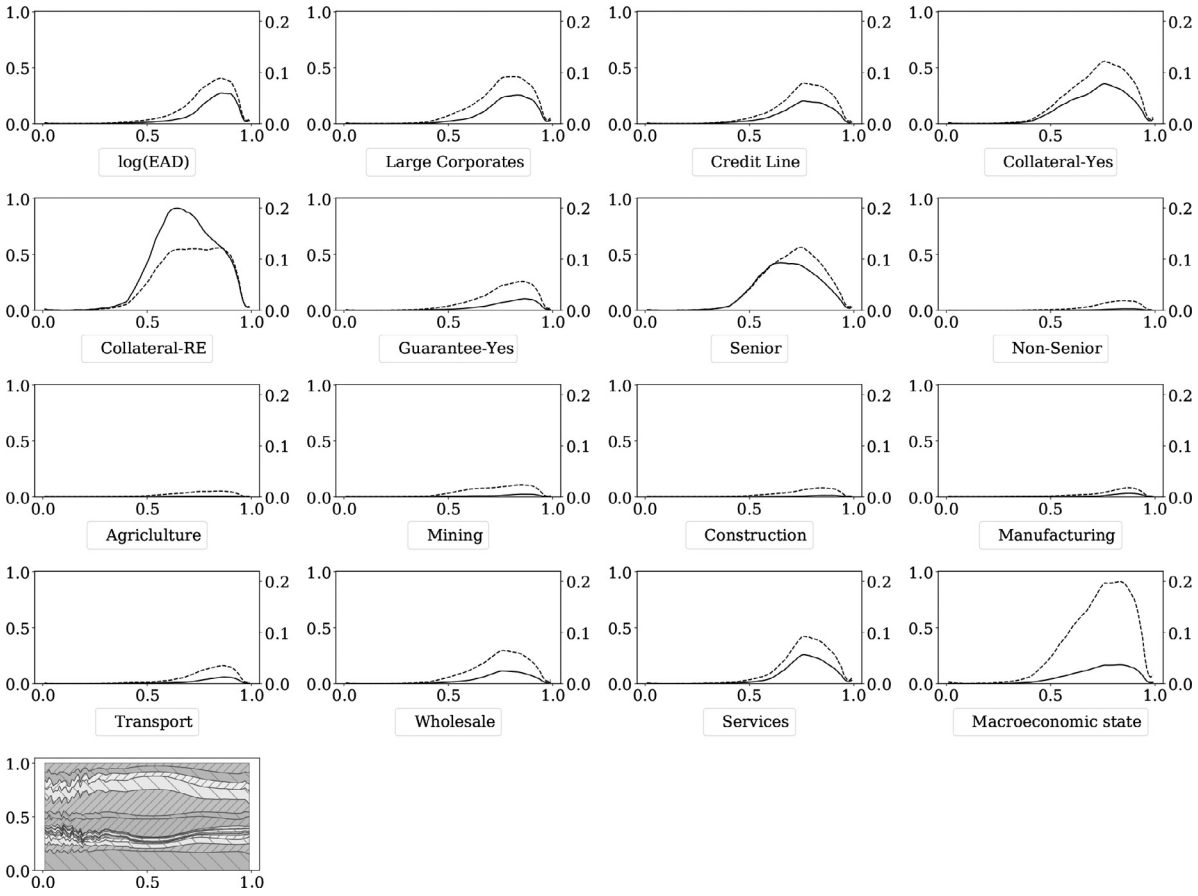

Fig. 7. $FI_{\tau}^{Joint}(x_{rs})$ and $FI_{\tau}^{Second}(x_r)$ | European sample. Note: These plots show the estimated importance of every variable for every quantile. The horizontal line shows the different quantiles, whereas the vertical line illustrates the importances. The left y-axis and the dashed line refers to $FI_{\tau}^{Joint}(x_{rs})$, whereas the right y-axis and the solid line refers to $FI_{\tau}^{Second}(x_r)$. This allows us the evaluation of the overall joint effects for LGD prediction at a glance.

important to have a collateral for the loan as the proceeds from bankruptcies, excluding collaterals, might be lower and, thus, it probably reduces the losses faced by the bank. Overall, we find that collateralization and the macroeconomy have the highest joint effects on the LGD distribution.

An interesting and actively debated topic is whether machine learning methods can be used for regulatory purposes. This question cannot be answered yet, as no final regulations are published. However, numerous workshops and discussion paper have emerged over the last years, see e.g. Basel Committee on Banking Supervision (2019); Deutsche Bundesbank (2020); Paulsen et al. (2021). For a detailed discussion of various aspects of applying machine learning fo regulatory purposes, we refer to Fritz-Morgenthal et al. (2021). In general, explainability methods play a major role in this discussion. For example, Deutsche Bundesbank (2020) state that explainability methods are a promising answer to black-box models, making them less opaque and more comprehensible. From this standpoint, the presented feature importance in this paper is a step into the right direction, as we do not only obtain the main (first-order) importances, but also second-order and joint importances. This may be helpful in the discussion with regulators, as we are able to explain why the machine learning method is superior (non-linearity and joint effects) and we can attribute this to specific variables (macroeconomy and collateralization).

Another important regulatory aspect is the ability to generate downturn estimates, which is extensively researched, see e.g. Calabrese (2014); Krüger and Rösch, 2017; Betz et al. (2018). We argue to use realizations of the PCA components during an economic downturn as a reasonable adverse scenario. To do so, we randomly sample 10,000 portfolios containing 500 obligors each, defaulted from 2010 to 2016. We use the QRNN to predict the conditional distribution of every obligor, calculate their expected LGD using the mean over the conditional distribution and finally aggregate this on portfolio level using the mean over all obligor's expected LGDs. Formally defined as:

$$\widehat{LGD}^{Portfolio} = \frac{1}{500}\sum_{i=1}^{500}\left(\frac{1}{99}\sum_{t=1}^{99} Q_{\frac{t}{100}}(y_i|\boldsymbol{x}_i)\right) \tag{11}$$

In the baseline scenario we use the true PCA component realization, whereas in the adverse scenario we use the PCA component realization during a quarter of economic downturn. Table 6 shows the mean of Eq. (11) over all 10,000 simulated portfolios.

Although we did not calculate the direction of impact for the PCA components in Figs. 6 and 7, we can now clearly see how the macroeconomic state impacts the predicted LGDs. In the US American sample, the estimated LGDs over all 10,000 portfolios increase by roughly 28% reflecting downturn conditions. The increase in the adverse scenario is even more pronounced in the European dataset. This can be attributed to the fact, that the difference between average LGDs in the GFC and after the GFC is much more pronounced than in USA. Hence, we see a downturn LGD that almost doubles the baseline LGD. Admittedly, our strategy relies on historical economic downturns and is difficult to extend to unprecedented scenarios. Nevertheless, this analysis shows that the QRNN is able to provide reasonably high downturn estimates, reflecting the impact of the macroeconomic state during economic downturns.

Summarizing, the contribution of this paper is a straight forward extension of the economically useful method of quantile regression into several important directions. Furthermore, we com-

Table 6
Downturn estimates.

Quarter	tex	USA			textt	Europe		
		Baseline	Adverse	Increase		Baseline	Adverse	Increase
		0.271				0.201		
2007 Q4			0.338	24.78%				
2008 Q1			0.325	19.93%			0.330	64.17%
2008 Q2			0.346	27.68%			0.313	55.72%
2008 Q3			0.345	27.31%			0.340	69.15%
2008 Q4			0.323	19.19%			0.343	70.65%
2009 Q1			0.344	26.94%			0.401	99.50%
2009 Q2			0.346	27.68%			0.385	91.54%
2009 Q3							0.362	80.10%

Note: This table shows the expected LGD value assuming the PCA realization in the given quarters. The quarters are chosen according to the OECD crisis indicator, which indicates that the GFC lasts from 2007 Q4 to 2009 Q2 in the USA, whereas it is slightly shifted in Europe (2008 Q1 to 2009 Q3). The values are obtained by randomly drawing 10,000 portfolios containing 500 obligors each. The column *Baseline* represents the average of the predicted LGD value using the true PCA realizations for each obligor and, thus, is identical over the quarters. The column *Adverse* reports the average predicted LGD value assuming the PCA realizations of the given quarter for all obligors. The column *Increase* shows the percent increase of the mean LGD value of the adverse scenario compared to the baseline values. A positive value indicates larger LGDs due to the adverse PCA realizations.

pare the QRNN to various other methods and obtain superior results regarding quantile specific forecasts. The neural network approach shows a considerably better distributional fit for the whole LGD distribution, especially in the out-of-time perspective with an improved precision of the quantile forecasts of up to 30%. We find that classical quantile regressions provide non-monotone estimates of the LGD's conditional distribution, contrary to the QRNN. With respect to the important drivers of LGDs, we see diverging determinants in USA and Europe. In the US, the macroeconomic state is the single most important determinant. Seniority and the economic surrounding have the largest joint effects. The European LGD predictions rely more on loan-specific characteristics, such as different types of collateral. However, the largest joint effects are driven by the macroeconomy and the collateralization.

6. Conclusion

Recent literature shows that modelling the entire LGD distribution or quantiles thereof is more adequate than focussing just on means, and linear quantile regression outperforms ordinary least squares, fractional response regressions, beta regressions, regression trees and finite mixture models (see Krüger and Rösch, 2017). We extend this approach by allowing non-linearities and interactions in quantiles accomplished by the Quantile Regression Neural Network. This approach considerably enhances the modelling flexibility. The additional flexibility pays off in terms of a better distributional fit for in- and out-of-time samples with an improvement of up to 30% in quantile forecast precision. Machine learning models are prone to the conjecture that the researcher tries many different combinations and only reports the best, without ensuring that there is a broad superiority and to some extent robustness with respect to the architecture. We alleviate this problem by reporting positive Spearman's ρ, indicating that a good model in-sample is also a good model out-of-sample. Furthermore, we show that a monotonicity constraint can easily be implemented and standard linear quantile regression does not ensure monotonously increasing distribution functions. This also allows banks to use the QRNN on a loan level basis.

To the best of our knowledge, this paper is the first in credit risk to disentangle the impact of variables in neural networks in a highly interpretable fashion. The first order feature importance measure in Horel et al. (2018) and Nagl (2021) allows us to quantify the relative importance of all features and calculate the direction of their impact. We find that macroeconomic variables account for up to one third in the US American sample, underlining the dependency of LGDs on the economic surrounding. On the contrary, the largest first order feature importance in the European dataset were collaterals and hence, loan characteristics. The macroeconomy accounts for only 10–15% of the overall importance. Therefore, we document highly diverging impacts with respect to the macroeconomy in these two regions. This may give further evidence that systematic behaviour, expressed by macroeconomic variables, is clearly different in the US and Europe. By using the second order and joint impact feature importance measure, we can see why the QRNN outperforms its counterpart. We quantify strong joint effects of the macroeconomy with other variables as the main driver of the superiority. The contributions of this paper may have important implications for credit risk management, as in Europe and the United States the QRNN approach provides a higher precision in terms of quantile forecasts especially for higher quantiles. This suggests non-linear behaviour in quantiles of high LGD realizations. Furthermore, the empirical findings of high dependency of US American LGDs on the macroeconomy may have serious implications for banks and regulators to carefully account for this large impact. This points towards a highly cyclical behaviour of LGDs, which may result in higher losses of the US American banking system, especially in crisis periods.

Machine learning methods are often seen as problematic due to their black-box character, particularly from a regulatory perspective. The used feature importance measures are easy to implement and interpret and may enhance the adoption of machine learning approaches

CRediT authorship contribution statement

Ralf Kellner: Conceptualization, Methodology, Formal analysis, Writing – review & editing. **Maximilian Nagl:** Conceptualization, Methodology, Data curation, Software, Formal analysis, Writing – original draft. **Daniel Rösch:** Conceptualization, Methodology, Resources, Formal analysis, Writing – review & editing.

Acknowledgements

We thank participants of the 14th International Conference on Computational and Financial Econometrics 2020 in London and the Global Credit Data European Conference 2021 for fruitful discussions and helpful comments. Furthermore, we thank two anonyms referees for their comments which have substantially improved the paper.

Appendix A. Macroeconomic variables for the principal component analysis

Table A.1
Macroeconomic variables for US American loans.

(a) USA

Variable	Source
Economic Political Uncertainty | Three Component Model	https://www.policyuncertainty.com
Economic Political Uncertainty	https://www.policyuncertainty.com
Financial Stress Indicator	https://www.policyuncertainty.com
US Equity Market Volatility Index	https://www.policyuncertainty.com
Geopolitical Risk Index	https://www.policyuncertainty.com
Economic Uncertainty Related Queries	https://www.policyuncertainty.com
Financial Uncertainty Index	https://www.sydneyludvigson.com
Macroeconomic Uncertainty Index	https://www.sydneyludvigson.com
Real Uncertainty Index	https://www.sydneyludvigson.com
Unemployment Rate	https://fred.stlouisfed.org
Real Gross Domestic Product (yoy growth)	https://fred.stlouisfed.org
S&P/Case-Shiller U.S. National Home Price Index (yoy growth)	https://fred.stlouisfed.org
Industrial Production: Total Index US	https://fred.stlouisfed.org
Consumer Price Index for All Urban Consumers: All Items in U.S. City Average	https://fred.stlouisfed.org
CBOE Volatility Index: VIX	https://fred.stlouisfed.org
SP500	EIKON
M2 (yoy growth)	EIKON
TED Spread	https://fred.stlouisfed.org
Term Spread 10y-3m	https://fred.stlouisfed.org
Commercial and Industrial Loans, All Commercial Banks (yoy growth)	https://fred.stlouisfed.org

(b) Europe

Variable	Source
Economic Political Uncertainty Europe	https://www.policyuncertainty.com
Harmonized Unemployment Rate: Total: All Persons for the Euro Area	https://fred.stlouisfed.org
Real Gross Domestic Product for Euro area (yoy growth)	https://fred.stlouisfed.org
Real Residential Property Prices for Euro area	https://fred.stlouisfed.org
Total Industry Production Excluding Construction for the Euro Area	https://fred.stlouisfed.org
Consumer Price Index: Harmonized Prices: Total All Items for the Euro Area	https://fred.stlouisfed.org
VSTOXX Europe	EIKON
EUROSTOXX 50	EIKON
M2 Europe (yoy growth)	EIKON
Total Loans to Corporate Euro Area (yoy growth)	https://www.euro-area-statistics.org
Business Survey Industry	https://ec.europa.eu/eurostat
Business Survey Construction	https://ec.europa.eu/eurostat
Economic Sentiment Indicator	https://ec.europa.eu/eurostat
Business Climate Indicator	https://ec.europa.eu/eurostat
International Trade (yoy growth)	https://ec.europa.eu/eurostat
Labor cost nominal value	https://ec.europa.eu/eurostat
Turnover in industry, total - quarterly data	https://ec.europa.eu/eurostat
Building Permits	https://ec.europa.eu/eurostat

Note: The table shows the employed macroeconomic variables for the principal component analysis. They are in line with papers in the literature concerning the estimation of workout LGDs of corporate loans and some variables which also may be suitable to account for variations of workout LGDs over the business cycle.

Appendix B. Hyperparameter Optimization

This section elaborates in more detail how the QRNN relates to the linear quantile regression and the way we optimized the hyperparameters. Fig. B.1 shows the difference between the standard linear quantile regression (QR) on the left-hand side and the QRNN approach on the right-hand side.

The standard quantile regression relates all input features directly and linearly to only one quantile of interest. To describe a full set of quantiles, one has to fit one separate linear quantile regression to each of them, resulting in 99 models overall for our empirical analysis. On the right-hand side, the QRNN approach is illustrated. We use the same set of input variables and estimate the same set of quantiles, but there are three important differences compared to the standard quantile regression. First, we model the full discrete set of quantiles at once, reducing the required models to only one single model. Second, the relation between input features and quantiles is no longer direct, but described by non-linear transformations in the hidden layers of the QRNN. This allows for all kinds of non-linearity to be present in all quantiles simultaneously. Third, as the QRNN approach models the full set of quantiles, we can penalize non-monotonic quantile estimates, i.e., the estimated value must increase from the top ($Q_{0.01}$) to the bottom ($Q_{0.99}$). This is not possible in the standard QR approach as the models are fitted independently.

The QRNN network is somewhat special as the architecture requires that we have more output neurons (quantiles) than input neurons (features). Therefore, we choose to assume a so-called “baseline” structure, which ensures that the number of features in the hidden layers increases from one to the other. In classical applications, where only one output neuron is used, e.g., predicting probability of default or market returns, it has turned out that reducing the number of neurons in the hidden layer by half for each additional hidden layer seems

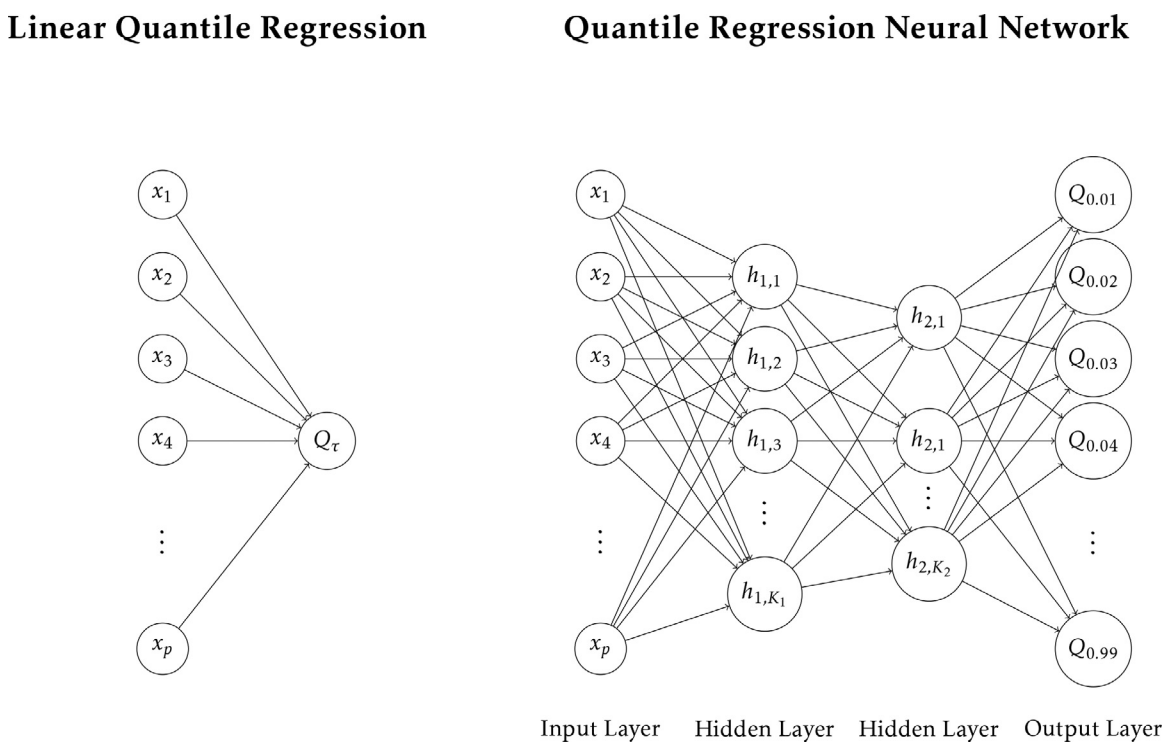

Fig. B.1. Graphical overview | QR vs. QRNN. Note: This figure shows the setup of the standard linear quantile regression and the QRNN. There are three important enhancements compared to the QR. First, we model all quantiles of interest at once. Second, we can easily allow for non-linearities in each of the quantiles. Third, we provide more monotonic quantile estimates compared to the QR approach.

Table B.1
5-fold Time Validation over time setup.

Parameter	Possible Values
Learning Rate	0.00001 / 0.0001/ 0.001/ 0.01
Dropout	0.10 / 0.20 / 0.30 /0.40
Multiple	1 / 2 / 4 / 8
L1 Loss	0.01 / 0.005 / 0.0005
Hidden Layer	1 / 2
Activation	sigmoid / tanh
Epochs	100/ 150/ 200/ 250/ 300

Note: The table shows different values for the hyperparameter search. As avoiding overfitting is of major concern, we put much emphasis on regularization parameters (L1) and different designs of Dropout Layers. With respect to the difficulties of training very deep neural networks, we do not use more than two hidden layers, but rather increase the number of neurons in each layer.

to be a robust and suitable baseline for most applications, see, e.g, Gu et al. (2020). As we have the opposite starting point, we double the number of neurons in each hidden layer. Another positive side effect is that we now only have to validate the multiple of this baseline structure, which makes the validation task much more efficient.[17] We use eight base neurons in the first hidden layer and 16 neurons in the second hidden layer. Furthermore, due to the vanishing gradient problem of deep neural networks, we evaluate no more than two hidden layers, but rather use a large number of neurons in these layers.

Table B.1 illustrates all used parameters during the 5-fold Time Validation. We estimate all possible combinations of the parameters and also apply a 3- and 10-fold approach, but the results remain the same. Various different learning rates and the adaptive moment-based (*Adam*) optimizer of Kingma and Ba (2014) are used. We further evaluate the levels of Dropout, ranging from very low to rather high. As an activation function, we use sigmoid and tanh, which are very common in neural network architectures and can be differentiated twice. The special part of our hyperparameter set are the multiples. As we assume a baseline structure with eight, respectively 16 neurons in the first and second hidden layer, these multiples are an efficient way to validate the shallowness of our network. The most narrow network with multiple equals one, which coincides with the baseline structure with only one hidden layer. The most shallow network with multiple equals eight, has 64 neurons in the first and 128 neurons in the second hidden layer. In general, more complex models, e.g., a larger multiple and more layers, are prone to overfitting. Hence, we try to find a balance with respect to complexity and support this task with Dropout Layers and weight regularization.

[17] Please note that we also used other baseline models and the classical way, e.g., using several different numbers of neurons without assuming a baseline structure. We find this increasing fashion to be the most robust and computationally efficient one.

Appendix C. Joint effects with macro variables

This Appendix shows the joint effects of all variables with the macroeconomic state. We can clearly observe differences among the two regions. In the United States the Seniority has the largest joint effects with the macroeconomic state, whereas in the European sample we observe large joint effects of the macroeconomic state and the collateralization.

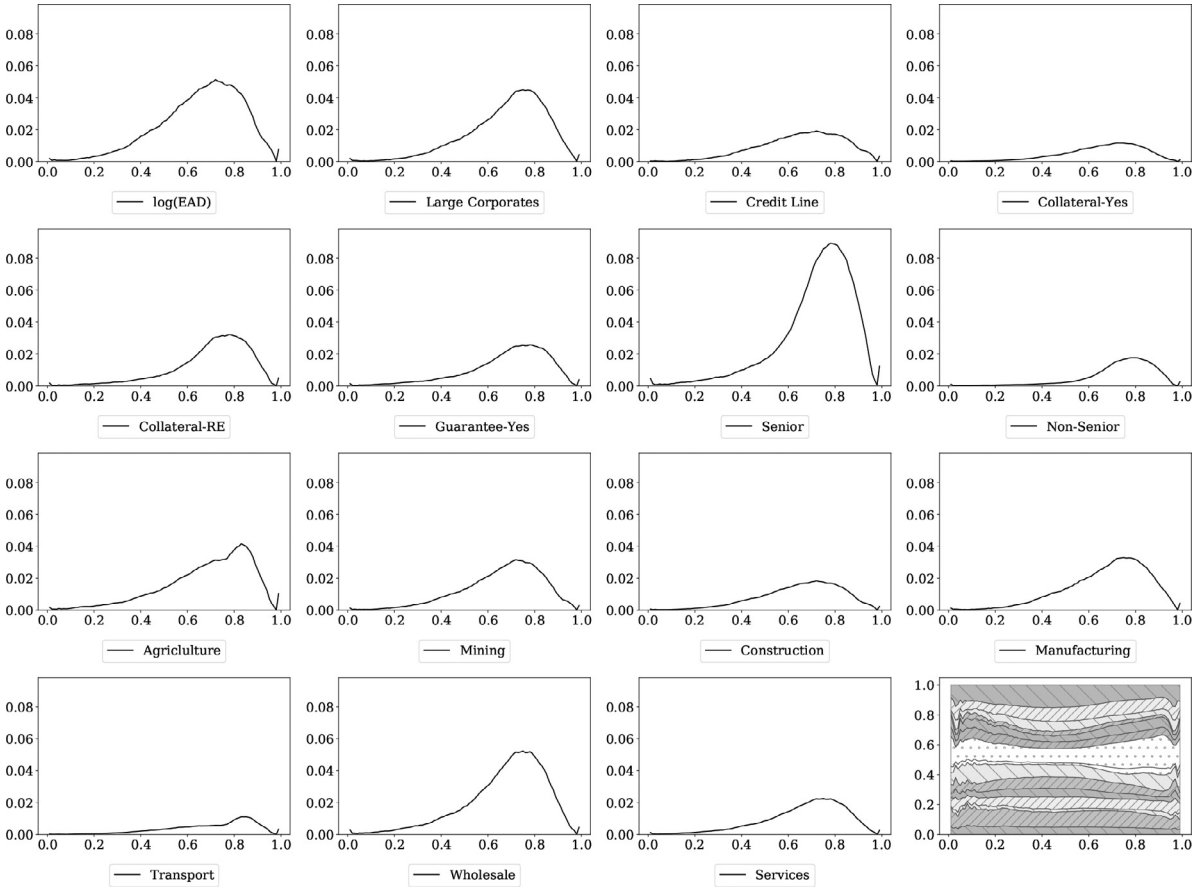

Fig. C.1. $FI_{\tau}^{Joint}(x_{jl})$ of the macroeconomic state | United States. Note: These plots show the estimated values of $FI_{\tau}^{Joint}(x_{jl})$ for every variable relating to the macroeconomic state and for every quantile. The last plot illustrates the importance of all variables in a stacked fashion. This allows us to evaluate which variables especially have a joint effect with the macroeconomic environment.

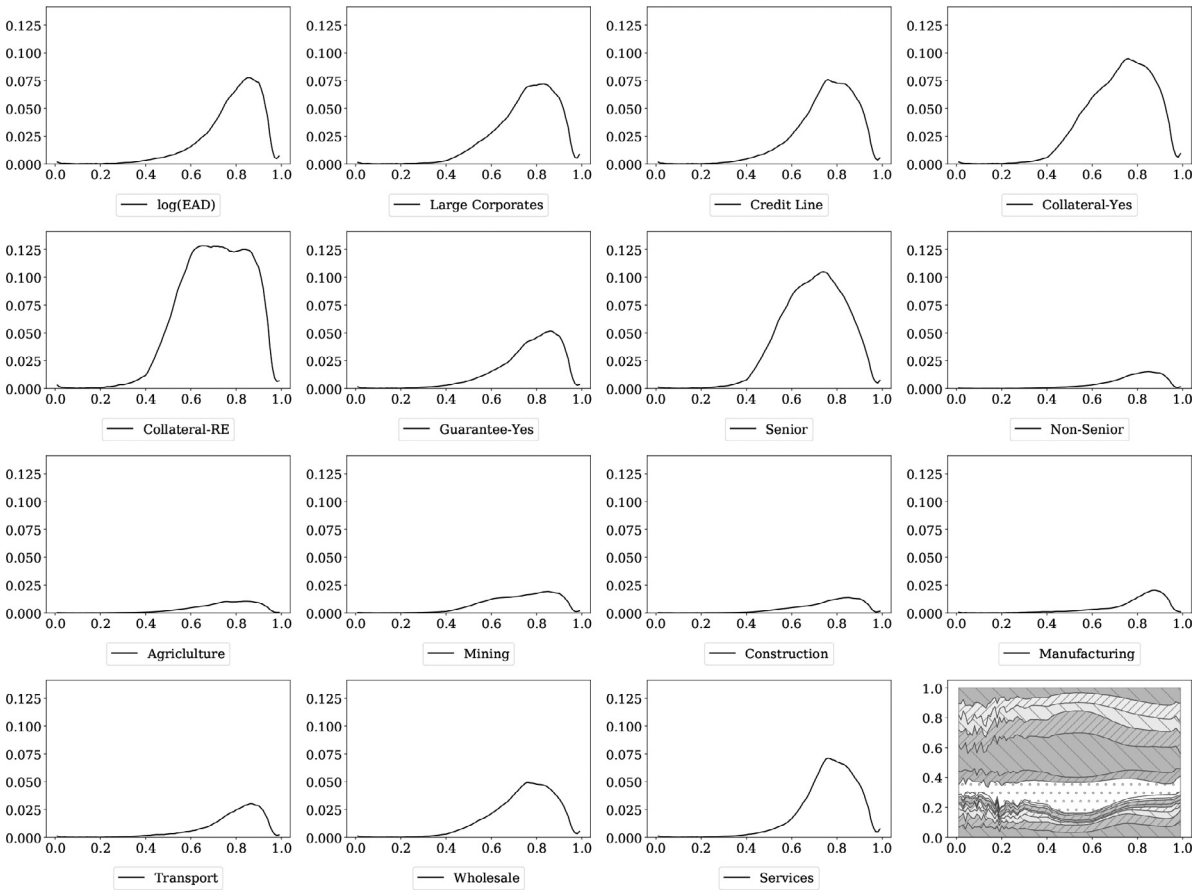

Fig. C.2. $FI_{\tau}^{Joint}(x_{jl})$ of the macroeconomic state | Europe. Note: These plots show the estimated values of $FI_{\tau}^{Joint}(x_{jl})$ for every variable relating to the macroeconomic state and for every quantile. The last plot illustrates the importance of all variables in a stacked fashion. This allows us to evaluate which variables especially have a joint effect with the macroeconomic environment.

References

Altman, E.I., Kalotay, E.A., 2014. Ultimate recovery mixtures. Journal of Banking & Finance 40, 116–129. doi:10.1016/j.jbankfin.2013.11.021.

Apley, D.W., Zhu, J., 2020. Visualizing the effects of predictor variables in black box supervised learning models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 82 (4), 1059–1086. doi:10.1111/rssb.12377.

Bakoben, M., Bellotti, T., Adams, N., 2020. Identification of credit risk based on cluster analysis of account behaviours. Journal of the Operational Research Society 71 (5), 775–783. doi:10.1080/01605682.2019.1582586.

Basel Committee on Banking Supervision, 2019. High-level summary: BCBS SIG industry workshop on the governance and oversight of artificial intelligence and machine learning in financial services.

Bastos, J.A., 2010. Forecasting bank loans loss-given-default. Journal of Banking & Finance 34 (10), 2510–2517. doi:10.1016/j.jbankfin.2010.04.011.

Bellotti, A., Brigo, D., Gambetti, P., Vrins, F., 2021. Forecasting recovery rates on non-performing loans with machine learning. Int J Forecast 37 (1), 428–444. doi:10.1016/j.ijforecast.2020.06.009.

Bellotti, T., Crook, J., 2012. Loss given default models incorporating macroeconomic variables for credit cards. Int J Forecast 28 (1), 171–182. doi:10.1016/j.ijforecast.2010.08.005.

Betz, J., Kellner, R., Rsch, D., 2018. Systematic effects among loss given defaults and their implications on downturn estimation. Eur J Oper Res 271 (3), 1113–1144. doi:10.1016/j.ejor.2018.05.059.

Betz, J., Krger, S., Kellner, R., Rsch, D., 2020. Macroeconomic effects and frailties in the resolution of non-performing loans. Journal of Banking & Finance 112, 105212. doi:10.1016/j.jbankfin.2017.09.008.

Brumma, N., Rainone, N., Crecel, R., 2020. Downturn lgd study 2020. Report. Available from https://www.globalcreditdata.org/library/downturn-lgd-study-2020

Brumma, N., Rainone, N., Crecel, R., 2020. Lgd report 2020-large corporate borrowers. Report. Available from https://www.globalcreditdata.org/news/how-can-banks-project-losses-in-the-current-covid-19-crisis-asks-global-credit-data-in-latest.

Calabrese, R., 2014. Downturn loss given default: mixture distribution estimation. Eur J Oper Res 237 (1), 271–277. doi:10.1016/j.ejor.2014.01.043.

Chen, Z., Chen, W., Shi, Y., 2020. Ensemble learning with label proportions for bankruptcy prediction. Expert Syst Appl 146, 113155. doi:10.1016/j.eswa.2019.113155.

Chernozhukov, V., Fernández-Val, I., Galichon, A., 2010. Quantile and probability curves without crossing. Econometrica 78 (3), 1093–1125.

Cowden, C., Fabozzi, F.J., Nazemi, A., 2019. Default prediction of commercial real estate properties using machine learning techniques. The Journal of Portfolio Management 45 (7), 55–67. doi:10.3905/jpm.2019.1.104.

Cybenko, G., 1989. Approximation by superpositions of a sigmoidal function. Mathematics of control, signals and systems 2 (4), 303–314.

Deutsche Bundesbank, 2020. The use of artificial intelligence and machine learning in the financial sector. URL: https://www.bundesbank.de/resource/blob/598256/d7d26167bceb18ee7c0c296902e42162/mL/2020-11-policy-dp-aiml-data.pdf.

Doshi, H., Elkamhi, R., Ornthanalai, C., 2018. The term structure of expected recovery rates. Journal of Financial and Quantitative Analysis 53 (6), 2619–2661. doi:10.1017/S0022109018000558.

Dumitrescu, E., Hue, S., Hurlin, C., Tokpavi, S., 2021. Machine learning for credit scoring: improving logistic regression with non-linear decision-tree effects. European Journal of Operational Research (forthcomming) doi:10.1016/j.ejor.2021.06.053.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. The Annals of Statistics 29 (5), 1189–1232. Publisher: Institute of Mathematical Statistics

Fritz-Morgenthal, S., Hein, B., Papenbrock, J., 2021. Financial Risk Management and Explainable Trustworthy Responsible AI. SSRN Scholarly Paper. Social Science Research Network, Rochester, NY.

Gambetti, P., Gauthier, G., Vrins, F., 2019. Recovery rates: uncertainty certainly matters. Journal of Banking & Finance 106, 371–383. doi:10.1016/j.jbankfin.2019.07.010.

Gambetti, P., Roccazzella, F., Vrins, F., 2020. Meta-learning approaches for recovery rate prediction. Working Paper.

Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings, pp. 249–256.

Goldstein, A., Kapelner, A., Bleich, J., Pitkin, E., 2015. Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. Journal of Computational and Graphical Statistics 24 (1), 44–65. doi:10.1080/10618600.2014.907095. Publisher: Taylor & Francis

Grunert, J., Weber, M., 2009. Recovery rates of commercial lending: empirical evidence for german companies. Journal of Banking and Finance 33 (3), 505–513.
Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. Rev Financ Stud 33 (5), 2223–2273.
Hartmann-Wendels, T., Miller, P., Tws, E., 2014. Loss given default for leasing: parametric and nonparametric estimations. Journal of Banking & Finance 40, 364–375. doi:10.1016/j.jbankfin.2013.12.006.
Horel, E., Giesecke, K., 2020. Significance tests for neural networks. Journal of Machine Learning Research 21 (227), 1–29.
Horel, E., Mison, V., Xiong, T., Giesecke, K., Mangu, L., 2018. Sensitivity based neural networks explanations. arXiv:1812.01029. ArXiv: 1812.01029
Hornik, K., 1991. Approximation capabilities of multilayer feedforward networks. Neural networks 4 (2), 251–257.
Horowitz, J.L., Lee, S., 2005. Nonparametric estimation of an additive quantile regression model. J Am Stat Assoc 100 (472), 1238–1249.
Hoshino, T., 2014. Quantile regression estimation of partially linear additive models. J Nonparametr Stat 26 (3), 509–536.
Huber, P.J., 1964. Robust estimation of a location parameter. The Annals of Mathematical Statistics 35 (1), 73–101.
Hwang, R.-C., Chu, C.-K., 2018. A logistic regression point of view toward loss given default distribution estimation. Quantitative Finance 18 (3), 419–435. doi:10.1080/14697688.2017.1310393.
Hwang, R.-C., Chu, C.-K., Yu, K., 2020. Predicting LGD distributions with mixed continuous and discrete ordinal outcomes. Int J Forecast doi:10.1016/j.ijforecast.2019.10.005.
Jing, J., Yan, W., Deng, X., 2020. A hybrid model to estimate corporate default probabilities in china based on zero-price probability model and long short-term memory. Appl Econ Lett 1–8.
Kalotay, E.A., Altman, E.I., 2017. Intertemporal forecasts of defaulted bond recoveries and portfolio losses. Rev Financ 21 (1), 433–463. doi:10.1093/rof/rfw028.
Kaposty, F., Kriebel, J., Lderbusch, M., 2020. Predicting loss given default in leasing: a closer look at models and variable selection. Int J Forecast 36 (2), 248–266. doi:10.1016/j.ijforecast.2019.05.009.
Khieu, H.D., Mullineaux, D.J., Yi, H.-C., 2012. The determinants of bank loan recovery rates. Journal of Banking & Finance 36 (4), 923–933. doi:10.1016/j.jbankfin.2011.10.005.
Kingma, D. P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
Koenker, R., 2005. Quantile regression. Camebridge University Press.
Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 46 (1), 33–50. doi:10.2307/1913643.
Koenker, R., Chernozhukov, V., He, X., Peng, L., 2017. Handbook of quantile regression.
Koenker, R., Ng, P., Portnoy, S., 1994. Quantile smoothing splines. Biometrika 81 (4), 673–680.
Krüger, S., Rösch, D., 2017. Downturn LGD modeling using quantile regression. Journal of Banking & Finance 79, 42–56. doi:10.1016/j.jbankfin.2017.03.001.
Kvamme, H., Sellereite, N., Aas, K., Sjursen, S., 2018. Predicting mortgage default using convolutional neural networks. Expert Syst Appl 102, 207–217. doi:10.1016/j.eswa.2018.02.029.
Leow, M., Mues, C., Thomas, L., 2014. The economy and loss given default: evidence from two UK retail lending data sets. Journal of the Operational Research Society 65 (3), 363–375. doi:10.1057/jors.2013.120.
Li, Q., Lin, J., Racine, J.S., 2013. Optimal bandwidth selection for nonparametric conditional distribution and quantile functions. Journal of Business & Economic Statistics 31 (1), 57–65.
Li, Q., Racine, J.S., 2008. Nonparametric estimation of conditional cdf and quantile functions with mixed categorical and continuous data. Journal of Business & Economic Statistics 26 (4), 423–434.
Li, Y., Chen, W., 2019. Entropy method of constructing a combined model for improving loan default prediction: a case study in china. Journal of the Operational Research Society 0 (0), 1–11. doi:10.1080/01605682.2019.1702905.
Loterman, G., Brown, I., Martens, D., Mues, C., Baesens, B., 2012. Benchmarking regression algorithms for loss given default modeling. Int J Forecast 28 (1), 161–170.
Lundberg, S.M., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 30. Curran Associates, Inc., pp. 4765–4774.
Luo, J., Yan, X., Tian, Y., 2020. Unsupervised quadratic surface support vector machine with application to credit risk assessment. Eur J Oper Res 280 (3), 1008–1017. doi:10.1016/j.ejor.2019.08.010.
Mai, F., Tian, S., Lee, C., Ma, L., 2019. Deep learning models for bankruptcy prediction using textual disclosures. Eur J Oper Res 274 (2), 743–758. doi:10.1016/j.ejor.2018.10.024.
Matuszyk, A., Mues, C., Thomas, L.C., 2010. Modelling LGD for unsecured personal loans: decision tree approach. Journal of the Operational Research Society 61 (3), 393–398. doi:10.1057/jors.2009.67.
Nagl, M., 2021. Does non-linearity in risk premiums vary over time? Working Paper 1–44.
Nazemi, A., Baumann, F., Fabozzi, F.J., 2021. Intertemporal defaulted bond recoveries prediction via machine learning. European Journal of Operational Research (forthcomming) doi:10.1016/j.ejor.2021.06.047.
Nazemi, A., Fabozzi, F.J., 2018. Macroeconomic variable selection for creditor recovery rates. Journal of Banking & Finance 89, 14–25.
Nazemi, A., Fatemi Pour, F., Heidenreich, K., Fabozzi, F.J., 2017. Fuzzy decision fusion approach for loss-given-default modeling. Eur J Oper Res 262 (2), 780–791. doi:10.1016/j.ejor.2017.04.008.
Nazemi, A., Heidenreich, K., Fabozzi, F.J., 2018. Improving corporate bond recovery rate prediction using multi-factor support vector regressions. Eur J Oper Res 271 (2), 664–675.
Paulsen, B., Misback, A., Sheesley, J., Uejio, D., Conyers-Ausbrooks, M., 2021. Request for Information and Comment on Financial Institutions' Use of Artificial Intelligence, Including Machine Learning.
Petropoulos, A., Siakoulis, V., Stavroulakis, E., Vlachogiannakis, N.E., 2020. Predicting bank insolvencies using machine learning techniques. Int J Forecast doi:10.1016/j.ijforecast.2019.11.005.
Qi, M., Yang, X., 2009. Loss given default of high loan-to-value residential mortgages. Journal of Banking and Finance 33 (5), 788–799.
Qi, M., Zhao, X., 2011. Comparison of modeling methods for loss given default. Journal of Banking & Finance 35 (11), 2842–2855. doi:10.1016/j.jbankfin.2011.03.011.
Ribeiro, M.T., Singh, S., Guestrin, C., 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, New York, NY, USA, pp. 1135–1144. doi:10.1145/2939672.2939778.
Ribeiro, M.T., Singh, S., Guestrin, C., 2018. Anchors: high-precision model-agnostic explanations. Conference on Artificial Intelligence (AAAI) 18, 1527–1535.
Rolnick, D., Tegmark, M., 2018. The power of deeper networks for expressing natural functions. In: International Conference on Learning Representations.
Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323 (6088), 533–536.
Salinas, D., Flunkert, V., Gasthaus, J., Januschowski, T., 2019. Deepar: probabilistic forecasting with autoregressive recurrent networks. Int J Forecast doi:10.1016/j.ijforecast.2019.07.001.
Sariev, E., Germano, G., 2019. Bayesian regularized artificial neural networks for the estimation of the probability of default. Quantitative Finance 1–18. doi:10.1080/14697688.2019.1633014.
Sigrist, F., Hirnschall, C., 2019. Grabit: gradient tree-boosted tobit models for default prediction. Journal of Banking & Finance 102, 177–192. doi:10.1016/j.jbankfin.2019.03.004.
Sopitpongstorn, N., Silvapulle, P., Gao, J., Fenech, J.-P., 2021. Local logit regression for loan recovery rate. Journal of Banking & Finance 106093. doi:10.1016/j.jbankfin.2021.106093.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research 15 (1), 1929–1958.
Sun, H.S., Jin, Z., 2016. Estimating credit risk parameters using ensemble learning methods: an empirical study on loss given default. Journal of Credit Risk.
Takeuchi, I., Le, Q.V., Sears, T.D., Smola, A.J., 2006. Nonparametric quantile estimation. Journal of Machine Learning Research 7 (Jul), 1231–1264.
Tanoue, Y., Yamashita, S., 2019. Loss given default estimation: a two-stage model with classification tree-based boosting and support vector logistic regression. Journal of Risk.
Tobback, E., Martens, D., Gestel, T.V., Baesens, B., 2014. Forecasting loss given default models: impact of account characteristics and the macroeconomic state. Journal of the Operational Research Society 65 (3), 376–392. doi:10.1057/jors.2013.158.
Tomarchio, S.D., Punzo, A., 2019. Modelling the loss given default distribution via a family of zero-and-one inflated mixture models. Journal of the Royal Statistical Society: Series A (Statistics in Society) 182 (4), 1247–1266. doi:10.1111/rssa.12466.
Wu, Q., Yan, X., 2019. Capturing deep tail risk via sequential learning of quantile dynamics. Journal of Economic Dynamics and Control 109, 103771. doi:10.1016/j.jedc.2019.103771.
Xu, Q., Liu, X., Jiang, C., Yu, K., 2016. Quantile autoregression neural network model with applications to evaluating value at risk. Appl Soft Comput 49, 1–12. doi:10.1016/j.asoc.2016.08.003.
Yao, X., Crook, J., Andreeva, G., 2017. Enhancing two-stage modelling methodology for loss given default with support vector machines. Eur J Oper Res 263, 679–689. doi:10.1016/j.ejor.2017.05.017.
Yashkir, O., Yashkir, Y., 2013. Loss given default modeling: a comparative analysis. The Journal of Risk Model Validation 7 (1), 25–59.