OR Spectrum
https://doi.org/10.1007/s00291-024-00787-7

ORIGINAL ARTICLE

Non-linearity and the distribution of market-based loss rates

Matthias Nagl[1] · **Maximilian Nagl**[1] · **Daniel Rösch**[1]

Received: 23 May 2023 / Accepted: 20 August 2024

Abstract

We synthesize the extended linear beta regression with a neural network structure to model and predict the mean and precision of market-based loss rates. We can incorporate non-linearity in mean and precision in a flexible way and resolve the problem of specifying the underlying form in advance. As a novelty, we can show that the proportion of non-linearity for the mean estimates is 14.10% and 80.37% for the precision estimates. This implies that especially the shape of the loss rate distribution entails a large amount of non-linearity and, thus, our approach consistently outperforms its linear counterpart. Furthermore, we derive trainable activation functions to allow a data-driven estimation of their shape. This is important if predictions have to be in a certain interval, e.g., $(0, 1)$ or $(0, \infty)$. Conducting a scenario analysis, we observe that our estimated distributions are more refined compared to traditional models, thereby demonstrating their suitability for risk management purposes. These estimated distributions can assist financial institutions in better identifying diverse risk profiles among their creditors and across various macroeconomic states.

Keywords Loss given default · Machine learning · Explainable artificial intelligence (XAI) · Distribution

We would like to thank the department editor, the associate editor and two anonymous referees for their comments and suggestions, which have substantially improved the paper.

✉ Maximilian Nagl
maximilian.nagl@ur.de

Matthias Nagl
matthias.nagl@ur.de

Daniel Rösch
daniel.roesch@ur.de

[1] Universität Regensburg, Chair of Statistics and Risk Management, Universitätsstraße 31, 93040 Regensburg, Germany

Published online: 21 September 2024

1 Introduction

The current economic climate is characterized by different challenges. Financial markets have experienced significant turmoils recently due to global economic uncertainty, geopolitical tensions, fluctuations in currency exchange rates and a different interest rate environment. The European Banking Authority (2022) states that these factors also impact financial institutions due to increased funding costs for banks and an overall asset quality deterioration. These circumstances put banks' internal risk management again into the focus of regulators and politics. The largest share of risk a bank faces is determined by its credit risk. Following the latest data, credit risk accounts for 83.3% of risk-weighted assets of 131 major EU banks as of June 2022, underlining its importance (European Banking Authority 2022). The expected loss (EL) of credit risk related assets can be split into three components. Probability of Default (PD) quantifies the probability that a obligor will not meet his agreed obligations. Exposure at Default (EAD) defines the amount of outstanding obligations. Finally, Loss Given Default (LGD) refers to the percentage share of outstanding debt that is lost, given the obligor defaults. This paper focuses on so-called market-based LGDs, which are relevant for publicly traded instruments such as bonds. They are defined as one minus the ratio of the market price 30 days after default over the outstanding amount. Especially market-based LGDs entail challenging characteristics such as bounded support, skewed distribution, and heteroskedasticity (Gambetti et al. 2019). Moreover, the estimation of market-based LGDs has also macro-economic implications. According to SIFMA Research (2022) the US corporate bond market has a volume of 10 trillion USD and therefore the estimation of the LGDs can be essential for the financial stability of the economy. The Basel Accord allows banks to use their own models to provide estimates for LGDs as well as the other components of the expected loss (Basel Committee on Banking Supervision 2017). Therefore, academia aims at providing guidance on how to estimate LGDs and which methods to use, see, e.g., Altman and Kalotay (2014), Kalotay and Altman (2017), Bellotti et al. (2021), Kellner et al. (2022) or Gürtler and Zöllner (2023).

In this paper we utilize market-based LGDs sourced from Moody's Default and Recovery Database spanning from January 1990 to March 2021. This dataset stands out as the most comprehensive when focusing on market-based LGDs. It comprises a range of LGD-specific variables, including seniority, industry sector, and default type, among other pertinent information. Additionally, we extend this dataset by incorporating commonly used macroeconomic and uncertainty-related variables described in the literature. Past literature uses classical statistical models, such as fractional response regression, to predict LGDs using this data, see, e.g., Bastos (2010). These methods commonly focus only on the mean LGD predictions, neglecting the challenging characteristics of LGD distributions. Furthermore, most recent literature reinforces the evidence of non-linearity of drivers for mean market-based LGD estimates using machine learning models, see, e.g., Bastos and Matos (2021), Olson et al. (2021), Bellotti et al. (2021) or Sopitpongstorn et al. (2021). Most of these methods have in common that only

the conditional *mean* is estimated. Although machine learning models increase the flexibility of the modelling framework and are capable of incorporating some distributional aspects, none of the studies explicitly account for these aspects nor investigate their drivers and potential impact.

Making decisions based only on one location parameter, e.g., by using standard machine learning algorithms, may not be holistic in the sense that further distributional characteristics carry important information. For example the dispersion can enhance the understanding of the underlying mechanics between drivers and the whole distribution and be supportive of the managerial decision process. However, they are frequently neglected in the literature.[1] For example, different parts of the distribution can be interpreted as scenarios for banks and risk managers. In that sense, lower quantiles of the distribution can be interpreted as good scenarios for banks, i.e., a low loss realization. On the contrary, higher quantiles imply higher losses. Hence, risk managers can conduct a scenario analysis to investigate how their loan portfolios face losses in adverse, normal, and good scenarios based on the individual loss distribution of their obligors. This enables managers also to reveal different risk profiles among obligors based on their individual distribution. As future realizations are unknown, comparing (predicted) quantiles can provide risk managers with a good indication of how likely low and high LGD realizations can realize and how individual obligors compare to each other. Against this background, we suggest jointly modeling mean and precision, i.e., a dispersion parameter, to allow for varying shapes of the distribution. This paper doesn't aim for a competitive horse race of various methods but contributes by proposing a novel method for explaining market-based LGDs. It focuses on understanding the factors influencing LGDs mean and precision, an area that has been underexplored so far. Rather than predicting future mean LGDs, it uncovers hidden relationships in the LGD distribution, aiding scenario analysis and deriving implications. Thus, it is designed as a non-parametric tool for exploring these connections.

We contribute to the literature in four important ways. First, we combine beta regression by Ferrari and Cribari-Neto (2004) and its extension by Smithson and Verkuilen (2006) and Simas et al. (2010) with artificial neural networks. This contributes to the literature on market-based LGDs, as the combination is the first to allow non-linearity in the mean and precision of the LGD distribution. It extends the work of Gambetti et al. (2019) as we detect a large amount of non-linearity especially in the shape of the distribution. Furthermore, our combination achieves a substantially better performance in and out-of-sample compared to its linear counterpart. Hence, non-linearity plays an important role for the mean and precision of the loss rates.

Second, we use Accumulated Local Effects (ALE) plots by Apley and Zhu (2020) to unveil the non-linear relationships in mean and precision. Furthermore, we

[1] There are studies which include distributional characteristics of LGD into their modeling strategy, see, e.g. Calabrese (2014), Altman and Kalotay (2014), Kalotay and Altman (2017), Krüger and Rösch (2017), Betz et al. (2018), Hwang and Chu (2018); Hwang et al. (2020) or Kellner et al. (2022). However, they do not explicitly model the drivers of different parameters of the LGD distribution.

quantify the amount of non-linearity in the mean and precision estimation according to Apley and Zhu (2020) and Nagl (2021). To the best of our knowledge, this paper is the first to incorporate, visualize and quantify non-linearity in the precision of market-based loss rates.

Third, we derive trainable activation functions similar to He et al. (2015b) to increase the robustness of predictions, especially for unseen data. The trainable activation function offers the neural network a data-driven way of estimating the shape of final predictions, which is new to the finance and credit risk literature. In general, this contribution is not restricted to our discipline but may be beneficial for any other field of research where bounded outputs play an important role, e.g., demand or sales forecasting.

Fourth, we find the that accounting for non-linearity and interactions the modeled distributions differ compared to the beta regression such that they can be better distinguished. This enables risk managers and regulators to better quantify the individual risk of obligors in a straightforward and interpretable way.

The remainder of this paper is structured as follows. In Sect. 2, we give a summary of the relevant literature of LGD estimation. Data is presented in Sect. 3, while the methodology is described in Sect. 4. Our empirical results are discussed in Sect. 5 and Sect. 6 concludes.

2 Literature review

Concerning the special characteristics of LGD distributions, mainly advanced statistical methods are applied. These include for example beta regression, mixture regression, and factorial regression among many others, see, e.g., Gambetti et al. (2019) or Sopitpongstorn et al. (2021) as recent examples. Due to the increased computational power and progress in academia, machine learning models become more and more apparent. They were used by academics at the beginning, but become increasingly applied by practitioners and heavily discussed by regulators. However, the vast majority of these studies focus only on mean estimates of LGDs, see Bastos and Matos (2021), Kellner et al. (2022) or Gürtler and Zöllner (2023) for a recent overview. Nagl et al. (2022) focus on the uncertainty quantification of machine learning models for market-based LGDs and finds that the aleatoric uncertainty, i.e., uncertainty in the data itself, is more significant than epistemic uncertainty, i.e., uncertainty due to a limited sample size. This study emphasises the importance of estimating LGD distributions instead of just point estimates. Focussing on the distribution of LGDs is not entirely new to the literature. Closely related to our paper is Gambetti et al. (2019) who use a linear generalized beta regression to model the distribution of market-based LGDs. However, they do not consider non-linearity in mean or precision in their model. Furthermore, related to the contributions of our paper are Krüger and Rösch (2017) and Kellner et al. (2022). They both estimate quantile regression or machine learning-based extensions thereof, focusing on workout LGDs. They emphasize the importance of accounting for distributional aspects in these LGDs. Their approaches require a sizable dataset to reliably estimate multiple quantiles to describe a full distribution. Furthermore, some papers use classical

statistical models, such as mixture models for LGDs. Altman and Kalotay (2014), Calabrese (2014), Kalotay and Altman (2017), and Betz et al. (2018) utilize these mixture models, consisting of a combination of different distributions, to disaggregate the estimation of a full conditional distribution into subparts. However, with an increasing number of components, these models become less interpretable, and the drivers of LGDs are linearly connected to the components, reducing flexibility. A novel approach proposed by Gürtler and Zöllner (2023) suggests that the modality type of workout LGDs is crucial in determining the best estimation method. Using cluster analysis, they identify three clusters/modality types suitable for their sample. While this approach outperforms traditional models, it requires a sizable dataset, as the entire dataset is divided into clusters, resulting in less data to fit various models. Overall, the methods proposed to account for distributional aspects necessitate a considerable dataset size, making them especially useful for workout LGDs, where data is typically much larger compared to market-based LGDs as used in our paper.

Recent literature on LGD estimation has seen significant growth, particularly with the widespread application of machine learning models. However, several gaps remain. Firstly, while many studies concentrate on mean LGD predictions, they often overlook distributional characteristics, including aleatoric uncertainty, identified as a primary source of uncertainty in LGD estimation by Nagl et al. (2022). Moreover, approaches addressing distributional aspects are typically only viable for large datasets, such as workout LGDs, rather than market-based LGDs. Hence, this paper aims to bridge this gap by synthesizing existing evidence and proposing a suitable method for market-based LGDs. Secondly, there is a notable absence of studies focusing on potential non-linearity in the shape of LGD distributions. Such non-linearities could significantly impact scenario analysis by allowing for different distributional shapes across categories.

3 Data

We use bond loss given defaults from Moody's Default and Recovery Database (Moody's DRD). The examined data contain 2315 market-based LGDs and related bond characteristics ranging from January 1990 until March 2021. Finding suitable drivers of market-based LGD's mean and precision is a challenging task. Gambetti et al. (2019) synthesizes the evidence of the literature on important predictor variables. We follow Gambetti et al. (2019) and use the same features as a starting point. The variables can be divided into three subgroups, bond characteristics, macroeconomic, and uncertainty related determinants. The bond specific characteristics are coupon rate, maturity, the seniority of the bond as well as the issuer's industrial sector. Furthermore, we include the severity of the default type, the defaulted amount and a dummy variable which indicates whether the bond is backed by guarantees. We use several macroeconomic related variables. These include the industrial production returns computed monthly, the S &P 500 returns[2]

[2] https://fred.stlouisfed.org/series/SP500

as well as the recession indicator[3] provided by the National Bureau of Economic Research. Furthermore, delinquency rates in commercial and industrial loans[4] are included quarterly. Following Gambetti et al. (2019), we gather the American default rates from Moody's database and control for withdrawal effects by using the number of defaults registered in a given month divided by the number of firms followed in the same period. We include both rates because delinquency is commonly used if a borrower misses a single payment. Default is usually triggered when a borrower fails to keep up with the loan repayments agreed upon or in some other way fails to fulfill the terms of the loan. Hence, both indices focus on financial distress, but vary in degree and time dimensionality. The third set of variables reflects different types of uncertainty. This may be of particular interest when the focus is on the uncertainty around the estimated means, modeled via their precision. Therefore, we include the VIX,[5] as a proxy for the uncertainty in the stock market. To reflect financial uncertainty the financial uncertainty index[6] derived by Jurado et al. (2015) and Ludvigson et al. (2021) is added. Furthermore, we take the policy uncertainty into account by extending the dataset with the news-based economic policy uncertainty[7] provided by Baker et al. (2016). The last two uncertainty measures are uncertainty based on forecast dispersion of the consumer price index[8] to reflect the inflation and the expenditures of federal and state/local purchases.[9] Those rely on the dispersion of forecasts computed from the Federal Reserve Bank of Philadelphia's Survey of Professional Forecasters. For further details on the variables, we refer to Gambetti et al. (2019). Similar to Olson et al. (2021) all macroeconomic variables and uncertainty indices are lagged by one quarter to ensure predictive properties.

Similar to Görgen et al. (2022) Fig. 1 shows the slightly negatively skewed bimodal distribution of the realized market-based LGDs in our sample. The average LGD is 61.40%. The lowest LGD is only half a percent, while the highest one is close to 100%. Overall, we recover the stylized empirical features of bond-related LGDs such as bounded support and skewed distribution.

Taking a closer look at the correlations of the uncertainty measures in Table 1, one can observe that the highest correlation is between VIX and financial uncertainty with a value of 76.64%. The other correlations are moderate to low.

As this database is often used for investigating market-based LGDs, different periods are frequently used in the literature, see, e.g. Altman and Kalotay (2014), Kalotay and Altman (2017), Hwang and Chu (2018), Gambetti et al. (2019), Hwang et al. (2020), Sopitpongstorn et al. (2021) just to name a few.[10] Therefore, stylized

[3] https://fred.stlouisfed.org/series/USREC
[4] https://fred.stlouisfed.org/series/DRALACBS
[5] https://fred.stlouisfed.org/series/VIXCLS
[6] https://www.sydneyludvigson.com/macro-and-financial-uncertainty-indexes
[7] http://www.policyuncertainty.com/global_monthly.html
[8] https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/dispersion-forecasts
[9] https://www.philadelphiafed.org/surveys-and-data/rfedgov
[10] The actual number of observation varies between these studies as they do not use the same set of bond characteristics. For example, Hwang and Chu (2018), Hwang et al. (2020) and Sopitpongstorn et al. (2021) report a higher number of observations by not considering the coupon rate, contrary to Gambetti et al. (2019) and our approach. However, the stylized descriptive facts remain comparable.

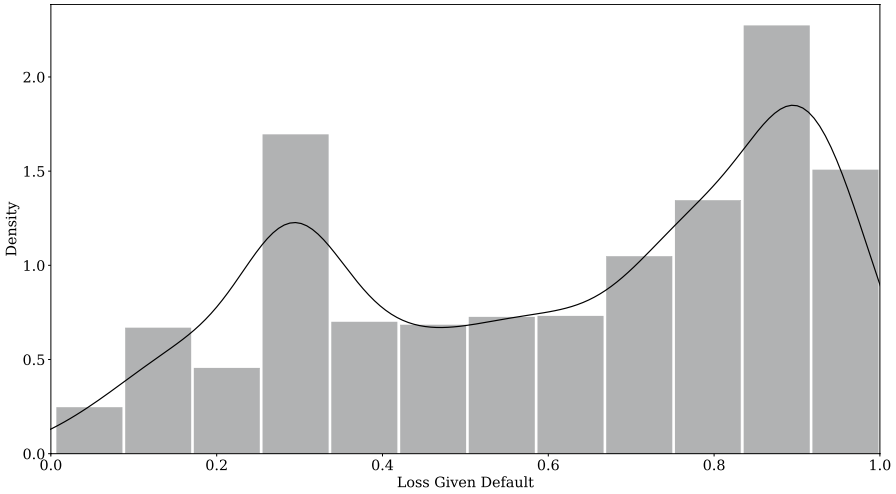


Fig. 1 Histogram of LGDs

Table 1 Upper triangle of the correlation matrix of uncertainty features in percentages

	VIX	Fin. unc.	News-based EPU	CPI unc.	F.S.L. exp. unc.
VIX	100.00	76.64	40.14	41.24	12.34
Fin. unc.		100.00	27.17	47.54	− 8.59
News-based EPU			100.00	29.48	− 0.79
CPI unc.				100.00	7.62
F.S.L. exp. unc.					100.00

All displayed values are expressed as percentages

facts such as the LGD increasing for lower seniority and differences for average LGDs across industries are well known. Hence, we move the discussion of these facts to Appendix A. Table 6 displays a overview of summarizing statistics for the total dataset. Tables 7, 8 and 9 give an overview of the descriptive statistics according to the seniority, industry sector and the default type.

4 Methods

The distribution of LGDs ranges from 0 to 1 and can be skewed and multimodal. As a starting point, we rely on the beta regression because of its flexibility and the fact that the distributional assumption matches the range of the LGDs. We use the alternative definition with two parameters $0 < \mu < 1$ and $\phi > 0$ to model the LGD, Y, with support $0 < Y < 1$. μ corresponds to the mean of Y, whereas ϕ is the precision parameter. Following Ferrari and Cribari-Neto (2004) the density of the beta distribution is:

$$f(y;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\cdot\Gamma((1-\mu)\phi)} \cdot y^{\mu\phi-1} \cdot (1-y)^{(1-\mu)\phi-1}, \tag{1}$$

where $\Gamma(.)$ denotes the Gamma function. The parameters μ and ϕ can be linked to the first two central moments of Y by:

$$E(Y) = \mu \tag{2}$$

$$Var(Y) = \frac{\mu(1-\mu)}{1+\phi} \tag{3}$$

By Eq. (3) we see, that for a fixed μ the variance decreases in ϕ. Especially the extension to a generalized form, where the precision ϕ can be modeled as a dependent variable in addition to the mean μ, is useful to analyze LGD uncertainty. To model the density in Eq. (1) one can formulate a regression model called the beta regression as stated by Ferrari and Cribari-Neto (2004), Smithson and Verkuilen (2006), Simas et al. (2010). In this regression model, the input variables are weighted by their regression coefficients and transformed to match the desired support of μ and ϕ. Usually for μ the logistic and for ϕ the exponential function is used. The regression coefficients are estimated by maximum likelihood optimization. In this approach, μ and ϕ are modeled as a function of the explanatory variables. We maximize the sum of the log likelihood over N bonds of Eq. (1), where y_i represent the LGD of the i-th of N bonds[11]:

$$\begin{aligned} &LL(y_1,\dots,y_N;\mu_1,\dots,\mu_N,\phi_1,\dots,\phi_N) \\ &\quad= \sum_{i=1}^{N} \big(log(\Gamma(\phi_i)) - log\big(\Gamma(\mu_i\phi_i)\big) - log\big(\Gamma(1-\mu_i)\phi_i\big) \\ &\qquad + \big(\mu_i\phi_i - 1\big)y_i + \big((1-\mu_i)\phi_i - 1\big)\big(1-y_i\big). \end{aligned} \tag{4}$$

Despite the flexibility of the beta regression, it is limited in that the relationships between the predictors and the dependent variable have to be specified beforehand. To resolve this restriction we propose the Beta Regression Artificial Neural Network (BRANN) and its extension, the Generalized Beta Regression Artificial Neural Network (G-BRANN).

Modeling approaches of beta regression neural networks The potential modeling approaches to connect neural networks with the beta regression can be divided into three cases. First, if the input variables are the same for both outputs, μ and ϕ, a vanilla feed-forward neural network with two output neurons can be used.[12] This allows interactions of the input variables between μ and ϕ. Hence, one would assume, that μ and ϕ can be explained by the exact same variables, which can interact with each other. Second, this can be relaxed by using skip connections, see, e.g., He et al. (2015a). Hence, the input variables of one parameter are a subset of the

[11] Because some of the explanatory variables are bond-specific for each observation i, i=1,...,N, we also subscript μ_i and ϕ_i in the log likelihood.

[12] For sparsity of notation, we skip subscripts for μ and ϕ here.

other and one would assume that μ and ϕ are driven by the same dynamics. Third, the input variables for the two parameters can be different and do not need to share any input variables. So each parameter can be modeled by a separate neural network, a μ-sub-network and a ϕ-sub-network, and then merged together. This is the most flexible approach, and, thus, we follow this path. In BRANN, which is based on a beta regression, ϕ is modeled as a constant. So the ϕ -sub-network can be represented by a neural network with a constant as input and no hidden layers. Similar to the beta regression, ϕ has to be a positive value, thus, we impose a transformation $\zeta(.)$ which is the exponential function following Ferrari and Cribari-Neto (2004).

BRANN The μ-sub-network can consist of L layers, $l = 1, \ldots, L$ with K_l neurons each. The first layer is called the input layer. This layer takes the input matrix $X \in \mathbb{R}^{N \times p}$, which typically consists of N observations with p exogenous features, and feeds every observation into the network. Each layer l takes as input the output of the previous layer $o_{l-1} \in \mathbb{R}^{K_{l-1} \times 1}$ and weights it by multiplying it with a weights matrix $W_l \in \mathbb{R}^{K_l \times K_{l-1}}$ and adding a bias vector $b_l \in \mathbb{R}^{K_l \times 1}$. The weighted output of the previous layer is activated by a non-linear activation function $\psi(.)$

$$o_l = \psi(W_l o_{l-1} + b_l) \tag{5}$$

This produces the output o_l of the current layer l. For the last layer L the weighted output of the previous layer is activated by a function $\iota(.)$. Since the desired range for μ is between 0 and 1, the logistic function is chosen for $\iota(.)$ to ensure the output of the μ-sub-network, o_μ, stays in the desired interval from 0 to 1.

$$o_\mu = \iota(o_L) = \frac{1}{1 + e^{-o_L}} \equiv \hat{\mu}_{BRANN} \tag{6}$$

Since the ϕ parameter of BRANN is modeled as a constant, the ϕ-sub-network in BRANN can be formalized as follows:

$$o_\phi = \zeta(r) = e^{w_\phi r + b_\phi} \equiv \hat{\phi}_{BRANN} \tag{7}$$

with $r \in \mathbb{R}$ as the constant input, usually $r = 1$, $w_\phi \in \mathbb{R}$ and $b_\phi \in \mathbb{R}$ are weight and bias for r. In a final step o_μ and o_ϕ are merged to get predictions of μ and ϕ.

G-BRANN For G-BRANN we use the same procedure as for BRANN, but increase the flexibility in the ϕ-sub-network by allowing multiple input variables, that can differ from the input variables of the μ-sub-network. Hence, we allow our model to incorporate non-linearity, which can be different in the μ- and the ϕ-sub-network. Since the μ-sub-network is the same as in Eq. (6), only the ϕ-sub-network changes to a network with Z layers where $z = 1, \ldots, Z$.

$$o_z = \psi(W_z o_{z-1} + b_z) \tag{8}$$

$$o_\phi = \zeta(o_Z) \equiv \hat{\phi}_{G-BRANN} \tag{9}$$

Trainable activation functions Outliers can impact the optimization tremendously and cause problems in the whole estimation process. This is especially true for the ϕ-part and out-of-sample predictions. The usual link function for the ϕ-part of the

generalized beta regression is the exponential function to ensure positivity of $\hat{\phi}$, but this link function can be too steep or too flat. Because we do not know the best suiting link function, we choose a data-driven approach and give G-BRANN the flexibility to learn the last activation function for ϕ from the data. We introduce three trainable activation functions in the spirit of the Parametric Rectified Linear Unit by He et al. (2015b). Hence, the last activation function for ϕ, labeled as $\zeta(.)$ differs for G-BRANN. The first one, the trainable exponential function (t-exp), can be modeled in terms of the exponential function, with the addition that the steepness of the curve is defined by the parameter a. This trainable parameter makes it possible for the G-BRANN to determine how steep the function should be in a data-driven fashion. If a is equal to Euler's number, the function results in the original link function. The trainable exponential function with a as the trainable parameter can be calculated as follows:

$$\zeta(o_Z) = e^{o_Z \cdot log(a)} = a^{o_Z} \tag{10}$$

The second function, trainable softplus (t-soft), is based on the softplus function with a steepness parameter q. Analogous to the first function, the network can learn how steep the activation function should be. The trainable softplus function with q as a trainable steepness parameter can be represented as:

$$\zeta(o_Z) = \frac{log(1 + e^{q \cdot o_Z})}{q} \tag{11}$$

The last activation function is called the trainable sigmoid function (t-sig), which is an extension of the adjustable generalized sigmoid as described in Apicella et al. (2021) as we introduce an additional shifting parameter c. The sigmoid is bounded on the open interval from zero to one. Because the ϕ parameter of a beta distribution can be any positive number a few adaptions have to be made. A multiplicative constant h can be used to stretch the sigmoid function to the open interval from zero to h, which can be any positive number. A common problem of the sigmoid function is saturation. To resolve that, we introduce two additional trainable parameters. The parameter s is the steepness parameter of the sigmoid function. For decreasing negative s the sigmoid function tends toward a step function. The last parameter c is the shifting parameter. This can be helpful if the output of the function tends towards the lower bound. Since zero is the lower bound of the sigmoid function and the ϕ of a beta regression, the log likelihood can explode for very low ϕ.

The trainable logistic function with h, s, and c as height, scale, and shift parameters are defined as:

$$\zeta(o_Z) = \frac{h}{1 + e^{s \cdot o_Z}} + c \tag{12}$$

Graphical illustrations of the activation functions can be found in Figs. 13, 14 and 15 in Appendix B. Table 2 summarizes the proposed activation functions.

A simple graphical illustration of BRANN and G-BRANN with an arbitrary number of hidden layers is provided in Fig. 2. In this example we take the variables $\{x_1, x_2, x_3, x_4\}$ as input variables for modeling the μ-sub-network. Each input is

Table 2 Trainable activation functions

Function	Formula	Parameters
Trainable exponential function	$\zeta(o_Z) = e^{o_Z \cdot log(a)} = a^{o_Z}$	a: Steepness parameter
Trainable softplus function	$\zeta(o_Z) = \frac{log(1+e^{q \cdot o_Z})}{q}$	q: Steepness parameter
Trainable logistic function	$\zeta(o_Z) = \frac{h}{1+e^{s \cdot o_Z}} + c$	h: Height parameter s: Scale parameter c: Shift parameter

This table illustrates the different definitions of the novel trainable activation functions. Each of the parameters is trained during the model fit and can be estimated data-driven

weighted and non-linearly transformed in the hidden layers. The ϕ-sub-network gets the constant input r, which usually has the value $r = 1$ as the sub-network models ϕ by weighting the constant without a non-linear transformation. This results in a constant ϕ for every observation. G-BRANN can consist of the same μ-part, but this is not mandatory. The ϕ-part shares the input variable x_3 with the μ-sub-network, but has additional variables x_5 and x_7. These input variables are also non-linearly transformed in the hidden layers. The objective function for both network types is the same as for the beta regressions, defined in Eq. (4). For stability reasons, we minimize the mean of the negative log likelihood instead of the sum, so the gradients do not tend to explode. For every market-based LGD i, we model individual values of $\hat{\mu}_i$ and $\hat{\phi}_i$. For BRANN and the linear beta regression, $\hat{\phi}_i$ is constant for all observations.

Accumulated Local Effects plots As BRANN and G-BRANN rely on neural networks, they are black-box by nature. However, the body of literature focusing on explanation methods has grown fast. Bastos and Matos (2021) provide a comprehensive overview of recent XAI techniques for credit risk. They show that financial institutions can use these techniques to (probably) comply with regulatory concerns of recovery rate predictions. Similar conclusions are drawn by Kellner et al. (2022). We use Accumulated Local Effect (ALE) Plots by Apley and Zhu (2020) to analyze BRANN and G-BRANN. This method is a common choice for visualizing feature effects in credit risk. One example is Bellotti et al. (2021), who use ALE plots on workout LGDs or Barbaglia et al. (2021) also using ALE Plots to analyze the drivers of probability of defaults of European mortgage. Multiple XAI methods, including ALE plots as well as Shapley values, are compared by Bastos and Matos (2021).

To compute the ALE plots, we first divide the range Z of one specific predictor $X_j \in \mathbb{R}^{N \times 1}$, where $j = 1, \ldots, p$, into a grid with $k \in [0, 1, \ldots, K]$, where K is the number of total bins. Following Apley and Zhu (2020), Z_k is chosen as the $\frac{k}{K}$ quantile of the empirical distribution of X_j, where Z_0 is the minimum and Z_K is the maximum. Let S_k define a set of observations that lies between the boundaries Z_{k-1} and Z_k. Furthermore, n_k denotes the number of observations in S_k, and $k(X_j)$ is an index that indicates in which bin a given value of X_j falls. The (uncentered) ALE can then be written as

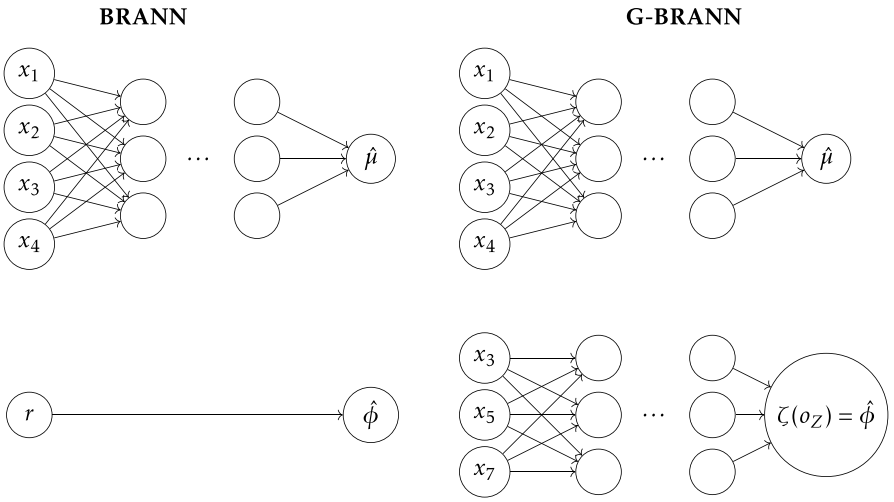


Fig. 2 Graphical illustration BRANN and G-BRANN. Notes: This figure illustrates a stylized network structure of BRANN and G-BRANN. The advantage of these structures is that we can allow different input variables for each of the distributional parameters of the beta distribution

$$g_{ALE}(X_j) = \sum_{k=1}^{k(X_j)} {n_k}^{-1} \sum_{i \in S_k} \left[f(Z_k, X_{\rho,i}) - f(Z_{k-1}, X_{\rho,i}) \right]. \tag{13}$$

$X_\rho \in \mathbb{R}^{N \times p-1}$ defines the set of variables without the variable j and $f(.)$ denotes the neural network's predictor before the final transformation. For each observation i we obtain a prediction assuming X_j to be the upper and lower limit of the interval, i.e., Z_{k-1} and Z_k, and calculate its difference. These differences are summed over all observations in the bin and weighted by the number of observations in that bin, n_k, to obtain the (uncentered) local effect of X_j. Finally, we accumulate these weighted summed differences up to a given value of X_j using the outer sum. This result is centered such that the mean effect of X_j is zero:

$$\Theta_{ALE}(X_j) = g_{ALE}(X_j) - N^{-1} \sum_{i=1}^{N} g_{ALE}(X_{j,i}) \tag{14}$$

The ALE plots have many advantages. Among other things, they are fast to compute and unbiased. Therefore, they can be used even if features are correlated in contrast to many other XAI techniques, such as partial dependence plots, see Apley and Zhu (2020). ALE plots are centered so that the mean effect of the features is zero. Therefore, the y-Axis of the ALE can be interpreted as the main effect of the independent variable at a certain point in comparison to the average predicted value. Furthermore ALE plots provide a R^2-like measure, which describes up to which degree the prediction can be explained by main order, second order, etc. effects. The proposed $R^2_{ALE,m}$ by Apley and Zhu (2020) can be formalized as follows:

$$R^2_{ALE,m} = \frac{var\{\sum_{J\subsetneq\{1,\ldots,d\},|J|\leq m} \Theta_{ALE}(X_J)\}}{var\{f(X)\}} \tag{15}$$

where m describes up to which order of effects the R^2_{ALE} is calculated. Therefore, it holds that $R^2_{ALE,d} = 1$. Nagl (2021) extended this approach by introducing $R^2_{ALE,linear}$, which measures how much linearity the prediction of the model contains. The $R^2_{ALE,linear}$ is defined as:

$$R^2_{ALE,linear} = \frac{var\{\sum_{j=1}^{p} \Theta^{linear}_{ALE}(X_j)\}}{var\{f(X)\}} \tag{16}$$

where $\Theta^{linear}_{ALE}(X_j)$ can be calculated by fitting a linear regression on $\Theta_{ALE}(X_j)$, which are the first order effects. Therefore, $1 - R^2_{ALE,linear}$ quantifies the degree of non-linearity in the prediction. All ALE plots are generated with a grid size of $K = 10$ and we calculate the ALE Plots for o_L (Eq. (5)) and o_Z (Eq. (8)), i.e., before the final transformation in the output layer.

5 Results

5.1 Feature selection and model estimation

Feature selection The selection of important drivers for market-based LGDs is not trivial. Therefore, we use the selection of Gambetti et al. (2019) as a starting point of our analysis. However, as our study additionally uses data after 2015 (about 7 years more), we follow an iterative process to select the most relevant to our sample. We divide our dataset randomly into an in-sample (80%) and an out-of-sample (20%) groups. For the feature selection process, we further divide the in-sample set into a training (70%) and testing set (30%). During this process, various feature sets are calibrated on this training set and predictions are generated for the testing set. The value of the loss function, i.e. the negative log-likelihood, on the testing set serves as metric for the feature selection process. Therefore, we select our feature to be suitable to predict out-of-sample data. Alternatively, we can use the same approach as Gambetti et al. (2019) and follow a step-GAIC approach. Then, however, we would select our feature only on training data, i.e. in-sample. As we want a model which can also predict out-of-sample/time data, we opt for selecting features by out-of-sample losses. Using AIC instead of the negative log-likelihood, does not change our results.

Since the μ-part is well researched in terms of drivers, see Gambetti et al. (2019), we choose the same variables for the μ-part, which are best performing in Gambetti

et al. (2019).[13] For the ϕ-part we apply the forward selection algorithm using the generalized linear beta regression, assuming the same selection for the μ-part from the previous step, following Gambetti et al. (2019).[14] Please note that the final set of features is robust to different splitting points in the training (70%) and testing (30%) set. Table 3 shows the final variables for the μ- and ϕ-parts.

Model estimation To find the optimal parameters for BRANN and G-BRANN, we conduct a random 5-fold cross-validation of the training sample. In summary, we draw 500 different configurations of hyperparameter values. Similar to Kellner et al. (2022) we adopt the multiple approach to find the number of neurons in each subpart of our network. The baseline for the multiple approach is (32, 16), i.e., we use maximum two hidden layers. In this approach we sample a multiplier for the baseline network structure instead of directly sampling the number of neurons in each hidden layer.[15] The descending number of neurons in each hidden layer is inspired by Gu et al. (2020). Furthermore, we use Stochastic Gradient Decent (SGD) as an optimizer and ReLU as an activation function in all hidden layers, which is in line with the literature, see, e.g., Gunnarsson et al. (2021) or Nagl et al. (2022). To increase the robustness of our results, we fit every constellation 10 times and use the average of these repetitions in the hyperparameter search. This eliminates the impact of the random weight initialization in the first step of the training phase.[16] Hyperparameters are the learning rate, the multiple, the dropout rate, the number of hidden layers, and our novel trainable activation functions for the ϕ-part in G-BRANN. In addition to that we include a MaxNorm kernel constraint of 3.0 as recommended by Srivastava et al. (2014) for dropout in neural networks. The search space and the final values are reported in Table 4.

Interestingly, the estimated coefficient of the t-exp activation function differs to Euler's number, indicating that G-BRANN selects a different shape of this activation function to be optimal.[17]

Our main metric for the performance comparison is the log likelihood, as it measures the performance concerning the distributional fit. However, we additionally include two common metrics from the literature, namely the MSE and the Pseudo-R^2 to assess how well the *mean* estimates perform.[18]

[13] In contrast to Gambetti et al. (2019) we dropped the Real Gross Domestic Product(GDP) as it worsens the performance considerably. This may partly be traced back to the recent crisis, where we observed large variations in GDP but almost no variation in the resulting LGD in these quarters.

[14] Alternatively, one could also use BRANN and G-BRANN for feature selection, but this would include a hyperparameter search in each step. As the aim of this paper is not to find the ultimate selection of drivers of market-based LGDs, we think that our approach is sufficient. Overall our selection recovers recent findings by Sopitpongstorn et al. (2021), Nazemi et al. (2021) and Bastos and Matos (2021).

[15] For example, if we sample a multiplier of 4 in a two hidden layer network, we have (128, 64).

[16] We find that 10 repetitions are enough in our setup. The differences in the means of 10 independent repetitions are negligible, and, thus, we find our results robust and reproducible.

[17] As robustness, we also conduct a hyperparameter search for G-BRANN where we use only the standard activation function in the output layer of ϕ. Overall, the trainable activation function outperforms the standard functions consistently.

[18] The MSE measures the quadratic difference between true and predicted mean LGDs. The Pseudo-R^2 is derived by Ferrari and Cribari-Neto (2004) and quantifies the squared correlation between the linear predictors of the model and the true realization.

Table 3 Selected variables for the subnetworks

Variable	μ-subnetwork	ϕ-subnetwork
Bond characteristics		
Coupon rate	✓	✓
Maturity	✓	✓
Industry sector	✓	✓
Seniority	✓	✓
Default type	✓	
Macroeconomic variables		
Recession indicator	✓	✓
Industry production	✓	✓
S &P 500 Returns	✓	✓
Delinquency rates	✓	
Default rate	✓	✓
Uncertainty measures		
Financial uncertainty	✓	✓
CPI uncertainty	✓	✓
News-based EPU	✓	
Uncertainty relative to Federal/state/local expenditures	✓	

The NBER-based recession indicators for the United States are retrieved from the Federal Reserve Bank of St. Louis (FRED) website as well as the industrial production and S &P 500 and the delinquency rate on All Loans. Following Gambetti et al. (2019), we gather the American default rates from Moody's database and control for withdrawal effects. The uncertainty measures are retrieved from the author's website. These include financial uncertainty (Jurado et al. 2015 and Ludvigson et al. 2021), the news-based EPU from Baker et al. (2016). Furthermore, we use survey-based proxies of uncertainty by including the inflation uncertainty measure for the United States (CPI uncertainty) and a proxy of uncertainty relative to both federal and state/local purchases

Table 5 illustrates that BRANN and G-BRANN outperform their linear counterparts by a large margin in terms of sum and mean log Likelihood. This also holds for MSE, and Pseudo-R^2 in-sample as well as out-of-sample. Bold values indicate the best value, underlined values indicate the second best. Overall, the neural networks are first or second-best choice for every metric. To remain comparable with Gambetti et al. (2019), we report the sum of the log likelihood. Hence, the values of the out-of-sample dataset are smaller due to the smaller sample size. Looking at the mean log likelihood, we observe that the values are, as expected, slightly smaller but comparable to the in-sample values. Overall, we see substantial improvements in the log-likelihood by the BRANN and G-BRANN models.

One might argue that the comparison of the "raw" likelihood is not fair as we do not control for the larger number of parameters in the neural networks and we should rely on metrics like the Akaike Information Criterion (AIC) instead. This could be

Table 4 Setup and final values of the hyperparameter search

Parameter	Distribution	BRANN	G-BRANN	
		μ-part	μ-part	ϕ-part
Learning rate	$U^c \sim [0.001, 0.1]$	0.0396	0.0091	
Dropout rate	$U^c \sim [0.05, 0.50]$	0.4385	0.3435	0.3144
Hidden layer	$U^d \sim [1, 2]$	2	2	2
Multiple	$U^d \sim [1, 8]$	1	3	8
Activation function	t-exp, t-sig, t-soft	–	–	t-exp ($\alpha = 3.37$)

The table shows different values for the hyperparameter search. U^c labels the continuous uniform distribution, whereas U^d labels the discrete uniform distribution. We observe that G-BRANN requires a wider network structure for the μ-part and a wide and deep structure for the ϕ-part

the case for in-sample values due to overfitting resulting in higher likelihood values. However, if we interpret this along with the (mean) likelihood values of the out-of-sample data, we do not see evidence for overfitting as we implemented various regularization techniques and rely on cross-validation. Furthermore, the number of parameters in neural networks does not necessarily coincide with the complexity. Recent literature shows that overparameterized neural networks even generalize better than those with a lower number of parameters, see, e.g. Belkin et al. (2019) or Yang et al. (2020).

To counteract concerns that our results are not robust to an out-of-time exercise and in comparison to other well-known machine learning models, we conduct a robustness check in Appendix C. The search space of possible hyperparameters including the final results can be found in Table 10. As displayed in the subtables of Table 11 we observe that the superiority of (G-) BRANN holds also for future predictions, and regarding the mean estimate, we observe similar performance.

5.2 Drivers of μ and ϕ

Bond related drivers of μ

The following figures unveil the relationship between the selected variables and the predicted mean of the LGD distribution. To allow a better comparison with the traditional approach, i.e. linear models, we also add their relationship. Overall, the ALE plots in the μ-part from G-BRANN and BRANN point in the same direction. Therefore, only the ALE plots of G-BRANN are presented in the following.[19] As the number of observations varies across the value range of the drivers, areas with a low number of observations should be interpreted with caution. Nonetheless, we have enhanced the robustness of our interpretations by refitting the models and recalculating the ALE Plots 100 times.

[19] The plots for BRANN as well as the plots of the control variables, such as industry sector, seniority, and default type are available upon request.

Table 5 Evaluation Metrics

(a) In Sample

	Beta regression	BRANN	Generalized linear Beta regression	G-BRANN
Σ Log Likelihood	747.470	1211.769	968.669	**2085.335**
Ø Log Likelihood	0.404	0.654	0.523	**1.126**
MSE	0.038	**0.022**	0.038	**0.022**
Pseudo-R^2	0.464	**0.688**	0.447	0.636

(b) Out of Sample

	Beta regression	BRANN	Generalized linear beta regression	G-BRANN
Σ Log Likelihood	165.536	232.737	227.342	**300.546**
Ø Log Likelihood	0.358	0.503	0.491	**0.649**
MSE	0.041	**0.031**	0.041	0.030
Pseudo-R^2	0.403	**0.549**	0.398	0.538

This table shows the average performance metric of BRANN and G-BRANN over 100 repetitions and their linear counterparts. The first row shows the sum of log likelihood to be comparable to the literature. The second row shows the average log likelihood, whereas the third row shows the mean squared error. The last row displays the Pseudo-R^2 following Ferrari and Cribari-Neto (2004). We observe that the neural network related methods consistently outperform the linear models in every performance metric. To remain comparability to Gambetti et al. (2019) we report the sum of the log likelihood. Bold values indicate the best, whereas underlined values the second best performance

Bold values indicate best values

Figure 3 shows that the slopes of the linear model (dashed line) coincide with the (linear) evidence from the literature, i.e., LGD increases with maturity and decreases with coupon rate, see, e.g., Gambetti et al. (2019). However, the ALE plots of G-BRANN reveal a more nuanced picture. We observe that a higher maturity results in higher mean LGDs up to a certain point, but this increasing effect vanishes for bonds with a maturity greater than roughly 20 years. Generally, the positive relationship could be explained by sell-side pressure originating from institutional investors who usually hold bonds with longer maturity, see Jankowitsch et al. (2014). This effect decreases for maturities greater than 20 years and even gets negative. But this negative effect could be due to the small number of bonds with very long maturities. For BRANN the effect of the maturity is almost constant after the 10 years mark. For the coupon rate, we observe a u-shaped relationship, as the LGD decreases for coupon rates up to 9%, but increases afterwards. A negative relationship is plausible as bonds with a higher coupon rate could be of higher value given there is a reasonable probability that all cash flows can be collected during the resolution of the bond, see Jankowitsch et al. (2014). However, a higher coupon rate also indicates higher risk, and, thus, for riskier bonds, the market expects higher losses as the probability that all cash flows can be recovered may be lower.

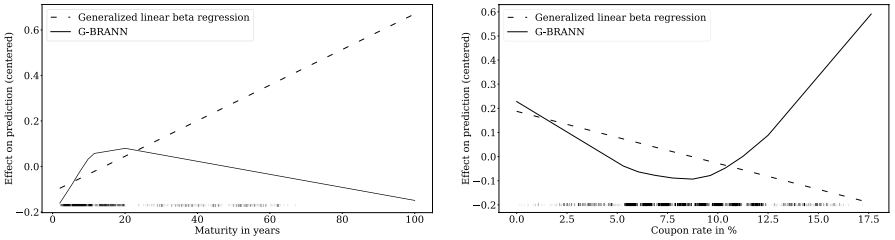


Fig. 3 ALE plots of bond characteristics I μ. Notes: The figures show the ALE plots for G-BRANN in solid black and the generalized linear beta regression in dashed black. We initialize the G-BRANN 100 times and calculate the final median ALE plots. We also include a rug plot at the bottom to visualize the distribution of the underlying data

Macroeconomic drivers of μ

Figure 4 shows that the default rate has the largest impact of the macroeconomic drivers on market-based LGDs, which appears logarithmic. For S &P 500 returns the linear model finds a (counterintuitive) positive relationship, whereas G-BRANN finds a (intuitive) negative relationship. Similarly, higher industry production is associated with higher losses in the linear model, but has a negative effect in the G-BRANN model. The last macroeconomic variable delinquency rate has an intuitive positive sign in the linear model, but a counterintuitive relation in the G-BRANN model. This is similar to Gambetti et al. (2019), where the delinquency rate has no significant impact in their best model. The counterintuitive sign might be due to correlations in the macroeconomic variables. The problem of counterintuitive signs when incorporating many of them is well-known in the credit risk literature. Figlewski et al. (2012) find that many macroeconomic variables change their signs and have even statistically significant counterintuitive signs if a large selection of them is included.

Uncertainty related drivers of μ

Figure 5 shows the ALE plots of uncertainty related variables. We find a substantial positive impact of financial uncertainty, which is similar to Gambetti et al. (2019). The effect is linear in G-BRANN and nearly identical with its linear counterpart. Financial uncertainty has the largest effect of the uncertainty related drivers on market-based LGDs. For the text-related news-based EPU index we find a positive relationship, which is close to the linear model from an index level of 100 onwards. CPI uncertainty shows a U-shaped relation indicating that market-based LGDs decrease for low levels of uncertainty but increase sharply after a certain point. FSL uncertainty exhibits a negative effect, which is in line with the linear model, but in G-BRANN the effect is more extreme up to a value of 2. Then its slope is similar to the linear counterpart. Overall, we find larger non-linearities in uncertainty-related variables than in macroeconomic variables.

Bond related drivers of ϕ

Following the definition of precision ϕ, the estimated sign is inversely connected to the variance of the LGDs. As a consequence of this an estimated negative effect increases the variance of the resulting LGD distribution.

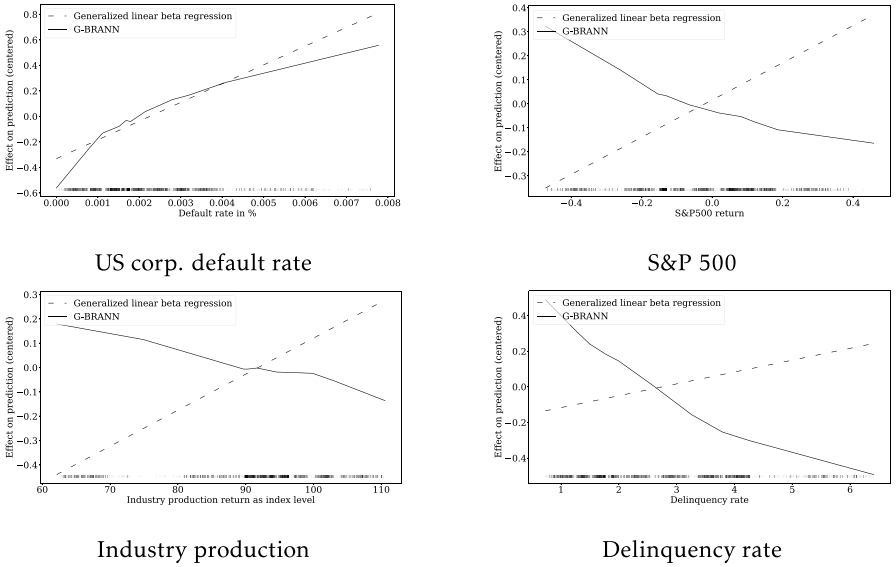


Fig. 4 ALE plots of macroeconomic variables. Notes: The figures show the ALE plots for G-BRANN in solid black and the generalized linear beta regression in dashed black. We initialize the G-BRANN 100 times and calculate the final median ALE plots. We also include a rug plot at the bottom to visualize the distribution of the underlying data

Figure 6 shows the impact of the two bond characteristics. We find a small negative effect of maturity on the precision of market-based LGDs. This implies that bonds with longer maturities are associated with less variance in the LGD estimate. G-BRANN recovers an increasing effect up to a certain coupon rate. Therefore for low coupon rates we have a negative effect on the precision, which becomes less negative as the coupon rate approaches 5%. For coupon rates between 5% and 10% we have a decreasing positive effect on the precision, which becomes slightly negative for higher coupon rates. Due to the inverse relationship between precision and variance we expect higher variances of the LGDs for low and high coupon rates.

Macroeconomic drivers of ϕ

Turning to the influence of the macroeconomy on the precision of market-based LGDs, Fig. 7 illustrates their impact. We find a positive relationship between S &P 500 returns and precision, implying that the variance decreases for higher returns. This is somewhat contrary to Gambetti et al. (2019), but they used the level of the S &P 500 and not the (stationary) returns. This positive relationship for very high returns can be partly explained as we included the recent Covid-19 pandemic in our sample, where we observe large positive returns, although the LGD realization stagnated. The positive effect is less pronounced than its linear counterpart. For default rates we have an increasing effect on the prediction for very small default rates, turning negative afterwards, which is consistent with the linear model. For delinquency rates, we find a similar picture as for the S &P 500 returns. The impact

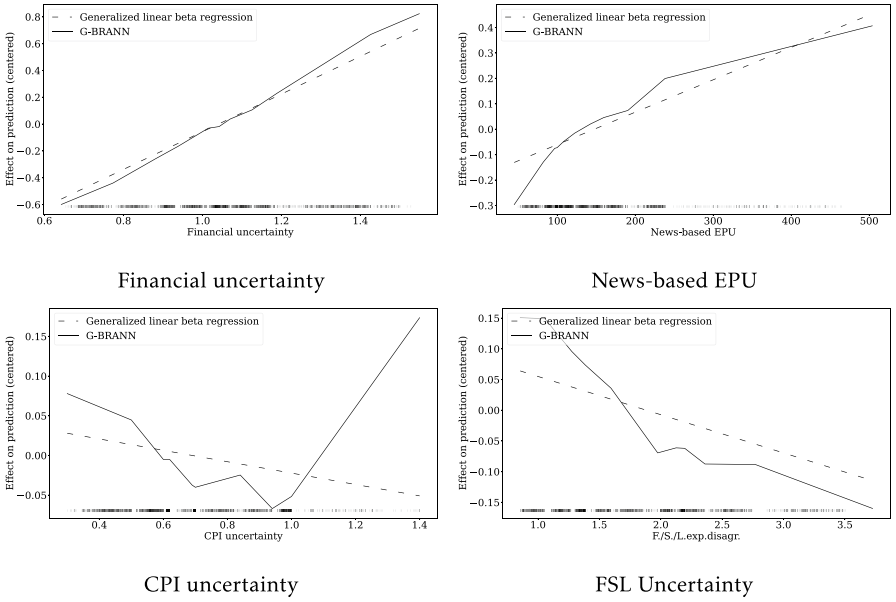


Fig. 5 ALE plots of uncertainty related variables. Notes: The figures show the ALE plots for G-BRANN in solid black and the generalized linear beta regression in dashed black. We initialize the G-BRANN 100 times and calculate the final median ALE plots. We also include a rug plot at the bottom to visualize the distribution of the underlying data

on the prediction has the same direction as the linear model but is more conservative. The industry production has a small, constant negative effect on average.

Uncertainty related drivers of ϕ

In our selection, only two uncertainty-related measures were selected for the final model. We find that financial uncertainty has a almost linear effect on the precision, so that high financial uncertainty correspond to lower variance of the LGD estimation. For this variable the effect is less strong than than the effect modeled by the generalized linear beta regression. CPI uncertainty on the other hand shows a negative trend for increasing uncertainty, but the overall effect is comparable small.

Non-linearity in the estimation of μ and ϕ

ALE plots are a powerful tool to visualize the modeled effects of features on the prediction. Due to the connection of ALE plots to a functional-ANOVA-like-decomposition ALE plots are capable to quantify the goodness of fit to the prediction due to an arbitrary order of effects according to Apley and Zhu (2020). We calculate the the $R^2_{ALE,1}$ for the parameter of the modeled distribution. Therefore we can measure how well the prediction can be approximated by the first order (main) effects, which are visualized in Figs. 3, 4, 5, 6, 7 and 8. For μ the $R^2_{ALE,1,\mu}$ is 0.9017. This means, that 90.17% of the prediction is due to (non-)linear main effects and only the remaining 9.83% are a result of (non)-linear higher order effects such as interactions. This picture changes for ϕ. Here the $R^2_{ALE,1,\phi}$ is 0.3531. Therefore, the most part of the prediction is due to (non-)linear higher order effects. Using the $R^2_{ALE,linear}$ derived by

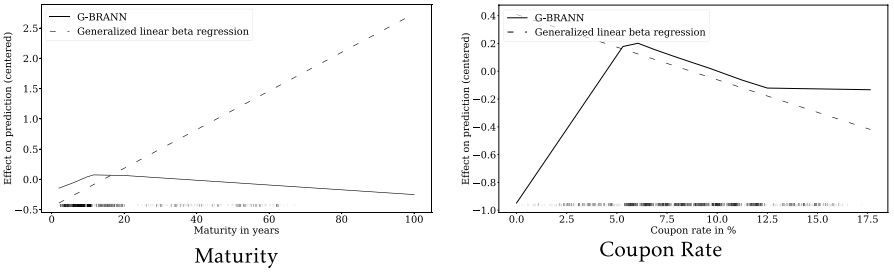


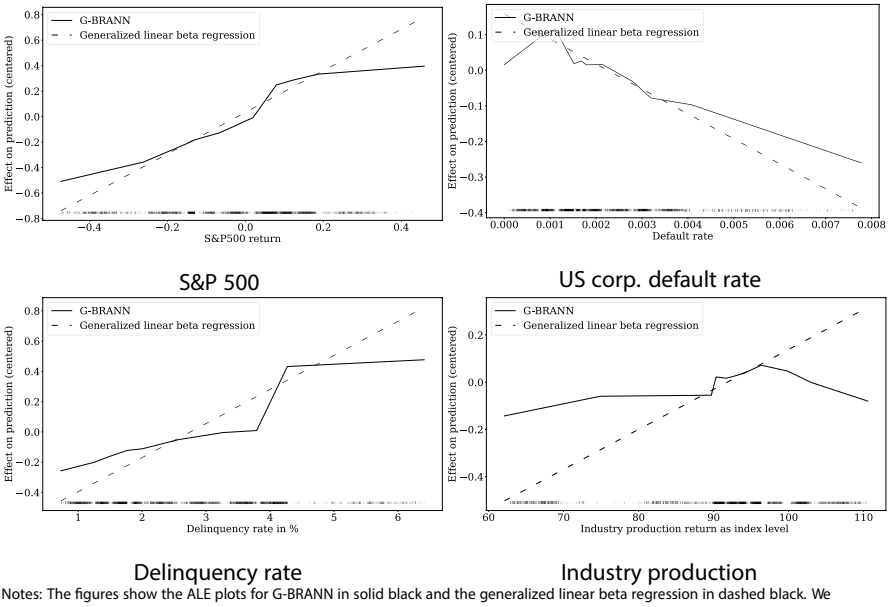
Fig. 6 ALE plots of bond characteristics | ϕ. Notes: The figures show the ALE plots for G-BRANN in solid black and the generalized linear beta regression in dashed black. We initialize the G-BRANN 100 times and calculate the final median ALE plots. We also include a rug plot at the bottom to visualize the distribution of the underlying data

Nagl (2021), we can divide the $R^2_{ALE,1}$ further. As $R^2_{ALE,linear}$ measures the how well the prediction can be explained by linear first order effects, the difference between $R^2_{ALE,1}$ and $R^2_{ALE,linear}$ is the improved approximation by non-linearity in the first order effects. For G-BRANN $R^2_{ALE,linear,\mu}$ is 0.8590, which indicates that the improvement in the first order effects by non-linearity is only about 4%. For the precision parameter this becomes more pronounced. $R^2_{ALE,linear,\phi}$ is only 0.1963, which means, that 80.37% of the ϕ predictions is due to non-linearity and higher order effects. More specific the increase due to non-linearity in the first order effects is more than 15%. Therefore, the non-linearity has a tremendous effect for the estimation of ϕ even in the first order effects.

5.3 Scenario analysis

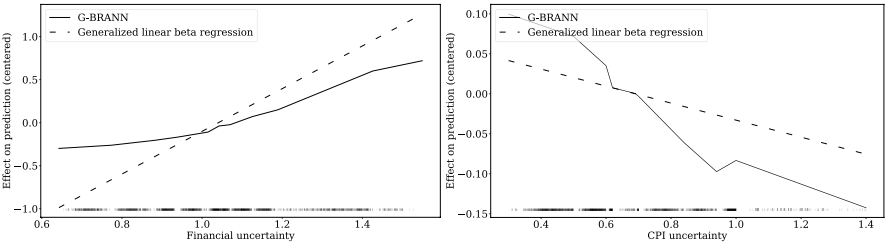
The remaining part of this section focuses on the implications of our findings for risk management. As stated in Kellner et al. (2022) only considering the mean or median does not allow to differentiate between risk profiles in a holistic way. Therefore, the whole distribution should be taken into account to derive risk profiles across possible realizations of the LGD.

Assume a bank aims at investigating the implications of favorable and unfavorable scenarios in their credit risk assessment. These scenarios can be easily derived using low and high quantiles of individual LGD distributions, predicted by G-BRANN. To examine this, we compare the trained beta regression and G-BRANN as described in Sect. 5.1 and predict the μ and ϕ for every sample in the training data. Figure 9 shows the results for the different types of seniority. To obtain a representative distribution of each of them, we take the the median μ and ϕ for every seniority. The left hand side of Fig. 9 shows the estimated distributions by the beta regression and on the right side the distributions calculated by G-BRANN are displayed. In the beta regression models all samples have the same value for ϕ, whereas G-BRANN allows individual values of ϕ. Please recall that the overall fit of G-BRANN in terms of likelihood is considerably higher compared to the beta regression. This holds also for every individual category, such as seniority. Therefore, we are confident that the estimated distributions by G-BRANN are superior as well.



Notes: The figures show the ALE plots for G-BRANN in solid black and the generalized linear beta regression in dashed black. We initialize the G-BRANN 100 times and calculate the final median ALE plots. We also include a rug plot at the bottom to visualize the distribution of the underlying data.

Fig. 7 ALE plots of macroeconomic variables | ϕ. Notes: The figures show the ALE plots for G-BRANN in solid black and the generalized linear beta regression in dashed black. We initialize the G-BRANN 100 times and calculate the final median ALE plots. We also include a rug plot at the bottom to visualize the distribution of the underlying data



Notes: The figures show the ALE plots for G-BRANN in solid black and the generalized linear beta regression in dashed black. We initialize the G-BRANN 100 times and calculate the final median ALE plots. We also include a rug plot at the bottom to visualize the distribution of the underlying data.

Fig. 8 ALE plots of uncertainty related variables | ϕ. Notes: The figures show the ALE plots for G-BRANN in solid black and the generalized linear beta regression in dashed black. We initialize the G-BRANN 100 times and calculate the final median ALE plots. We also include a rug plot at the bottom to visualize the distribution of the underlying data

From a risk manager's perspective a more differentiated picture across categories provides a valuable information to derive individual risk profiles. Therefore, the less theses distributions overlap between categories, the more refined can the derived risk

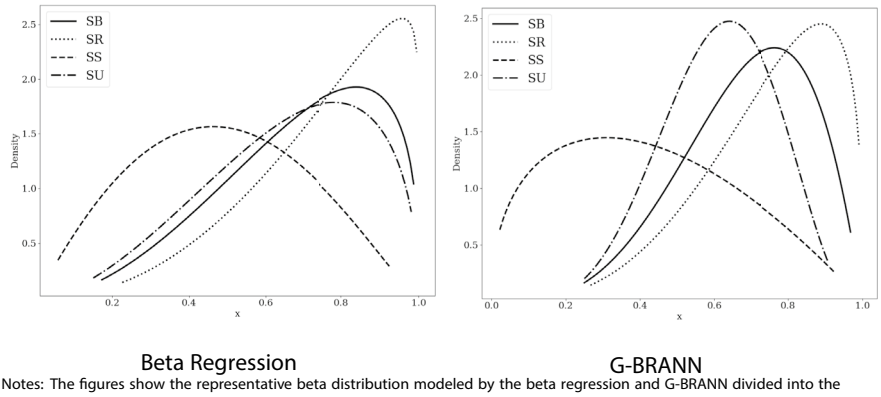


Fig. 9 Beta distributions per seniority. Notes: The figures show the representative beta distribution modeled by the beta regression and G-BRANN divided into the seniorities. SB, SR, SS and SU refer to Senior Subordinated, Subordinated, Senior Secured and Senior Unsecured

profiles be. Overall, using the beta regression the distributions overlap more than using the G-BRANN. Therefore, G-BRANN allows a more refined picture of the different distribution. Again, the fit in terms of likelihood is superior for every category and, thus, the less overlapping distributions suit the data more. To quantify this effect, we calculate the area where the distributions overlap. Since the integral of these distributions is always one, the overlapping area is naturally bounded from zero to one, where one means that one distribution envelops the other distribution. We calculate the overlapping area for every pair of seniority levels. On average the overlapping area of the beta regression is 0.7443 in contrast to 0.6402 for G-BRANN.

We redo the same analysis for the industry types. Typically, the LGDs vary across different industries due to differences in collateralized assets or guarantees. Therefore, a risk manager appreciates a models that allows for a clear distinction between LGDs in different industries. Similar to the seniority, the fit of G-BRANN is superior in every industry compared to the beta regression, which is currently industry standard.

Figure 10 shows the estimated distribution for the most common industries in our sample. Overall, we observe a similar picture. G-BRANN produces much more differentiated distributions than the beta regression.

Calculating the mean overlapping are the beta regression results in 0.6919 and G-BRANN 0.4444. The individual overlappings are even more different. G-BRANN predicts a distribution for the Utilities sector that does not have any overlapping with the Nonbank Finance sector. On the contrary, the beta regression shows a comparable large overlap. In our data, LGDs from the Utilities sector have the lowest average LGD, whereas the Nonbank Finance sector has the highest according to Table 8. Again, this shows that G-BRANN reflects the empirical features of our data much better.

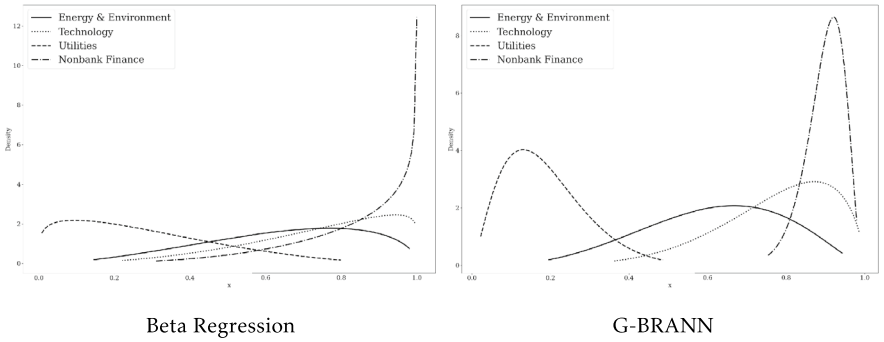


Fig. 10 Beta distributions per industry type. Notes: The figures show the representative beta distribution modeled by the beta regression and G-BRANN divided into the industry types. Here just a selection of industry types are shown

Lastly, risk managers do not only want to differentiate between industry types or seniorities, but also between different macroeconomic states. Therefore, we provide a scenario analysis which focuses on the economic surrounding. We choose three quarters with different average realized LGDs. As "good" scenario we rely on the macroeconomic state in Q1 2004 with an average LGD of 0.43, which is comparatively low. The "average" scenario is Q4 2005 with a mean LGD of 0.62 which is very close to the average of our whole dataset. The "bad" case is Q3 2008, which is a quarter of the Global Financial Crisis that is reflected by the very large mean LGD of 0.90. The "good", "average" and "bad " states are also reflected by the macroeconomic variables, such as the S &P 500 return or the US corp. default rate.

Figure 11 illustrates a clear separation between the good and the bad scenario for G-BRANN, whereas the distributions modeled by the beta regression overlap by 0.3502 compared to 0.0095 for G-BRANN. Again, the fit in every macroeconomic state of G-BRANN is considerably better than by the beta regression and, thus, this clear separation is more aligned with the data. Furthermore, the clear separation between good and bad economic states is also economically more plausible.

Figure 12 visualizes the overlapping area of Fig. 11 to allow for a easy comparison. We observe that the difference between "average" and "bad ", the G-BRANN has less than half of overlapping and for "good " vs. "bad", we see overlapping close to zero. Therefore, G-BRANN offers a data-driven and flexible way to derive tailored scenario analysis for risk management tasks and allows for a clear and economic plausible differentiation between macroeconomic states. The detailed results of every pairwise overlapping for every scenario analysis can be found in Appendix D in Figs. 16 and 17.

6 Conclusion

In recent years a broad stream of literature emerged which shows that mean LGD estimates are non-linearly connected to well-known drivers, see, e.g., Bastos and Matos (2021), Bellotti et al. (2021), Nazemi et al. (2021), Olson et al.

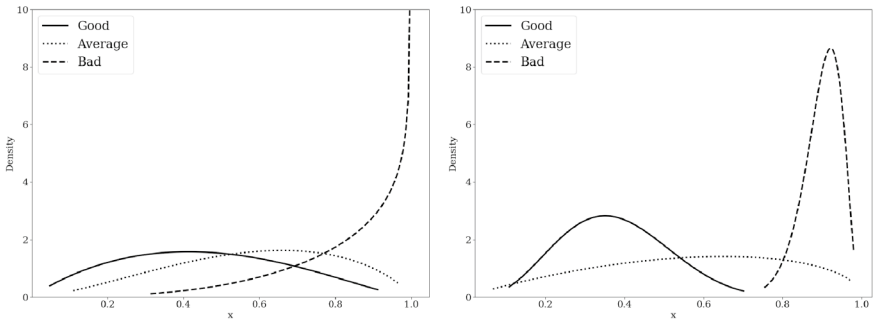


Fig. 11 Beta distributions per macroeconomic states. Notes: The figures show the representative beta distribution modeled by the beta regression and G-BRANN divided into different macroeconomic sates

(2021), Sopitpongstorn et al. (2021) or Xia et al. (2021). The drivers of the LGD distribution's precision (variance) are considerably less investigated as noted by Gambetti et al. (2019). They find that there are several variables with effect on the precision by using a generalized linear beta regression. We extend this approach by allowing non-linearity in mean *and* precision by combining the generalized linear beta regression with a neural network structure. This allows us to incorporate these little-noticed characteristics such as bounded support, skewed distribution, and heteroskedasticity directly into our modeling framework. Furthermore, we derive novel trainable activation functions to address the bounded support problem in the LGD distribution's mean and precision. We implement a data-driven way to characterize the actual shape of the precision predictions which increases the robustness. By accessing Moody's Default and Recovery Database from January 1990 until March 2021, we incorporate the most recent evidence in market-based LGD realizations. We observe strong non-linearity in the prediction of the precision parameter. Therefore, especially this parameter benefits from non-linear modeling.

Modeling the precision and thus, the form of every obligor's LGD distribution enhances the capability of risk managers in several important ways. First, lower and high quantiles can be used to derive good and bad scenarios in a data-driven way. Therefore, the impact of portfolio losses beyond expectation values can be quantified. Hence, our approach provides a flexible and data-driven tool to derive conservative estimates, i.e., higher quantiles, concerning regulator's margin of conservatism. Second, by comparing the individual distributions of obligors, risk managers can reveal differences in the obligor's risk profiles by comparing extreme losses, for example in terms of Value-at-Risk (VaR). This enables banks to better quantify the riskiness of their business in terms of potential losses. Our scenario analysis reveals that the distributions modeled by the beta distribution lack of distinctiveness compared to G-BRANN. Thus, scenario analysis with beta regression could lead to inadequate loss estimation. Furthermore, the application of BRANN and G-BRANN to workout LGDs would be interesting, as they entail similar challenges. As the data on this kind of LGD is considerably larger, one could even consider a multilevel

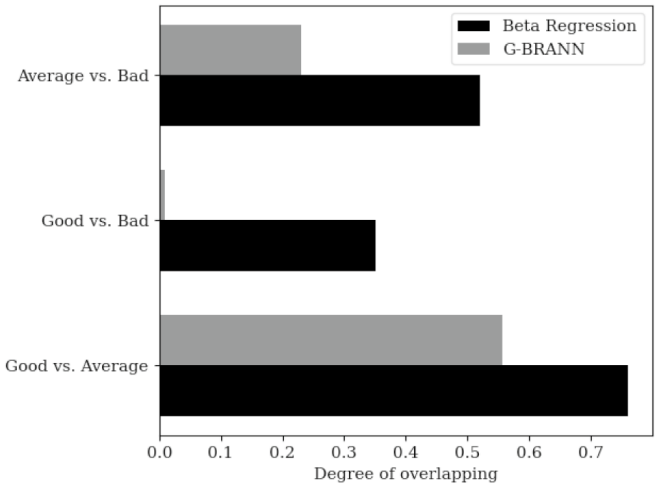


Fig. 12 Degree of overlapping across the macroeconomic states. Notes: The figure shows the degree of overlapping of the representative beta distribution modeled by the beta regression and G-BRANN divided into macroeconomic states

approach by fitting our proposed methods on different seniority levels, industry sectors or default types.[20]

Our approach of combining well-known statistical methods with neural networks and the novel derived activation functions can not only be used for credit risk-related problems but to more general and broader set of problems in business and economics, e.g., demand or retail forecasting.

Descriptive statistics

Table 6 shows the descriptive statistics of LGDs across the whole sample. Over the 2,315 LGDs we have a slightly negative skewed distribution with a median LGD of 68 %. The following tables should provide an overview of the LGD distribution across the categorical values.

Dividing the LGDs in their seniority the picture changes in a few regards. While the skewness for senior secured and senior unsecured remains relatively close to zero, the skewness for senior subordinated and subordinated decreases close to -1, which indicates moderate skewness. Furthermore, the average LGD per category is different. Higher levels of seniority tend to have lower LGDs on average and at median. For the industry sector there are differences as well. High LGDs in particular can be observed for technology and for nonbank finance

[20] We thank an anonymous reviewer for suggesting this potential application of BRANN and G-BRANN.

companies. By far the lowest mean LGD with low standard deviations are located in the utility sector followed by the banking companies. Those two sectors are the only sectors with highly positive skewness. This low mean LGD in the utilities sector corresponds to the high recovery rates in Gambetti et al. (2019). On the contrary low LGDs in the banking sector are quite different from the observed ones by Gambetti et al. (2019), but one must take into account that the used sample size in the banking sector in this paper is more than six times the sample size used in Gambetti et al. (2019). The remaining industry sector show slightly higher or similar LGDs to the average LGD over the whole sample.

Conditioning the LGDs on the default type there are major differences compared to table 6 noticeable. First of all, there are some default types, that barely occur. Some of them occur only once or twice in the observed period of more than three decades. For the slightly bigger categories it is visible that payment moratorium has the lowest average LGD and the smallest standard division by far. The biggest category Chapter 11 shows the second highest average LGD. Only Chapter 7 provides higher average LGD, but also a very small sample size. Most of the conditional distributions are negatively skewed except the category distressed exchange, which is moderately positive skewed and show low average LGDs.

Trainable activation functions

In the following the trainable activation functions from Sect. 4 are represented graphically. Figure 13 illustrates the impact of the steepness parameter a. For a larger value for a, $a = 3.37$, the curve is even steeper than the original exponential function. The value for Fig. 13 is chosen according to the trained G-BRANN in table 4. The trainable parameter q of the trainable softplus function in Fig. 14 controls for the curvature of the function. For increasing q the trainable softplus function tends towards a relu activation function.

The last of the proposed trainable activation functions is the trainable sigmoid function. Setting the parameters $c = 0$, $s = -1$ and $h = 1$ it results in the ordinary sigmoid function bounding the output on an intervall from 0 to 1. The parameter c shifts the function vertically as displayed by the upper function of Fig. 15. Increasing the scale parameter s towards 0 the output tends to flatten the input so that changes in the input less affect the output. The last parameter h defines the upper bound of the output, so that the trainable sigmoid function can result in higher values than the ordinary sigmoid function. The following figure compares the original sigmoid function with the sigmoid function, which is trained for the robustness section. Here the chosen parameters are $c = 1.88$, $s = -1.45$ and $h = 2.68$.

Table 6 Descriptive statistics of LGDs across the whole sample

	N	Min	Median	Mean	Max	St.Dev	Skewness
LGD	2315	0.50	68.00	61.40	99.99	28.11	− 0.35

All displayed values except the sample size are expressed as percentages

Table 7 Descriptive statistics of LGDs according to the seniority of the defaulted bond

	N	Min	Median	Mean	Max	St.Dev	Skewness
Senior secured	195	0.50	47.5	49.42	99.25	28.91	0.0594
Senior unsecured	1599	0.50	65.0	59.59	99.97	28.36	− 0.2206
Senior subordinated	360	0.50	79.0	72.02	99.99	23.87	− 0.9923
Subordinated	161	0.87	74.0	70.17	99.87	23.74	− 0.9030

Robustness

Overall, this paper does not seek to conduct a horse race and contributes to the literature by finding a new best method of predicting market-based LGDs. Our aim is at the rationale behind the drivers of *mean* LGDs and the drivers of the *precision* of the LGDs, which is considerably less investigated. Hence, our neural networks can be seen as explanation models to unveil drivers of the mean and variance of the market-based LGDs. We want to use these relations to generate estimates in scenario analysis and derive implications. Therefore, they are not built as a *prediction* model for predicting future *mean* LGDs, but as a non-parametric way to unveil hidden relationships between drivers and the market-based LGD distribution. However, to counteract the concerns that our derived neural networks are not suitable to predict future realizations, we conduct an out-of-time benchmark exercise.

For this purpose we split the the whole sample into a training sample and a test sample. For the training sample only LGDs until end of 2007 are used. The remaining observations act therefore as an out-of-time sample. For BRANN, G-BRANN, Neural Networks and Random Forests we optimized the hyperparameters using 5 fold cross validation and run each fold 10 times to ensure stable results for every fold. For all models with hyperparameters we draw 500 constellations each by a random search approach. Table 10 shows the setup and the resulting parameters for the robustness section.

For each model type we choose those hyperparameters, which return the lowest negative log likelihood or mean squared error (MSE), respectively, averaged over the 5 hold out folds. The (extended) beta regression and (G-) BRANN are optimized by the mean of the negative log likelihood in Eq. (4), while the objective of the remaining models is to minimize the MSE. Since the Pseudo-R^2 is based solely on the μ-part of the (extended) beta regression and (G-) BRANN, which represents the predicted LGD, it can be calculated for all models.

Table 8 Descriptive statistics of LGDs according to the industry sector of the defaulted bond

	N	Min	Median	Mean	Max	St.Dev	Skewness
Banking	276.0	7.92	28.96	35.52	99.75	19.13	2.6608
Capital industries	471.0	0.75	72.50	66.58	99.87	25.51	− 0.6049
Consumer industries	307.0	0.50	71.50	63.78	99.99	26.07	− 0.6349
Energy & environment	296.0	0.50	66.25	63.37	99.97	24.12	− 0.5381
Media & publishing	164.0	1.00	56.50	55.70	99.99	28.96	− 0.0344
Nonbank finance	261.0	14.00	90.00	74.06	99.87	29.63	− 1.3282
REIT	17.0	36.65	76.48	66.03	98.12	21.37	− 0.0421
Retail & distribution	164.0	0.50	68.25	63.85	99.50	25.45	− 0.6727
Technology	224.0	1.00	79.75	71.46	99.62	25.34	− 1.2007
Transportation	84.0	4.75	77.75	66.94	98.25	23.72	− 0.8157
Utilities	51.0	6.25	16.00	18.84	80.00	12.65	2.7352

Table 9 Descriptive statistics of LGDs according to the default type

	N	Min	Median	Mean	Max	St.Dev	Skewness
Chapter 11	749	0.75	85.00	73.84	99.99	25.14	− 1.2891
Chapter 7	7	54.00	96.00	89.32	99.47	16.19	− 2.2681
Distressed exchange	554	0.50	29.00	39.82	98.00	21.33	0.8111
Grace period default	26	2.00	49.94	46.17	92.00	22.22	− 0.0041
Missed interest payment	700	1.00	73.50	66.70	99.99	24.19	− 0.6654
Others	94	1.00	67.00	62.36	99.75	29.36	− 0.3277
Payment moratorium	35	14.87	16.83	16.79	17.95	0.55	− 1.8553
Prepackaged chapter 11	150	0.50	76.75	65.41	99.64	28.81	− 0.5668

For comparability some categories are displayed consolidated, but feed to the network separately

However, this does not apply to the log likelihood calculation. For the out-of-time comparison we form a portfolio of 100 randomly drawn bonds and evaluate the MSE, Pseudo-R^2 and, if possible, the log likelihood. This procedure was repeated 10 times and their average is provided in table 11.

Comparing the values for the log likelihood in- and out-of-time, we observe that in both samples one of our neural network approaches outperform the linear beta regressions. Therefore, we can argue that the good performance illustrated in Table 5 can be recovered when we focus on future predictions. Overall the non-linear models recover the distribution of market-based LGDs best. While focusing only on the μ -part, i.e. only on mean predictions, we observe that the Random Forest performs best in-sample and third best out-of-time. This is similar to findings in the literature, see, e.g., Kaposty et al. (2020); Bellotti et al. (2021); Nazemi et al. (2021). However, as previously noted, the aim of this paper is not to predict the mean of market-based LGDs best as done by various studies, e.g., Bastos (2010); Loterman et al. (2012); Qi and Zhao (2012); Bastos and Matos (2021); Olson et al. (2021); Nazemi et al. (2021) among many others. The contribution of this paper is to model the precision of the

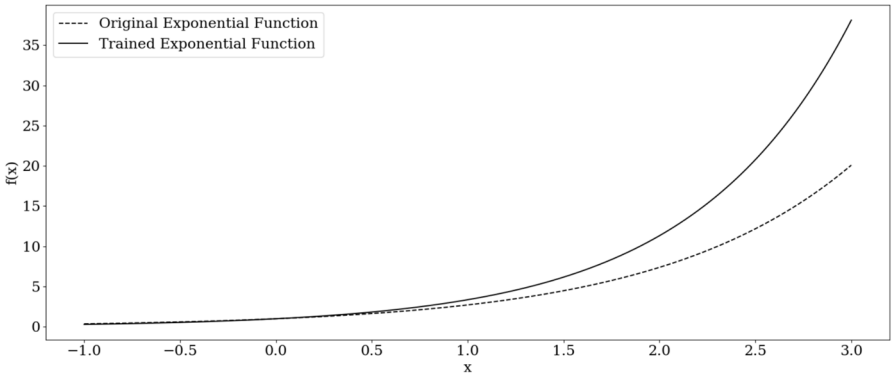


Fig. 13 Trainable exponential function

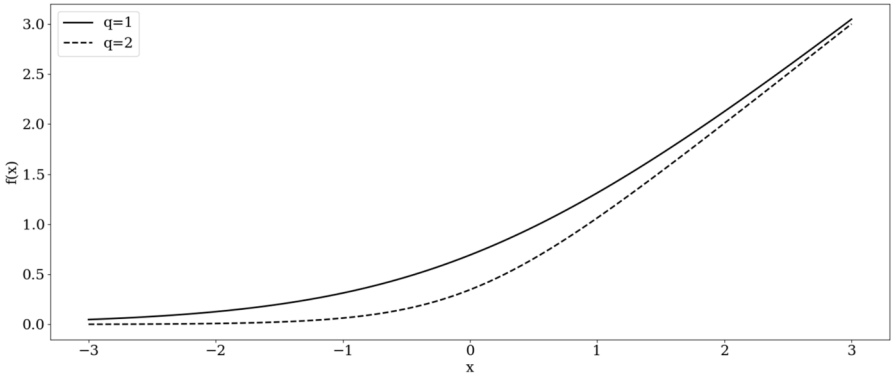


Fig. 14 Trainable softplus function

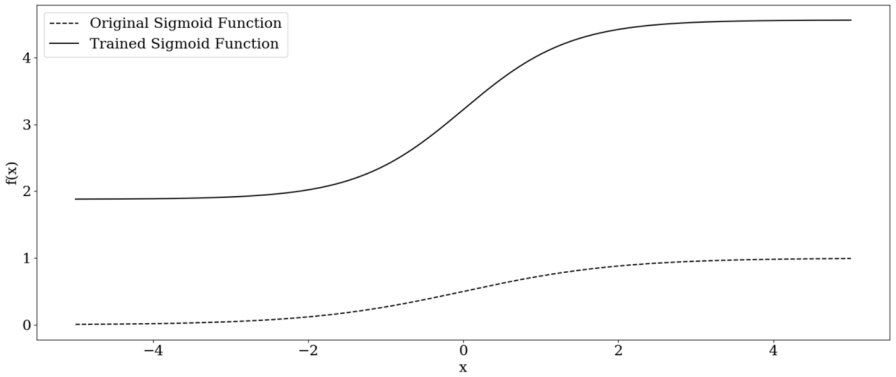


Fig. 15 Trainable sigmoid function

Table 10 Setup and final values of the hyperparameter search - robustness

Model	Parameter	Distribution	Final parameter
BRANN	Learning rate	$U^c \sim [0.001, 0.1]$	0.0893
	Dropout rate	$U^c \sim [0.05, 0.50]$	0.3289
	Hidden layer	$U^d \sim [1, 2]$	1
	Multiple	$U^d \sim [1, 8]$	1
G-BRANN	Learning rate	$U^c \sim [0.001, 0.1]$	0.0636
	Dropout rate μ	$U^c \sim [0.05, 0.50]$	0.4314
	Dropout rate ϕ	$U^c \sim [0.05, 0.50]$	0.1827
	Hidden layer μ	$U^d \sim [1, 2]$	2
	Hidden layer ϕ	$U^d \sim [1, 2]$	2
	Multiple μ	$U^d \sim [1, 8]$	1
	Multiple ϕ	$U^d \sim [1, 8]$	7
	Activation function	t-exp, t-sig, t-soft	t-sig
Neural network	Learning rate	$U^c \sim [0.001, 0.1]$	0.0796
	Dropout rate	$U^c \sim [0.05, 0.50]$	0.4951
	Hidden layer	$U^d \sim [1, 2]$	1
	Multiple	$U^d \sim [1, 8]$	6
Random forest	Number trees	$U^d \sim [10, 250]$	90
	Splitsamples	$U^d \sim [2, 10]$	3
	Leafsamples	$U^d \sim [1, 10]$	1
Regression tree	Splitsamples	$U^d \sim [2, 10]$	7
	Leafsamples	$U^d \sim [1, 10]$	10
Ridge regression	Regularizationparameter	$U^c \sim [0.0, 10]$	0.0023
Lasso regression	Regularizationparameter	$U^c \sim [0.0, 0.02]$	0.0001
Elastic net	Ratio	$U^c \sim [0.0, 1]$	0.1711
	Regularizationparameter	$U^c \sim [0.0, 0.02]$	0.0129

The table shows different values for the hyperparameter search. U^c labels the continuous uniform distribution, whereas U^d labels the discrete uniform distribution, in which the upper bound was excluded. For the random forest and the regression tree the splitsamples and the leafsamples refer to the minimum number of samples to split respectively to include in a leaf

market-based LGDs distribution in a straightforward non-linear way, which is a novelty in the literature of LGD modeling.

Scenario analysis

The following provides the detailed results of our scenario analysis in Sect. 5. The estimated overlappings are illustrated via heatmaps. The lower triangle refers to the results of G-BRANN and the upper triangle to the results of the beta regression. In general, a lower value refers to less overlapping and vice versa.

Table 11 Evaluation Metrics

	LogLikelihood	MSE	Pseudo-R^2
(a) In Sample			
Beta regression	434.178	0.039	0.433
BRANN	441.049	0.039	0.447
Generalized beta regression	**487.562**	0.039	0.423
G-BRANN	478.603	0.037	0.485
Neural network	–	0.045	0.398
Random forest	–	**0.005**	**0.948**
Regression tree	–	0.023	0.681
Linear regression	–	0.039	0.472
Ridge regression	–	0.040	0.459
Lasso regression	–	0.039	0.471
Elastic net	–	0.044	0.417
(b) Out of time			
Beta regression	15.723	0.057	0.339
BRANN	8.838	0.061	0.262
Generalized beta regression	– 21.463	0.056	0.361
G-BRANN	**20.651**	0.058	0.286
Neural network	–	0.115	0.173
Random forest	–	0.053	0.426
Regression tree	–	0.068	0.300
Linear regression	–	0.062	0.379
Ridge regression	–	0.052	0.450
Lasso regression	–	0.061	0.386
Elastic net	–	**0.047**	**0.461**

Bold values indicate best values

Figure 16 shows the overlapping estimates for the different seniority types and macroeconomic states. Overall, we can see that G-BRANN has in every constellation a lower overlapping, which implies that GBRANN helps to differentiate between seniority and industry type better than the standard beta regression. Please recall that the fit of G-BRANN is better in any seniority type or macroeconomic state. Therefore, we argue that the less overlapping better represents the underlying data.

Figure 17 shows the overlapping estimates for the different industry types. Interestingly, G-BRANN shows very less overlapping for the banking industry with any other industry type. On the contrary, the Beta Regression shows medium overlapping. It is well known that the banking industry differs from other industry types due to their special business model and their impact on financial stability. It seems that the difference is also visible in the LGD estimates in our sample. Similar to Fig. 16 G-BRANN shows considerable less overlappings and, thus, allows for a better differentiation between industry types.

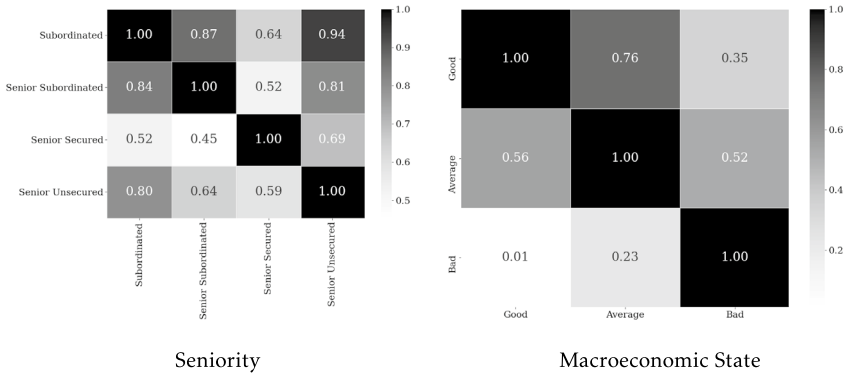


Fig. 16 Overlappings | Seniority & Macroeconomic State. Notes: The figures show the overlapping of the estimated distributions of GBRANN on the lower triangle and the Beta Regression on the upper triangle. A lower value indicates less overlapping

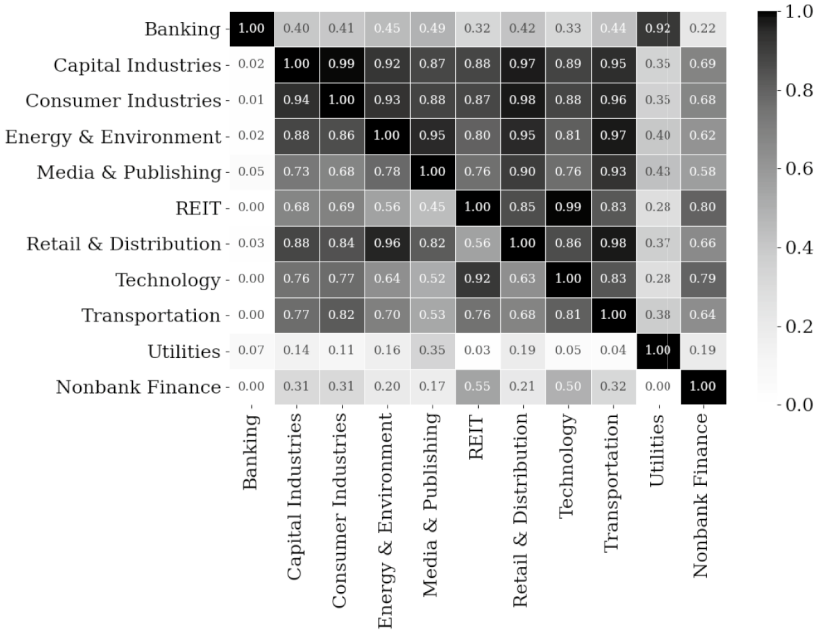


Fig. 17 Overlappings | Industry. Notes: The figure shows the overlapping of the estimated distributions of GBRANN on the lower triangle and the Beta Regression on the upper triangle. A lower value indicates less overlapping

Funding Open Access funding enabled and organized by Projekt DEAL.

References

Altman EI, Kalotay EA (2014) Ultimate recovery mixtures. J. Bank. Finance 40:116–129

Apicella A, Donnarumma F, Isgrò F, Prevete R (2021) A survey on modern trainable activation functions Neural Networks 138:14–32. https://doi.org/10.1016/j.neunet.2021.01.026

Apley DW, Zhu J (2020) Visualizing the effects of predictor variables in black box supervised learning models. J. R. Stat. Soc. Ser. B (Stat. Methodol.) 82(4):1059–1086

Baker SR, Bloom N, Davis SJ (2016) Measuring economic policy uncertainty. Q. J. Econ. 131(4):1593–1636

Barbaglia L, Manzan S, Tosetti E (2021) Forecasting loan default in Europe with machine learning. J. Financ. Econom. 21(2):569–96

Basel Committee on Banking Supervision (2017) Basel III: Finalising post-crisis reforms. Technical report, Bank for International Settlements

Bastos JA (2010) Forecasting bank loans loss-given-default. J. Bank. Finance 34(10):2510–2517

Bastos JA, Matos SM (2021) Explainable models of credit losses. Eur. J. Operat. Res. 301(1):386–94

Belkin M, Hsu D, Ma S, Mandal S (2019) Reconciling modern machine-learning practice and the classical bias-variance trade-off. Proc. Natl. Acad. Sci. 116(32):15849–15854

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

Betz J, Kellner R, Rösch D (2018) Systematic effects among loss given defaults and their implications on downturn estimation. Eur. J. Oper. Res. 271(3):1113–1144

Calabrese R (2014) Downturn loss given default: Mixture distribution estimation. Eur. J. Oper. Res. 237(1):271–277

European Banking Authority (2022) Risk Assessment of the European Banking System. Technical report

Ferrari S, Cribari-Neto F (2004) Beta regression for modelling rates and proportions. J. Appl. Stat. 31(7):799–815

Figlewski S, Frydman H, Liang W (2012) Modeling the effect of macroeconomic factors on corporate default and credit rating transitions. Int. Rev. Econ. Finance 21(1):87–105

Gambetti P, Gauthier G, Vrins F (2019) Recovery rates: Uncertainty certainly matters. J. Bank. Finance 106:371–383

Gu S, Kelly B, Xiu D (2020) Empirical asset pricing via machine learning. Rev. Financ. Stud. 33(5):2223–2273

Gunnarsson BR, van den Broucke S, Baesens B, Óskarsdóttir M, Lemahieu W (2021) Deep learning for credit scoring: Do or don't? Eur. J. Oper. Res. 295(1):292–305

Gürtler M, Zöllner M (2023) Heterogeneities among credit risk parameter distributions: The modality defines the best estimation method. OR Spectr. 45(1):251–287

Görgen, K., Nazemi, A., Schienle, M. (2022) Robust Knockoffs for Controlling False Discoveries With an Application to Bond Recovery Rates

He, K., Zhang, X., Ren, S., Sun, J. (2015a) Deep Residual Learning for Image Recognition. arXiv:1512.03385 [cs]

He, K., Zhang, X/, Ren, S., Sun, J. (2015b) Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision, pp 1026–1034

Hu Z (1992) The study of neural network adaptive control systems. Control Decis. 7:361–366

Hwang R-C, Chu C-K (2018) A logistic regression point of view toward loss given default distribution estimation. Quant. Finance 18(3):419–435

Hwang R-C, Chu C-K, Yu K (2020) Predicting LGD distributions with mixed continuous and discrete ordinal outcomes. Int. J. Forecast. 36(3):1003–22
Jankowitsch R, Nagler F, Subrahmanyam MG (2014) The determinants of recovery rates in the US corporate bond market. J. Financ. Econ 114(1):155–177
Jurado K, Ludvigson SC, Ng S (2015) Measuring uncertainty. Am. Econom. Rev. 105(3):1177–1216
Kalotay EA, Altman EI (2017) Intertemporal forecasts of defaulted bond recoveries and portfolio losses. Rev. Finance 21(1):433–463
Kaposty F, Kriebel J, Löderbusch M (2020) Predicting loss given default in leasing: A closer look at models and variable selection. Int. J. Forecast 36(2):248–266
Kellner R, Nagl M, Rösch D (2022) Opening the black box-Quantile neural networks for loss given default prediction. J. Bank. Finance 134:106334
Krüger S, Rösch D (2017) Downturn LGD modeling using quantile regression. J. Bank. Finance 79:42–56
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
Ludvigson SC, Ma S, Ng S (2021) Uncertainty and business cycles: Exogenous impulse or endogenous response? Am. Econ. J. Macroecon. 13(4):369–410
Nagl, M. (2023) Does non-linearity in risk premiums vary over time? Working paper
Nagl M, Nagl M, Rösch D (2022) Quantifying uncertainty of machine learning methods for loss given default. Front. Appl. Math. Stat. 8:1076083
Nazemi A, Baumann F, Fabozzi FJ (2021) Intertemporal defaulted bond recoveries prediction via machine learning. Eur. J. Operat. Res. 297(3):1162–77
Olson LM, Qi M, Zhang X, Zhao X (2021) Machine learning loss given default for corporate debt. J. Empir. Financ. 64:144–159
Qi M, Zhao X (2012) Comparison of modeling methods for loss given default. J. Bank. Finance 35(11):2842–2855
SIFMA Research (2022) SIFMA Research Quarterly -1Q22 US Fixed Income Markets - Outstanding. Technical report, Securities Industry and Financial Markets Association (SIFMA)
Simas AB, Barreto-Souza W, Rocha AV (2010) Improved estimators for a general class of beta regression models. Comput. Stat. Data Anal. 54(2):348–366
Smithson M, Verkuilen J (2006) A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. Psychol. Methods 11(1):54–71
Sopitpongstorn N, Silvapulle P, Gao J, Fenech J-P (2021) Local logit regression for loan recovery rate. J. Bank. Finance 126:106093
Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: A simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15(1):1929–1958
Xia Y, Zhao J, He L, Li Y, Yang X (2021) Forecasting loss given default for peer-to-peer loans via heterogeneous stacking ensemble approach. Int. J. Forecast 37(4):1590–1613
Yang Z, Yu Y, You C, Steinhardt J, Ma Y (2020) Rethinking Bias-Variance Trade-off for Generalization of Neural Networks. In: Proceedings of the 37th International Conference on Machine Learning. PMLR, pp 10767–10777