Proceedings of the 10th Polish Symposium on Physics in Economy and Social Sciences (FENS 2019)

Application of Machine Learning Method under IFRS 9 Approach to LGD Modeling

U. Grzybowska* and M. Karwański

Department of Applied Mathematics, SGGW-Warsaw University of Life Sciences, ul. Nowoursynowska 159, 02-776 Warsaw, Poland

Doi: 10.12693/APhysPolA.138.116 *e-mail: urszula_grzybowska@sggw.edu.pl

The aim of our research was to show new methods that can successfully be applied by banks in their internal risk calculations. The methods concern one of the key risk parameters, Loss Given Default (LGD). The proposed approach is admissible under IFRS 9 standard. We have applied gradient boosting algorithm which is a classification algorithm and a transitional generalized linear model to forecast LGD values based on explanatory variables and lagged LGD values. We have introduced a Markov chain structure into our data and built an infinitesimal generator to forecast LGD values based on migration matrices for any period $t > 0$. Performance of both applied methods was examined by ROC curves. The calculations were done on real data in SAS 9.4.

topics: LDG, gradient boosting, Markov chains, transitional GLMM.

1. Introduction

The research has been motivated by introduction of new International Financial Reporting Standard 9 (IFRS 9) which is in force since January 1st, 2018, [1]. IFRS 9 was issued in 2014 and has replaced International Accounting Standard 39 (IAS 39) as a response to the Global Financial Crisis of 2007–2010. The main aim was to enforce adequate provisions in line with expected rather than incurred losses. IFRS 9 requires a three-stage approach that results in a new expected loss model. Our objective is to propose a method for calculating long term LGD (behavioral life-time estimation) that is consistent with IFRS 9 objectives. LGD is the amount of money a bank or other financial institution loses when a borrower defaults on a loan.

Financial institutions build loan loss provisions for expected losses due to default or impairment. Expected losses are calculated using the following formula:

$$\mathrm{ECL} = \mathrm{PD} \times \mathrm{EAD} \times \mathrm{LGD}, \tag{1}$$

where ECL is Expected Credit Loss, PD is Probability of Default, EAD is Exposure at Default (an estimation of the amount outstanding in case of default), LGD is the estimated percentage of exposure that the bank loses in case of borrower's default.

IFRS 9 standard assumes a three stage or three bucket model for impairment [2]. Stage 1 consists of financial instruments that are low risk, have low credit risk at the reporting date or have not had a significant increase in credit risk since initial recognition. For the assets in this bucket one–year ECL is calculated, i.e., the expected loss that results from default possible within one year after the reporting date. Stage 2 consists of financial instruments that have experienced a significant increase in credit risk since initial recognition but do not have objective evidence of impairment. The bucket in Stage 3 includes assets that show objective evidence of impairment at the reporting date. For the assets in Stage 2 and 3 lifetime LGD is calculated. In our research we are interested in modeling expected lifetime LGD for instruments in Stage 2.

LGD=1 - $\mathrm{Recover}_{\mathrm{rate}}$ is expressed in percent, i.e., 0% means no loss (0), 100% means total loss (1). LGD probability distribution function in general exhibits non normal behavior. It is usually a bimodal distribution with point masses at 0 and 1. The LGD values can both decrease or increase in time. LGD modelling is not an easy task. In contrast to PD modeling there has not much research been done so far in this field.

Tiziano Bellini in his book [3] described main ideas and possibilities to be developed within IFRS 9 standard for LGD modeling. The aim of methods discussed there are severity identification, (average) LGD estimation and variables' selection. The most typical applications include Tobit regression, Beta regression, zero-one inflated mixture models [3, 4]. Also hazard models can be applied. Bellini also considers application of machine learning methods to variables' selection and prediction of LGD. The methods discussed are decision trees, random forests, and boosting methods. The ideas proposed by Bellini and described in his book should be developed but also new approaches should be tested. Especially machine learning methods and their application requires more attention. Due to newly relaxed restrictions previously imposed by supervising authorities the application of machine learning in LGD estimation is allowed now but it is still a new topic of research.

The aim of our research was to show that machine learning methods can successfully be applied in LGD estimation. We have applied gradient boosting algorithm which is a classification algorithm. As our data has got a time series structure and machine learning algorithms are not a suitable tool for dealing with long-run time dependent variables, we introduce Markov chain structure into our data. The objective variable is LGD and its values will be divided to form Markov chain states. We have also applied regression transitional model as a benchmark. Both methods include multistage approach and are novel in credit risk management.

The paper is organized as follows: we first describe our data. Then, in Sect. 3 we explain in details the idea of our multistage modeling. Section 4 is devoted to description of our results. In Sect. 5 we present the conclusions of our research and some ideas that can be developed.

2. Data

The data was collected in one of Polish banks in the years 2013–2015 for credit transactions. It concerned overdrafts in the segment of Small and Medium-sized Enterprises (SME). As a target variable the LGD values for pre-defaulted cases were used. They were calculated in internal bank models. We chose a pool of 4364 observations describing 358 objects (clients) of one business line. A detailed description of the data selection and cleaning process is beyond the scope of this manuscript. LGD values should range from 0 to 1. Values beyond this range were floored to zero or one. The model building procedure has been started by specifying the maximum model. This model used all 185 variables both these taken into account in bank's internal models and also variables indicated by business experts. Subsequently, any other model considered was created by deleting variables from the maximum model according to specified accuracy measure. We have used the Akaike Information Criterion (AIC) as a model fit measure. Additional attention was paid to the possibility of undesirable collinearity of variables. This allowed the automation of the variable selection process. The final model consisted of 37 variables.

The LGD values are measured on interval scale. We have divided them into disjoint rating categories to build migration matrix states. We have used a pseudo quantile binning to create classes of equal number of observations [5]. For the split for 8 bins it was: A01 $= [0, 0.205)$, A02 $= [0.205, 0.389)$, A03 $= [0.389, 0.462)$, A04 $= [0.462, 0.576)$, A05 $= [0.576, 0.698)$, A06 $= [0.698, 0.77)$, A07 $= [0.77, 0.9)$, and A08 $= [0.9, 1]$.

3. Methods

Markov chains have already been successfully applied in PD modeling [6–8]. Our aim was to apply migration matrices for lifetime LGD forecasting required by IFRS 9. The transition matrices encountered in LGD modeling are not absorbing. Therefore, results obtained for PD estimation cannot be directly applied here. Not much research has been done so far neither is application of Markov chains in LGD estimation known to us.

The most common route is to calculate changes in probability using a discrete timescale cohort method [6, 7]. In fact, our method is compatible with IFRS 9 and it relies on migration matrices using a continuous timescale. We assume that the ratings follow Markov chain and we apply two methods to evaluate (based on available data) a one period migration matrix and use it to obtain an infinitesimal generator Λ of this Markov chain. The infinitesimal generator Λ will be used to forecast LGD values based on migration matrices for any period $t \in (0, T]$ [9].

Once the generator Λ is known, a t-period transition matrix $P(t)$ for any $t > 0$ can be obtained by the formula

$$P(t) = \mathrm{e}^{(t\Lambda)} = \sum_{n=0}^{\infty} \frac{(t\Lambda)^n}{n!} = I + (t\Lambda) + \frac{(t\Lambda)^2}{2} + \ldots \tag{2}$$

Our aim was to build models that for given initial classes A01,...,A08 and the set of observed explanatory variables would predict future LGD values. That goes along with IFRS 9 expectations for lifetime LGD calculation. We have first calculated a one period transition probability matrix based on obtained division for relevant classes by a cohort method (Table I), [6, 7].

The entries of this matrix p_{ij} are probabilities of migrating from class A0i to A0j in one period. To forecast LGD values we have however used a matrix evaluated based on the whole set of data, where the outcome ratings were predicted based on given values of explanatory variables and lagged LGD values. It has got to be mentioned here that due to the structure of the migration matrix, with very high values on the diagonal, meaning only few migrations between ratings, one cannot model directly LGD ratings as outcomes of lagged ratings and covariates in machine learning approach. Therefore, in machine learning we have used LGD values, not ratings, as outcome of the model taking as input covariates and lagged LDG values. The predicted outcome values were divided into LGD ratings and migration matrix was calculated based on predicted outcomes and initial LGD ratings. Afterwards infinitesimal generator Λ was calculated, which then was used to calculate migration matrices at any time t (2). The entries of these matrices are probabilities that the LGD value will be in a given ratings A0j after the time t, provided at the beginning it was in a given class A0i.

TABLE I

Migration matrix evaluated by cohort method (only terms > 0.01% are shown)

	A01	A02	A03	A04	A05	A06	A07	A08
A01	96.07%	2.42%	1.49%					
A02	3.21%	94.97%	1.80%					
A03	0.59%	0.19%	96.25%	2.75%	0.19%			
A04			1.50%	96.98%	1.50%			
A05		1.25%	0.17%	0.71%	95.16%	2.32%		0.35%
A06					5.56%	91.09%	1.55%	1.78%
A07					1.36%	5.47%	93.15%	
A08					0.62%			99.37%

At first, we have applied a regression model. The regression model was a transitional generalized linear model (transitional GLMM), which is an extension of the regressive logistic model introduced by Bonney and described in [10]. We have applied gradient boosting as a machine learning method. It was introduced first in 2001 by J. Friedman [11]. The algorithm MART is described in [12].

Boosting is a way of fitting an additive expansion in a set of elementary basis functions which take the form

$$f\left(x\right)=\sum_{m=1}^{M}\beta_m b\left(x,\gamma_m\right). \tag{3}$$

In gradient boosting implementation, the tree predictions $T\left(x,\Theta\right)$ are used as basis functions. Typically these models are fit by minimizing a loss function, i.e.,

$$\sum_{(\hat{x},\hat{y})} L\Big(f_{m-1}(\hat{x})+b_m(\hat{x}),\hat{y}\Big)\approx\sum_{(\hat{x},\hat{y})} L\Big(f_{m-1}(\hat{x}),\hat{y}\Big)$$
$$\times\sum_{(\hat{x},\hat{y})}\frac{\mathrm{d}}{\mathrm{d}f}L\Big(f_{m-1}(\hat{x}),\hat{y}\Big)b_m\left(\hat{x}\right). \tag{4}$$

At each iteration m, one solves for the optimal basis function $b(x,\gamma_m)$ and corresponding coefficient β_m to add to the current expansion $f_{m-1}(x)$. This produces $f_m(x)$, and the process is repeated. Previously added terms are not modified.

One of commonly used loss functions is squared-error loss. Robust criteria, such as absolute loss, perform sometimes better. One such criterion is the Huber loss criterion used for M-regression:

$$L\left(y,f(x)\right)=$$
$$\begin{cases}\left|y-f\left(x\right)\right|^2, & \text{if } \left|y-f(x)\right|\le\delta\\ 2\delta\left|y-f(x)\right|-\delta^2, & \text{otherwise.}\end{cases} \tag{5}$$

The size of the constituent trees, $J\in\Theta$, is one of the key parameters of the algorithm. Another critical parameter is the number of boosting iterations M (regularization). Controlling the value of M is not the only possible regularization strategy. More sophisticated and efficient shrinkage techniques add a parameter norm penalty $\Omega\left(\theta\right)$ to the loss function $\tilde{L}_\theta\big(y,f\left(x\right)\big)=L_\theta\big(y,f\left(x\right)\big)+\nu\Omega\left(\theta\right)$. In this article the authors used $\Omega\left(\theta\right)=\frac{1}{2}\left|w\right|_2^2$, where w is commonly known as weight decay. The hyperparameter ν can be regarded as controlling the learning rate of the boosting procedure. At the end γ_j's are the fixed constants weights responsible for the optimal classification. Thus, a tree can be formally expressed as

$$T\left(x,\Theta\right)=\sum_{j=1}^{J}\gamma_j I(x)\quad\left(x\epsilon R_j\right), \tag{6}$$

where I is the indicator function.

In our calculations we have used squared error as a loss function, $\delta=0$, leaf size J equal to 15, and we have applied shrinkage parameter $\nu=0.2$. The training set consisted of 70% of observations. The performance of methods was examined by ROC curve [13], which is a popular tool applied for evaluating models with binary outcomes. The values of AUC [14] reflect numerically the performance of classification. The higher the value the better the method. Our outcomes are however not binary. Therefore, we have measured the performance of classification to the given base outcome rating (A01 to A08) against all remaining ratings.

4. Results

We have applied two methods to calculate migration matrices and infinitesimal generators of Markov chains. The migration matrices (Table II and Table III) we have evaluated differ from the initial transition matrix (Table I). The entries (but the first one) of the matrix obtained by gradient boosting (Table II) are much closer to the entries of the initial matrix (Table I) than the entries of the matrix given by transitional generalized linear model (Table III). On other hand, one can notice larger spread of values off the diagonal for the matrix obtained by the transitional generalized linear model. This spread facilitates the probability of distant migration far from the diagonal (extreme transition), which is particularly important in long-run forecasts.

The results of LGD estimation are presented in Fig. 1 and Fig. 2. Both figures present time evolution of LGD in terms of probability of achieving a given LGD rating. Each picture shows probabilities of reaching one target rating, from A01 to A08, provided the first observed LGD value was assigned to relevant initial rating A01 to A08.

TABLE II

One-year migration matrix evaluated by gradient boosting (only terms > 0.01% are shown).

	A01	A02	A03	A04	A05	A06	A07	A08
A01	76.76%	23.23%						
A02	0.77%	86.82%	12.39%					
A03	0.23%	0.14%	83.39%	16.22%				
A04			0.74%	98.03%	1.22%			
A05		0.70%		1.44%	88.70%	8.69%	0.0045	
A06					6.18%	75.43%	18.38%	
A07					0.94%	4.88%	84.25%	9.91%
A08						0.16%	0.04%	99.49%

TABLE III

One-year migration matrix evaluated by transitional model (only terms > 0.01% are shown).

	A01	A02	A03	A04	A05	A06	A07	A08
A01	93.51%	6.38%	0.09%					
A02	10.26%	78.37%	11.17%	0.18%				
A03	0.13%	8.98%	77.72%	12.79%	0.35%			
A04		0.20%	12.61%	72.64%	14.24%	0.27%	0.01%	
A05			0.29%	11.07%	76.36%	11.55%	0.68%	0.02%
A06				0.39%	18.90%	62.42%	17.42%	0.83%
A07				0.02%	1.23%	18.57%	66.35%	13.80%
A08					0.03%	0.70%	14.33%	84.92%

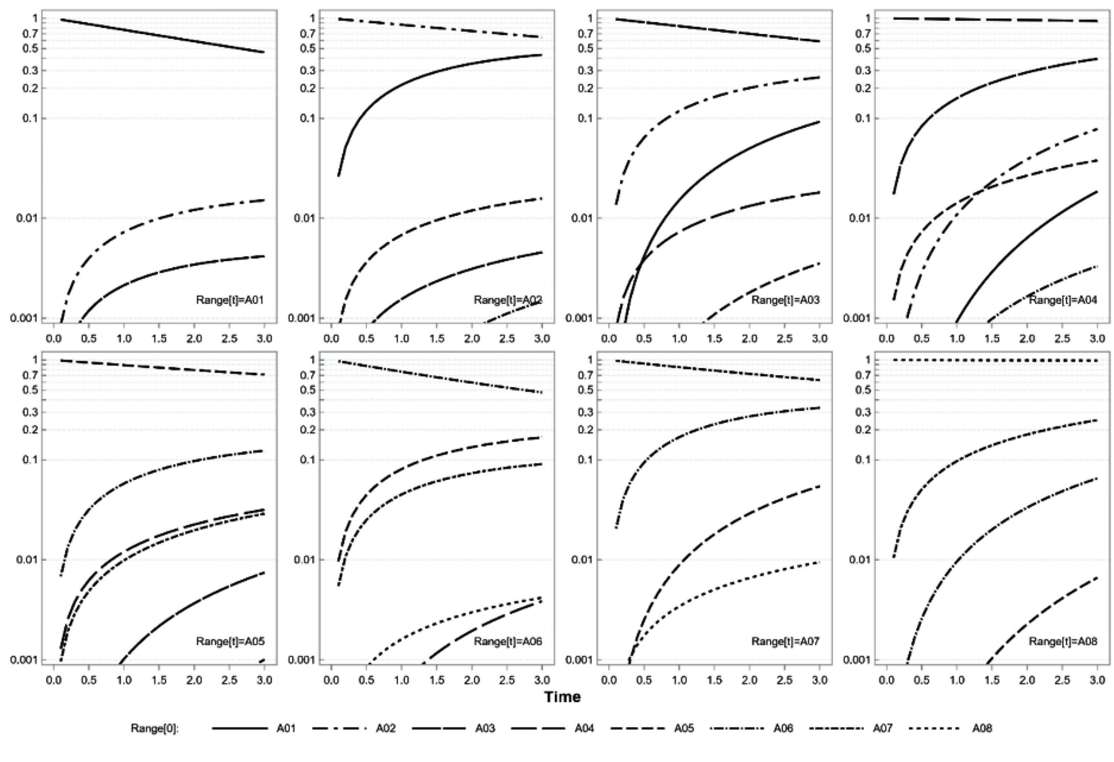

Fig. 1. Evaluated by the gradient boosting method probabilities that LGD values will reach a given rating when starting from any rating A01 to A08.

The dashed lines reflect initial LGD ratings. Both methods show that it is not possible to finally arrive at A01 rating if the initial rating is high (A06, A07, A08). It is also not possible to attain high rating A08, if the initial rating was low (A01, A02, A03). The probabilities and possibilities of migration between ratings differ depending on the method. The ROC curves presented in Fig. 3 and Fig. 4 show the performance of our methods as classification methods. The comparison of both

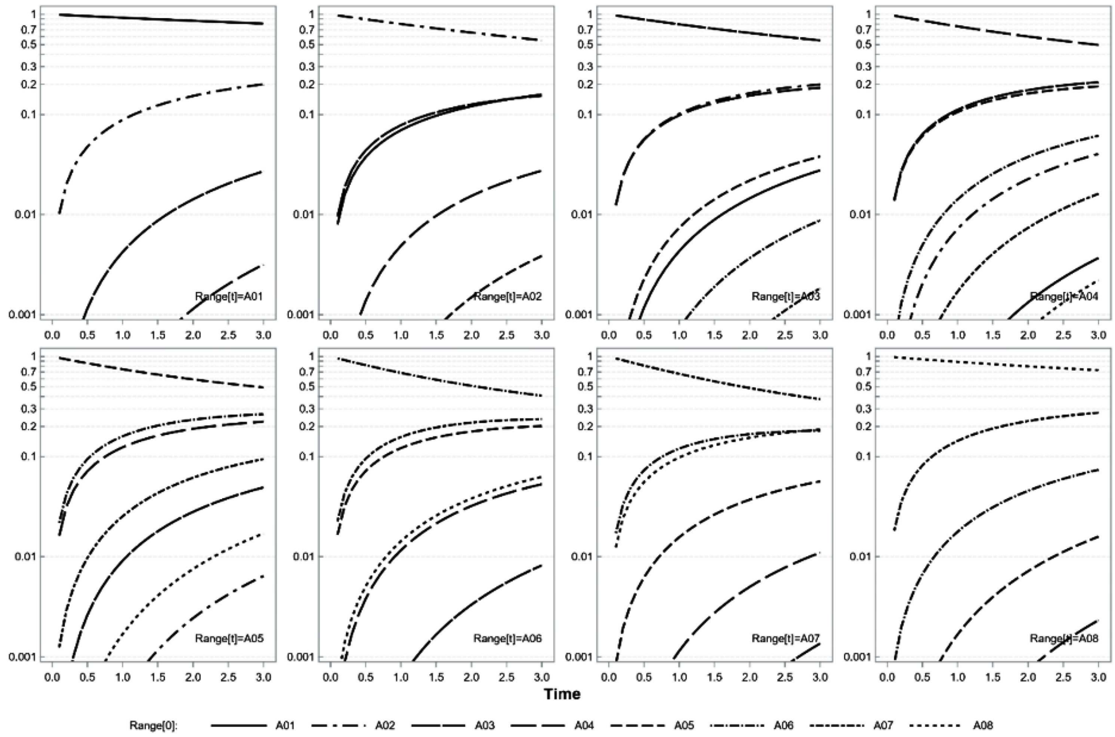

Fig. 2. Evaluated by the transitional model probabilities that LGD values will reach a given rating when starting from any rating A01 to A08.

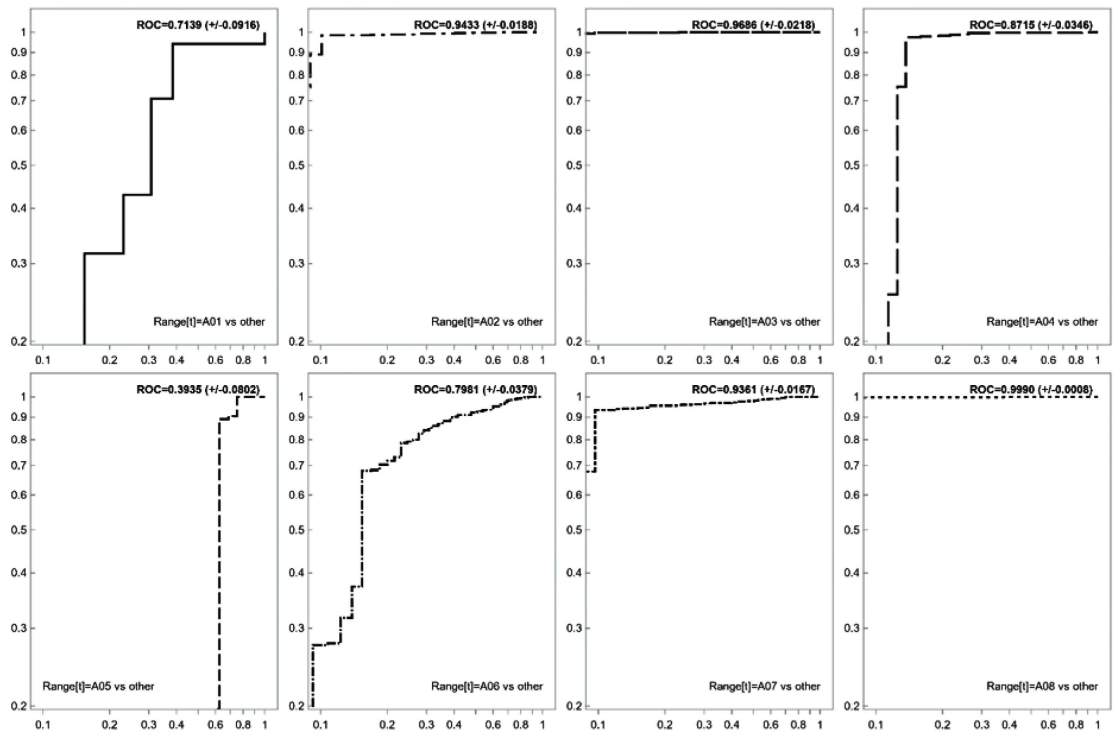

Fig. 3. ROC curves and AUC values for gradient boosting LGD classes. Each rating is tested against all remaining ratings.

figures reveals that gradient boosting performs much better than the transitional model. The values of AUC presented above each curve indicate outstanding classification to four ratings A02, A03, A07 and A08 for gradient boosting (Fig. 3). The classification to A01, A04 and A06 is also high. The performance of transitional model evaluated by ROC curves and AUC values is much worse (Fig. 4).

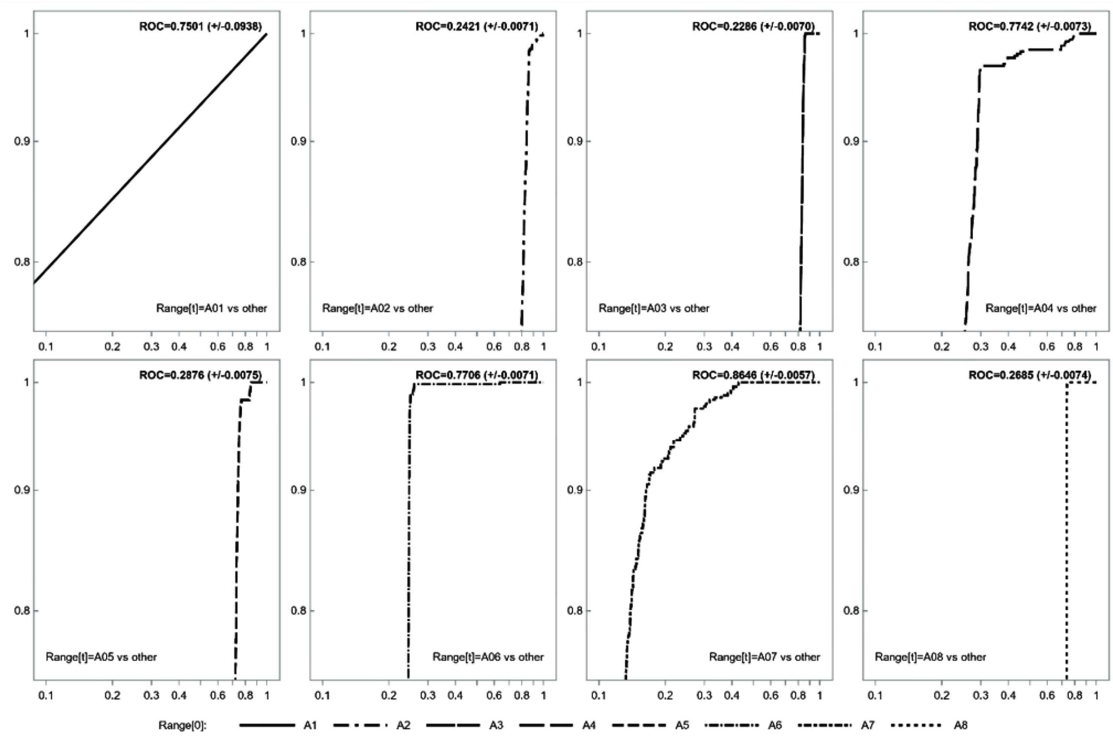

Fig. 4. ROC curves and AUC values for transitional model based LGD classes. Each rating tested against all remaining ratings.

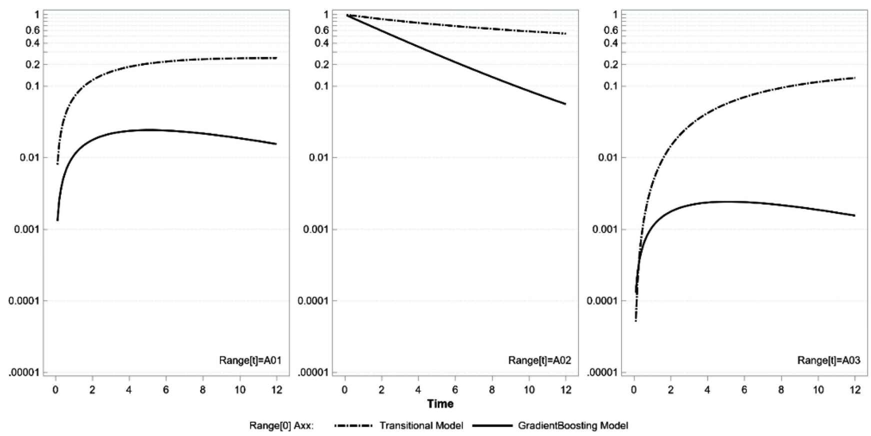

Fig. 5. Dynamics of the changes of migration from rating A02 (time = 0) to A01, A02 and A03 as a function of time.

The LGD estimates for a period longer than one year are shown in Fig. 5. The results show a systematical relationship between the change of rating and the rating level for transitional model. Gradient boosting estimates indicate the opposite behavior. Although most estimates for different periods indicated consistent results (the relationship between the change LGD and rating level), but in some cases, the result of non-monotonic dependence was obtained.

The machine learning method gives a transition matrix with a probability distribution centered around the diagonal. Projections of the transition probabilities based on infinitesimal generator built on this matrix give more diffuse results as shown in Fig. 1 and Fig. 2. It can be seen that the probability of staying in one rating decreases over time. Confirmation of this fact can be found in Figs. 1, 2 and Fig. 5 for A02 rating. Both distributions facilitate the probability of distant migration far from the diagonal (extreme transition), which is especially important in long-term forecasts.

Meanwhile, estimates using the regression method are more dispersed. It is difficult to assess the results because from the business side the most favorable is the trend of change towards the higher ratings. It can be concluded that the estimates based on transitional model give more weight to rating migration scenarios which are historically rarely observed.

5. Conclusions

The aim of our research was to show that machine learning methods can successfully be applied in estimation of lifetime LGD. In order to forecast LGD values we have constructed migration matrices of Markov chains using two approaches, a transitional model and a gradient boosting machine learning method. Both methods proposed within these approaches are new. The transitional model involves an estimation process using predefined ratings. However, the method has a rigorous and rigid assumption, therefore the rating changes are similar to the method based on frequency counts. Transitional models enable the inclusion of changes in the economic environment described by covariates, but the changes featured with these variables are minimized by the logistic loss function.

Estimates using machine learning method provide more efficient results than the transitional model. In addition, similarly as transitional model, this method also facilitates indirect estimation in a sequential manner. Constructed transition matrices are able to take into account dynamics of external factors (covariates) throughout the period but the loss function is more flexible because of dealing with continuous values of LGD.

The analysis of credit quality of enterprises is an important factor when assessing credit risk. It depends equally on the probability of default and the recovery rate in the event of default. Especially the latter parameter needs monitoring. To this purpose, rating change assessment models can be used that can highlight trends and thus improve forecasts.

The method proposed in the manuscript enables calculating LGD for arbitrary time in a sequential manner. Thanks to this it is possible to estimate the credit risk in accordance with IFRS 9 recommendations for the actual life time of the contract.

Estimates using the machine learning method give different results than that using the regression. In both cases, constructed transition matrices can take into account dynamics of external factors (covariates). The methods give results more resistant to changes in the economic environment, more similar to the TTC (Through the Cycle) methodology. There is a problem in finding another way of adequate incorporating time into the model. This can be done either by making the transition probabilities time dependent or by incorporating time directly into the rating. This is still an open problem for further research.

The effectiveness and accuracy of the methods used in credit management are critical. This study compares the results of statistical methods and machine learning on real data. Until now, the use of algorithmic methods was limited by the fact that the data were collected as repeated measures from the same sample at different points in time which led to the need for relationship analysis at several levels. Two cultures of analytical models: statistical based on probability distributions and algorithmic using elements of the cause-effect chain have their limitations. It seems that the use of an algorithmic approach will allow for more precise risk calculations.

Our research has shown that machine learning methods can be applied in life time LGD estimation. Similarly, as regression models, machine learning methods enable incorporating attributes into the model. Gradient boosting seems to perform better than transitional model. However, at this stage it is not possible to compare both considered approaches. One reason lies in the data set. The sequence of available observations is too short. The performance of methods should be examined also on other data sets.

References

[1] IFRS 9 — *Financial Instruments*.

[2] D. Beerbaum, S. Ahmad, "Credit Risk According to IFRS 9: Significant Increase in Credit Risk and Implications for Financial Institutions", *SSRN Electronic J.*, (2015).

[3] T. Bellini, *IFRS 9 and CECL Credit Risk Modelling and Validation*, Academic Press, 2019.

[4] G. Chawla, L.R. Forest Jr., S.D. Aguais, *J. Risk Manage. Financial Institutions* **9**, 3 (2016).

[5] SAS, *Base SAS(R) 9.4 Procedures Guide: High-Performance Procedures*.

[6] Y. Jafry, T. Schuermann, *J. Bank. Finance* **28**, 11 (2004).

[7] T. Schuermann, in: *Encyclopedia Quantitative Risk Analysis and Assessment*, 2008.

[8] J. Berglund, "Estimating Expected Lifetime of Revolving Credit Facilities in an IFRS 9 Framework".

[9] St. Trueck, S. Rachev, *Rating Based Modeling of Credit Risk*, Academic Press, 2009.

[10] A. Agresti, *Categorical Data Analysis*, John Wiley & Sons, 2002.

[11] J.H. Friedman, *Ann. Statist.* **29**, 5, (2001).

[12] T. Hastie, R. Tibshirani, J.H. Friedman, *The Elements of Statistical Learning*, Springer-Verlag, 2009.

[13] K.A. Spackman, in: *Proc. Sixth Int. Workshop on Machine Learning*, Ed. A.M. Segre, Elsevier, Cornell University, 1989, P. 160.

[14] T. Fawcett, *Pattern Recognit. Lett.* **27**, 861 (2006).