Original Scientific Paper

doi:10.5937/jaes14-11752

Paper number: 14(2016)4, 403, 465 - 476

VALIDATION OF LOSS GIVEN DEFAULT FOR CORPORATE

Miloš Vujnović*
Jubmes Banka, Serbia
Nebojša Nikolić
Jubmes Banka, Serbia
Anja Vujnović
Jubmes Banka, Serbia

This paper presents an contemporary approach for development and validation of Loss given default (LGD) in accordance with the Basel Accords standards. The modeling data set has been based on data on recoveries of outstanding debts from corporate entities in Republic of Serbia that defaulted. The aim of the paper is to develop a LGD model capable of confirming the validity of historically observed LGD estimates on the sample of corporate entities that defualted. The modelling approach in this research is based on average LGD without time or exposure weightening. The probability density function of realized empirical LGDs has been created by beta distribution usage. The validation process on proposed LGD model has been performed by throughout testing of: cumulative LGD accuracy ratio, mean square error calculation and regression analysis. On the basis of obtained results, the possibilities of application of the developed LGD model are proposed and discussed.

Key words: Loss Given Default; LGD; Model, Portfolio; Serbia

INTRODUCTION

In the context of credit risk modeling, the term “validation” includes the set of processes and activities that contribute to the standpoint that risk components adequately characterize relevant risk aspects, the risk components being the probability of default (PD), the loss given default (LGD) and the exposure at default (EAD). The validation framework includes all aspects of validation which are, in this context, defined by general principles of validation published by the Basel Committee on Banking Supervision [03].

LGD represents the credit risk parameter that plays an important role in contemporary banking risk management practice. It contributes as the key risk parameter in regulatory capital calculation according to IRB approach [4], as well as for banks’ internal risk management process. Primary reason for such incentive is the permission for the banks to use the real LGDs from experience instead of fixed regulatory LGDs. The aim of LGD estimate is to accurately and efficiently quantify the level of recovery risk inherited within a defaulted exposure. Contemporary risk management practice and regulation emphasizes and promotes the use of internal models for calculating credit risk parameters and capital calculation [04]. Basel II framework emphasizes three approaches to quantifying LGD: workout LGD, market LGD and implied market LGD. The visibility and attractiveness of LGD has also been recognized in new IFRS 9 standard. The new IFRS 9 standard extends the usage of LGD not only for calculation of risk weighted assets, as currently under Basel Capital Accord IRB approach, but for calculation of loan loss provision and allowances.

Clear definition of default is the prerequisite for the LGD estimation. Another basic prerequisite is the definition of LGD. Depending on definition of the time of default, LGD calculation may offer different results. If the model is to be used for capital calculation in accordance with Basel II standards, it is necessary to use the regulatory definition of default [04], according to which LGD is based on economic loss, where the bank must estimate LGD for each placement in such a way that it reflects the recession conditions, which is necessary in order to include all relevant risks. LGD estimates must be based on historical recovery rates and, where applicable, they should

**Jubmes Banka AD, Bulevar Zorana Đinđića 121, 11000 Belgrade, Serbia; milos.vujnovic@jubmes.rs*

be based on estimated market value of a collateral.

Basel II defines the validation as one of the requirements regarding LGD so that estimation of the same is acceptable for definition of the regulatory capital. Banks must have a sustainable system for validation of accuracy and consistency of the rating system, processes and all relevant components. Comparison between realized and estimated LGD must be performed regularly (at least once a year) in order to demonstrate that the realized LGD is within the expected value framework.

Although recent research led to advanced backtesting methods for PD models, the literature on similar backtesting methods for LGD models is much scarcer.

In this sense, the framework for backtesting of LGD model was offered by Loterman, Debruyne, Vanden Branden, Van Gestel, Mues [13]. Current LGD performance evaluation practices found in the literature have so far been usually limited to comparing internal LGD predictions and realized LGD observations using error-based metrics, correlation-based metrics or even classification-based metrics [12]. Most of the LGD studies focus on investigating the importance of various factors that affect LGD, for example, contract characteristics, borrower characteristics, industry conditions, and macroeconomic conditions. Very few studies of LGD explore the alternative modeling methodologies [01].

The research of Gurtler and Hibbeln [9] theoretically analyze problems arising when forecasting LGDs of bank loans that lead to inconsistent estimates and a low predictive power. The research present several improvements for LGD estimates, considering length-biased sampling, different loan characteristics depending on the type of default end, and different information sets according to the default status. It shows how the modeling data could be restricted in order to obtain unbiased LGD estimates.

The predictive power of any LGD model depends on proper choice (and availability) of the model input parameters obtained from obligor's information.

For a given input data set, the model calibration quality depends mainly on the proper choice (and availability) of explanatory variables (model factors), but not on the model used for fitting. Calibration of LGD models using distressed business cycle periods provide better fit than the data from total available time span [18].

The paper of Qi and Zhao [14] compared six modeling methods for LGD and found that non-parametric methods (regression tree and neural network) perform better than parametric methods both in and out of sample when over-fitting is properly controlled. The Farinelli and Shkolnikov [8] study pointed out that LGD follow beta distributions with means estimated from historical data. The shapes of the beta distributions vary across firms in such a way that the density function is concave if the corresponding credit instrument is backed by a collateral and convex otherwise.

The global financial crisis highlighted the fact that default and recovery rates of multiple borrowers generally deteriorate jointly during economic downturns. The vast majority of the literature, as well as many industry credit-portfolio risk models, ignore this and analyze default probabilities and recoveries in the event of default separately. The paper of Bade, Rösch and Scheule [05] is the first of its kind to assess the performance alternatives that incorporate the dependence between probabilities of default and recovery rates. In it four banks using different estimation procedures are compared.

Further researches of LGD are discussed and implemented by Antăo and Lacerda [02], Thomas, Matuszyk and Moore [17], Jankowitsch, Pullirsch and Vezˇa [11], Calabrese [06], Jacobs Jr. and Karagozoglu [10].

LOSS GIVEN DEFAULT ESTIMATION METHODOLOGY AND EXPERIMENTAL DESIGN

LGD estimation is the first step in the process of validation. However, LGD estimation may be a challenge mainly due to limited data availability. Basel II emphasizes the need of banks to develop and apply internal credit risk models and therefore quantitative models of LGD estimation represent the basis for application of IRB approach for corporate entities. The banks are required to enable LGD estimation based on the group of borrowers with similar characteristics. However, LGD estimation may be a challenge mainly due to limited data availability.

In view of the fact that LGD is one of the basic inputs of the credit risk model, the primary problem may occur in the case of small number of defaults.

Basel II framework published in relation to the validation principles [03] is considered to be sufficiently flexible so that even portfolios with a small number of default incidents are acceptable for application under the IRB systems. As in the case of other portfolios, they must fulfill minimum criteria established by the Basel framework that include requirements for a sensible, precise and consistent quantitative risk estimations. The choice of tools and techniques will considerably depend on the situation of the individual bank and the portfolio itself.

As financial institutions increasingly use economic capital to measure and manage their risk exposures and to optimize capital levels, issues relating to economic capital and loan asset pricing have become increasingly important. By varying the riskiness of assets and LGD levels it is possible however for institutions to target higher ratings without having to increase their capital levels. Indeed, there is even opportunity to decrease capital and subsequently loan price given a higher target rating for the lending institution. [16].

The process of determination of LGD estimation included activities of data collection, preliminary processing and analysis, then model building and estimation and, finally, model validation.

For necessary calculations in the validation process, certain concepts of validation in the form of measurement of the discrimination strength of the model and the adequacy of its calibration have been applied. Approaches that were found to best measure the calculation accuracy were chosen, on the basis of the validation process according to empirical results.

All necessary methods and models were applied on the relevant data obtained from a small bank's operations exposed to credit risk (measured by the balance amount and its share in the total balance amount of the banking sector) which predominantly performs its business activities with corporate clients (corporate entities) on the domestic market i.e. on the market of the Republic of Serbia.

Data collection and structure of dataset

In order to define the database which will serve as a source of data for the LGD calculation and validation, it was necessary to define the model objectives. In this sense, the model developed here is intended for use in validation of a LGD calculated for the purpose of its use as a basic credit risk parameter. As so, following implementation possibilities are considered: calculation of level of depreciation of exposure at a collective level in accordance with IAS requirements, determination of methodological approach for placement price calculation (interest or discount rate) based on adequate and accurate estimation of margin for undertaken risk, calculation of minimal capital requirements for credit risk by application of IRB approach, fulfilment of requirements of Tier II of the Basel II Capital Accord which means measurement of internal capital requirement for credit risk through ICAAP process by application of statistical methods and performance of stress testing of the bank's exposure to credit risk, as well as a quality measuring and management of profitability through calculation of return indices based on risk estimation and engaged bank's economic capital. In addition to the regulatory requirement, accurate predictions of LGDs are important for risk-based decision making, e.g. the risk-adjusted pricing of loans, economic capital calculations, and the pricing of asset-backed securities or credit derivatives [11].

Consequently, banks using LGD models with high predictive power can generate competitive advantages, whereas weak predictions can lead to adverse selection.

Data collection is the most demanding segment of the validation process. It was very important for the model integrity to make sure that the empirical data fulfill the requirements such as the representativeness of the segment for model application, the quantity (sufficient quantity to enable statistically important results), the quality (in order to avoid distortion as a result of unreliable data). In this context, it was necessary to perform preliminary analysis of the available database in order to enable insight into available data, deletion of double entries and identification of the nature of missing data.

Dataset representing the basis for creation of the possibility to perform adequate modelling had to fulfill the following conditions:

- obvious mistakes had to be removed;
- data on default and payments are available and reliable for all defaulted borrowers.

In the research, we used the possibility of acquisition of observations of different defaulted exposures for the purpose of their inclusion into the database using different starting points (default date) in time. This approach of time stratification is desirable as it decreases dependence of data on a particular calendar year i.e. the economic cycle present at the moment to which the data refer. This is especially important from the viewpoint of work with small databases, such is the available database, and with the aim of generating a quality, statistically important sample that may be used for the model creation.

The database was formed out of defaulted clients of a small bank operating on the market of the Republic of Serbia. The default status emerges if the firm in the subsequent year enters into material delinquency (more than 1% of exposure) on their obligation of more than 90 consecutive days past due. Such definition of default is compliant with [04]. Defaults are internal information of the bank and are recorded only for the bank's clients.

In view of the fact that the bank is small with a relatively limited database, the whole population was included in the analysis i.e. census research was performed in which all data were acquired for all elements of the population. Contrary to calculation of PD where one usually have large data sets, in LGD calculation data set is limited to only defaulted clients, which are usually scarce. In the research we used payment data for 217 exposures of 161 defaulted corporate entities in Serbia. Basic input dataset used for calculation of LGD and development of validation model is made of information about defaulted borrowers – corporate entities in the reference period of 5 years: 2009-2013. Only the corporate entities which have had material financial liabilities from credit-like products were taken into analysis. This length of data series satisfies Basel II compliance condition of minimum existence of 5 year of data history [04].

The most useful data for estimation of LGD come from the bank's own experience as LGD directly reflects the characteristics of the recovery process of an institution. Relevant data therefore mean the complete history of loss cases. The history of relevant loss data consists of the following:

- data on possible additional drawings after default;
- data on all recoveries related to exposure at default and risk mitigation instruments;
- data on all expenses coming out of the process of recovery of outstanding debts;
- all other information on the recovery process.

As the number of available data on losses was, as expected, small, cases that are not closed were also included in the analysis database. The decision to include such cases was made individually, on the basis of uncertainty of the final recovery due to incompleteness of the loss cases.

Analysis of availbale data represents the basis for adequate measurement of recovery risk. The analysis included the definition of adequate period and cases included in the calculation. This means that it should be determined on the basis of historical data when a case should be considered closed, although it is not formally closed. If it is recorded that, for example, five years after the date of entry in state of default and initiation of recovery process, any further flow of time increases the recovery rate only insignificantly, it can be concluded that such cases may be included in the analysis and the calculation of LGD.

Besides, the analysis of recovery time period was performed. The sample was divided into the segment of cases closed shortly after entry into default state (one year or less) and the segment of cases closing over a longer period of time. Cases closed over a shorter period of time, either through recognition of total loss (LGD=100%) or through full recovery (LGD = 0%), may be a result of technical default or fraud. As a matter of fact, such cases were discovered through expert analysis.

Period of recovery begins when the debtor enters status of default or when the debtor's case is transferred to debt recovery department. The period ends with the complete write-off of debt or the recovery and the return of debtor to active portfolio.

Availability of data is an important element of the analysis. The cases with missing data were excluded from the analysis in order to avoid poor quality results. Extreme results were detected by the expert estimation or the distribution analysis.

The structure of the dataset regarding the years from which the data used in research derive from is shown in the Table 1.

Table 1: Structure of the dataset by years

Year	Number	%
2009	44	20.28
2010	25	11.52
2011	15	6.91
2012	36	16.59
2013	97	44.70
Total	217	100.00

Data acquired and cleared represent the entire dataset.

Population included defaulted corporate entities for which complete information are available including receivable collection history after default.

Model building technique

Basel II requires conservative LGD estimates. LGD must be estimated in such a manner that it reflects the conditions of recession economy and cannot be lower than the long-run default-weighted average.

As in the case of rating philosophy, LGD philosophy defines expected behavior of LGD during the cycle. Under the PIT (point in time) system, LGD is a cyclical measure that describes expected LGD during the next 12 months. Contrary to this, according to TTC (through the cycle) philosophy, LGD is counter-cyclical and is defined as average LGD for a cycle during which it is relatively constant. Basel II assumes using of methodology similar to TTC philosophy in order to avoid the cyclicality in the dynamics of estimated capital. The majority of banks adopt counter-cyclical LGD philosophy i.e. the changes of LGD may be the result only of changes in the collateral characteristics and not the prediction of the LGD dynamics during the next year within the credit cycle.

There are four methods of calculation of LGD long-run default-weighted average on the portfolio level. Time-weighting is less desirable as it mitigates the impact of the years with high LGD rates at the cost of the low-default years, so that in practice weighting of all defaults is more frequently used. In the corporate segment, it is recommended to apply the calculation of averages based on the number of exposures that have come in default status, while exposure-weighting is more largely applied in the segment of individual customers.

Table 2: Long-term portfolio LGD average calculation methods

	Default count averaging	Exposure weighted averagin
Default weighted averaging	$LGD = \frac{\sum_{y=1}^{m}\sum_{i=1}^{n_y} LR_{i,y}}{\sum_{y=1}^{m} n_y}$ (1)	$LGD = \frac{\sum_{y=1}^{m}\sum_{i=1}^{n_y} EAD_{i,y}\times LR_{i,y}}{\sum_{y=1}^{m}\sum_{i=1}^{n_y} EAD_{i,y}}$ (2)
Time weighted averaging	$LGD = \frac{\sum_{i=1}^{n_y}\left(\frac{\sum_{i=1}^{n_y} LR_{i,y}}{n_y}\right)}{m}$ (3)	$\frac{\sum_{i=1}^{n_y}\left(\frac{\sum_{i=1}^{n_y} EAD_{i,y}\times LR_{i,y}}{\sum_{i=1}^{n_y} EAD_{i,y}}\right)}{m}$ (4)

where: i is a default observation, y is the year of default, n_y is the number of defaults in each year, m are years of observation, LR is the loss rate or LGD for each observation [07].

As the loss used for estimation of LGD for regulatory purposes represents an economic loss, the research took in consideration all relevant factors such are material effects of discounting, material direct and indirect expenses related to the process of recovery [04]. Direct external expenses included e.g. fees to the recovery agent, expenses of the collateral sale, expenses of the business operation. Indirect expenses represent the expenses borne by the recovery bank in the form of manpower and material spent on the part of the department authorized to recover problem debts. LGD estimation consisted of three steps. The first was collection of data, which included identification and collection of data necessary for generating the LGD estimation. The second step represented preliminary processing. This step included transformation of raw data into the form appropriate for estimation of LGD values. The last step was generating of LGD estimates by means of adequate collection of results of preliminary processing.

Starting from the fact that on the domestic market there are no market prices of receivables in the form of bank placements, the calculation based on the process of recovery was used for the needs of LGD calculation, with the focus on estimated collection transactions resulting from the process of recovery of bad exposures, with adequate discounting and estimation of exposure.

LGD is by application of this approach defined as one minus ratio of current value of recovery in relation to the book value at the time of default: where CF_{ti} represents net payments received

$$LGD = 1 - \frac{\sum_{i=1}^{n} \frac{CF_{t_i}}{(1-d)^{t_i}}}{Book\ value\ at\ the\ time\ of\ default} \quad (5)$$

through recovery at the time t_i after default and d represents the discount rate, while n is the total number of collection transactions.

The basic question regarding the use of collection transactions approach, for which there is no universal answer in theory or practice, is what discount rate should be used. Higher the discount rate, higher is the LGD.

Discount rate may vary depending on the source of repayment so that a number of discount rates may be appropriate upon estimation of the recovery rate for an individual company (contracted placement rate, creditor's capital price, non risk rate, etc.). Adequate discount rate may therefore be lower than the contracted rate that includes compensation to expected decrease of collection transactions related to contracted payments. The research used the discount rate which consisted of the cost of funds and risk premium which is determined based on debtor rating class from which the debtor defaulted.

A large number of parameters (and their components) define the sum of recovery and in the case of corporate entities they are: value of the company in default status, payments received from the process of bankruptcy (liquidation value of the company at the moment of realization, period of payments received from bankruptcy, bankruptcy expenses), value of collateral in default status (market or nominal value), recovery through sale of collateral (market value of the collateral at the time of realization, liquidation period, realization costs, price in the case of foreclosure sale), interest cost (cost of receivable refinancing before realization, interest loss due to write-off), cost of regulation of bad debt (administrative expenses, restructuring expenses and liquidation expenses).

Starting point of the recovery estimation is defining of the estimation basis, which is the value of collateral in the case of recovery from the same i.e. the liquidation value of debtor's property in the case of recovery from the bankruptcy process. In this sense, it is necessary to determine the value of collateral at any time of entering into default, especially in view of the fact that the same may be reduced by the lack of maintenance before default as a result of liquidity problem in the previous period.

Expenses related to the realization of the collateral (sales agent) or the bankruptcy (administrative receiver) represent another important elements of recovery.

In the case of corporate debtors, the following components of LGD parameters stand out:

- debtors: information on creditworthiness, collateral and transaction, basic information about debtor;
- security instruments: value, kind, collateral creditor;
- transactions: book value, assigned collateral, kind of product, interest rate, repayment structure.

There are two basic differences between validation of PD and LGD:

- PD values are tested against DR, defined for groups of debtors, while expected LGD values are tested against individually realized LGD;
- default or fulfilment of obligations is a discrete variable with two possible states, while LGD represents a continuous variable where realized LGD values may vary from 0% to 100%.

The results in literature show that models accounting for the correlation of default and recovery do indeed perform better than models ignoring it [05].

Mentioned differences dictate different approaches in LGD and PD testing. However, dimensions that have to be tested remain the same and they refer to estimation of discrimination strength of LGD system against risk, calibration of LGD ranking system against risk, realization of LGD ranking philosophy and homogeneity of LGD rating against risk. Estimation of discrimination strength of LGD system against risk is performed on the basis of cumulative LGD accuracy ratio (CLAR) which represents a modified accuracy ratio (AR) which is the measure of discrimination strength [15]. Calibration of LGD ranking system against risk meant various analyses on individual and aggregate level. Testing of homogeneity of LGD rating against risk related to validation on the portfolio segment level and performance of statistical significance tests.

So it was necessary that realized LGD should be compared with expected LGD one year before default in order to test accuracy of the ranking system. Reference point for LGD rating, one year before default is consistent with the period of risk of economic capital. Unlike observed default rate, realized LGDs are not known at the time of default. It usually takes a few years to realize recovery unless it is a practice of financial institution to sell outstanding debts soon after entering into default state. LGD testing is therefore usually performed with certain delay. Yet another basic challenge in LGD testing was the lack of relevant historic data. Annual number of defaults is limited in order to create statistical significance of backtesting results. So, aggregate instead of annual data were used in order to deal with this problem.

Limited data and large standard deviations related to “U” look of distribution of LGD (beta distribution) may create a considerable “disharmony” in performance of individual statistical tests. Grouping of LGD observations into predefined frames may help in reduction of this statistical “disharmony”. LGD frames must be used so that LGD ratings could be compared against realized LGDs in discrimination tests. The purpose of discrimination tests is to validate the correctness of ranking of exposure to LGD risk. If LGD rating may perform an effective discrimination as per LGD risk, it is expected that the majority of realized LGDs originate from highly predictable LGD frames. Cumulative LGD accuracy ratio serves as a measure of ability of such ranking.

First step was determination of the number of exposures to which LGD rating is attributed (predicted LGD) in each LGD frame. These LGD frames are called predicted LGD frames. Realized LGDs are then sorted in a decreasing order and grouped in such a manner that the number of LGDs in each frame is equal to the number of exposures to which LGD rating is attributed within each observed LGD frame. On the basis of distribution it was tested how many attributed LGD ratings originated from predicted LGD frame. This was also performed on the cumulative basis, each time including yet another frame with lower predicted and realized LGD. Mentioned counting was then performed for the worst two ratings, then for the worst three ratings, etc., thus creating CLAR curve.

Based on the performed activities, it was possible to calculate CLAR coefficient, which is similar to Gini coefficient. It was possible to draw CLAR curve through the results, similar to CAP curve. y axis of CLAR curve represents a cumulative percent of correctly attributed realized LGDs, while x axis shows a cumulative percent of observations in LGD frames. In the most favourable case of ranking, observations in predected and realized frames match and CLAR curve will be on the 45-degree line. CLAR coefficient will be equal to 2x surface below CLAR curve. CLAR coefficient is between 0 and 1, where 1 represents ideal discrimination strength. The higher the CLAR value, the higher the discrimination strength of LGD ranking system.

First step was the division of LGDs to discrete LGD ratings. Next step was segmenting of realized LGD on the basis of ordinal ranking. Realized LGDs were ranked from the most unfavorable (100%) to more favorable. After identification of frameworks of predicted and realized LGDs for each debtor in the sample, the discrimination test was performed.

This approach was applied further until all frameworks were involved.

In view of limited data availability and high volatility of historical LGD estimates, the results of these accuracy tests had to be carefully analyzed. Essentially, the results of any test or ratio must not be observed on standalone basis but in combination with the result of a comprehensive set of accuracy tests. At the level of individual exposure, a large number of standard tests were performed:

1. a diagram, with predicted LGDs on the horizontal axis and corresponding realized LGDs on the vertical axis, for visual testing of comprehensive relationship and emphasizing of bias in comprehensive LGD rating process;
2. mean square value calculated in the following manner:

$$MSE = \frac{\Sigma(realized\ LGD - predicted\ LGD)^2}{number} of\ observations$$

Regression analysis (e.g. R^2 regression as a measure of statistical strength of expression - square root of which is correlation between predicted and realized LGDs), through segments between predicted LGD (dependent variable) and realized LGD (independent variable) is performed and the results are tested.

Similar to PD testing, validation on the level of portfolio segments is useful for LGD testing in order to provide a deeper insight into LGD ranking. As the sample is small on the aggregate level, any division into segments may result in a loss of statistical strength which is larger than benefit that may arise out of such testing. Nevertheless, when it is possible to perform, the analysis on the segment level provides the indication of validity of homogeneity assumption of various LGD risk factors. Dimensions that should be considered are characteristics of exposure: security instruments, contract provisions, recovery priority and characteristics of debtors in the sense of their kind (companies, banks, states, etc.), industries, locations, etc.

As more and more information become available over time, a repeated estimation of LGD factors may be performed as a part of annual validation process. Validity of these factors must be confirmed by e.g. regression analysis. The analysis must be performed on aggregate data and not only on newly available data. The aim is to confirm that LGD factors continue to be justifiable as new data are entered into the database.

EMPIRICAL DATA RESULTS AND DISCUSSION

Staring point of the analysis is LGD calculation based on reference set of empirical data. The calculation was based on the entire portfolio i.e. there was no segmentation according to any particular criteria and taking in account the limited size of available data set. Although the period shorter than 7 years recommended by Basel II standards was used for LGD calculation, it can be considered as relevant and conservative because the period includes both economic expansion and recession. So it can be regarded that LGD calculation follows the TTC philosophy and is an acyclic measure, being determined on the basis of the time period that includes the recession, where determined LGD is not corrected for any expectations in the course of the next year. Approach based on recovery of outstanding debts was used for LGD calculation as the loan market values (classic placements on domestic market) are not available i.e. there are no market data, which limits the application of other two possible approaches (market and implicit market). LGD of individual outstanding debt is calculated on the basis of the recovery money transfers and estimated expenses of monetary resources and familiar external expenses. This is reflected through the process of discounting of all recorded recovery-based payments received in the period from debtor's entering default status or any subsequent payment due dates, the discounting rate being the one that includes the expenses of monetary resources for financing of such outstanding debts and the risk premium for the corresponding placements of the rating class from which the debtor entered default status.

The data necessary for LGD calculation included, besides basic information about debtor, the following categories of information: unique identifier of placement, date of entry into default status, date of exit from default status (if existing), value date of debt denomination, rating class from which debtor entered default status, risk premium for the corresponding risk class, price of monetary resources, recovery expenses, collateral data.

For LGD calculation on the level of the entire portfolio, an approach based on LGD rate average per individual outstanding debt without time or EAD-weighted averages was used. In order to avoid averaging of years with high LGD rates by data from the years with low LGD rates, no time-weighted averages were used, but the entire sample was used for LGD calculation based on the number of exposures that have entered the default status. EAD weighting was not performed in order to avoid distortion of calculated LGD through adding of large weight to larger amounts of outstanding debts, starting from the premise that the bank's approach is the same for each debt, regardless of its size. In the case of the sample used, this would lead to overestimation of LGD. In LGD calculation, no division was made according to the kind of outstanding debt i.e. the quality of security instruments or recovery priority. The reason for such approach lies in the fact that the recovery rate of problem debts in the domestic market is not remarkably conditioned by the existence of a collateral. This was also confirmed through the population analysis for LGD calculation of LGD that recorded that average LGD of secured outstanding debts (mortgage, pledged assets, etc.) is 58.20% and that of non-collateralized is 55.47%, i.e. even lower than that of the collateralized. By application of beta distribution of LGD rates, probability density function of empirical LGDs was created, which has expected “U” form (Figure 1.).

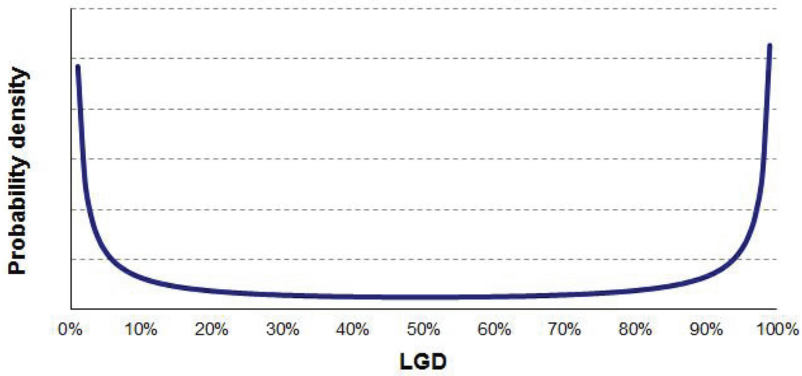

Figure 1: Probability density function of LGD

From the probability density function diagram it can be seen that grouping of determined percent is significant around 0% and 100%, which determines its “U” shape i.e. in the majority of default cases were either fully recoverable or totally unrecoverable. This is partially the consequence of definition of default i.e. the fact that the subject of modelling was every entry in default status - primarily the payment delay of 90 or more days (technical default), which did not necessarily result in an economic default. However, the use of this definition was necessary as the aim was to develop a model and demonstrate such approaches to validation that the model may be used for the needs of calculation of the capital requirement in accordance with IRB approach in whose root is the applied definition of default.

LGD validation is carried out through testing of CLAR, mean square error calculation and regression analysis of data from the LGD calculation sample.

CLAR serves as a measure of ranking ability against LGD risk. For the needs of CLAR calculation, LGD framework values were determined that define risk (five framework values, where 1 denotes the lowest risk and 5 denotes the highest risk). Also, the testing of adequacy of LGD prediction was performed based on estimated LGD one year before entry in default status of a particular placement and recorded empirical LGD of the placement. Table 3 shows LGD framework values according to predicted and realized LGDs.

Table 3: Value buckets (rating) of predicted and realized LGDs

Rating	Predicted LGD		Realized LGD		Number of observations
	Min LGD (%)	Max LGD (%)	Min LGD (%)	Max LGD (%)	
1	0.00	12.65	0.00	0.05	23
2	12.65	24.51	0.05	1.41	22
3	24.51	29.78	1.41	32.95	44
4	29.78	29.95	32.95	98.03	36
5	29.95	100.00	98.03	100.00	92

Value buckets are determined on the basis of analysis of LGD data. It was not possible to take completely equal number of observations as per framework, which is the consequence of the fact that no large number of individual observations is available (which is usually the case with LGD, especially in view of the fact that the source is a small bank) and having in mind significant grouping of LGD around 0% and 100% (detected also by "U" shape of LGD probability density curve). However, such division does not endanger the analysis as the distribution of the number of observations according to predicted and realized LGDs is the same. After distribution of LGD data according to the amount of the same, cumulative distribution of predicted LGD observations and cumulative distribution of correctly attributed LGD risk rating was determined. Correct prediction was considered to be an individual observation distributed in the same LGD rating according to predicted and realized LGD. Table 4. shows determined points of CLAR curve.

The area under CLAR curve is 0.4611 (Figure 2.), on the basis of which, by application of the rule of CLAR coefficient calculation the value of the same is obtained at 0.9221 (2×0.4611).

Table 4: Points of CLAR curve

Rating	Cumulative - predicted LGD observations (%)	Cumulative - correctly assigned realized LGD (%)
1	100.00	100.00
2	89.40	99.30
3	79.26	95.80
4	58.99	88.81
5	42.40	64.34

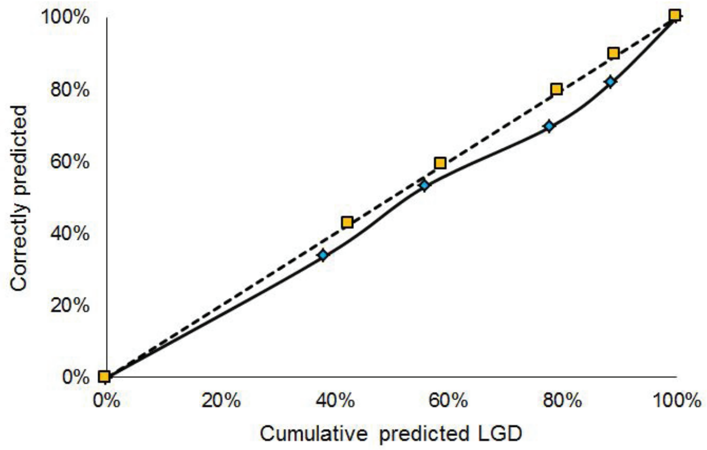

Figure 2: CLAR curve

As the value is very near to 1, it can be concluded that proposed ranking is well-differentiated against the LGD risk level.

Determined mean square error of the sample represents one of the approaches of the model calibration analysis which is used for LGD calculation and is 0.003256. As the mean square error is statistically important being less than 0.05 (predicted 95% reliability level) it can be concluded that average mean square deviation is not significant.

As an approach by which calibration of LGD model is analyzed, regression analysis of data from LGD calculation sample was performed. Realized LGD was determined as an independent variable, while predicted LGD was determined as a dependent variable. Table 5. contain relevant data on performed regression analysis.

Starting point of the regression analysis was the analysis of the residual i.e. of the figure representing deviation of empirical observations and results obtained through the model. Figure 3. shows points of empirical data and the regression line and it can be observed on the same that points are not significantly grouped around the regression line i.e. that differences between the analyzed data and the values predicted through the model are considerable. This was also confirmed by the calculation of determination coefficient (R^2) hich is very low.

Table 5: Regression analysis

Multiple R	R^2	Adjusted R^2	Standard error	Number of observations
0.214571451	0.046040907	0.041603888	0.077478107	217

	df	SS	MS	F	F significance
Regression	1	0.06228889	0.06228889	10.37654043	0.001474295
Residual value	215	1.290614284	0.006002857	-	-
Total	216	1.352903174	-	-	-

	Inception (α)	LGD
Coefficients	0.263126955	0.037246106
Standard error	0.008339015	0.011562575
t statistics	31.55372039	3.221263794
p-value	1.23131E-82	0.001474295
Lower boundary 95%	0.246690264	0.014455586
Upper boundary 95%	0.279564	0.060037

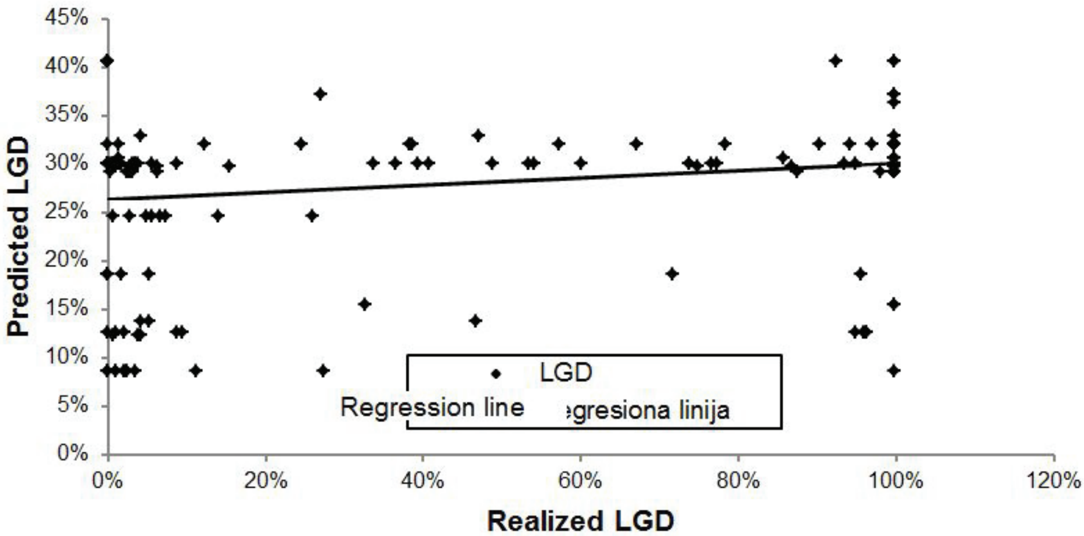

Figure 3: Regression line

The conclusion is that prediction of LGD based exclusively on realized LGDs contains a big amount of unexplained variability. This, however, does not imply that LGD calculation is not correct but indicates the fact that prediction of LGD cannot be based exclusively on dynamism of LGD in the previous period and that it is necessary to include other parameters in the LGD prediction model (e.g. macroeconomic variables, belonging to a particular economic sector, etc.) in order to increase the ability of explanation of the changes in predicted LGD (a larger sum of quadrate deviations explained by regression i.e. a bigger R^2). Nonetheless, the model does have coefficients that are statistically important (p-value smaller than 0.05, 95% reliability level) so that the change of value of the realized LGD will have the impact on the predicted LGD.

CONCLUSION

In our research we have built LGD model capable of validating LGD estimations. The model has been developed on dataset which comprised of five years of defaulted exposures to corporate entities in Republic of Serbia.

The aim of the research was to design a consistent and complete validation framework in order to confirm the soundness of the obtained results. During the LGD calculation process, as well as validation process various limitations and peculiarities have emerged, primarily with respect to data availability necessary for application of statistical analyses. All problems were successfully overcome and the final results produced the statistically profound model.

The quality of obtained results and the fact that the developed LGD validation model is based on actual data which have been proved to be statistically significant and sound. According to results, we can conclude that the developed and proposed model can be implemented and employed within a bank that operates in Serbia or in the region of South Eastern Europe.

On the basis of the conducted research in this paper, with application of appropriate techniques widely accepted in academic and industry practice the empirically appropriate process of validation of loss given default is proposed and implemented. The quality of the developed model is underlined by the fact that real data from the available database were used. Moreover fully replicable through application of methods and described approaches were employed during the model development.

REFERENCES

1) Altman, E., Resti, A., Sironi, A., (2005). Default recovery rates in credit risk modeling: A review of the literature and recent evidence. Journal of Finance Literature 1, 21-45
2) Antão, P., Lacerda, A. (2011). Capital requirements under the credit risk-based framework. Journal of Banking and Finance. 35, 1380-1390
3) Studies on validation of internal rating systems. (2005). Basel Committee on Banking Supervision Publications-Bank for International Settlement
4) International convergence of capital measurement and capital standards - A revised framework. (2006). Basel Committee on Banking Supervision Publications-Bank for International Settlement
5) Bade, B., Rösch, D., Scheule, H. (2011). Empirical performance of loss given default prediction models. The Journal of Risk Model Validation, 5 (2), 25-44,
6) Calabrese, R. (2012). Modelling Downturn Loss Given Default. UCD Geary Institute Discussion Paper Series, 7 (1),
7) Chalupka, R., Kopescni, J. (2008). Modelling Bank Loan LGD of Corporate and SME Segments: A Case Study. IES Working Paper, 27
8) Farinelli S., Shkolnikov M. (2012). Two Models of Stochastic Loss Given Default. arXiv.org
9) Gurtler, M., Hibbeln, M. (2013). Improvements in loss given default forecasts for bank loans. Journal of Banking and Finance, 37, 2354-2366
10) Jacobs Jr, M., Karagozoglu, K.M. (2011). Modelling Ultimate Loss Given Default on Corporate Debt. The Journal of Fixed income, 21 (1), 6-20
11) Jankowitsch, R., Pullirsch, R., Vezˇa, T. (2008). The delivery option in credit default swaps. Journal of Banking and Finance, 32, 1269–1285
12) Loterman, G., Brown, I., Martens, D., Mues, C., and Baesens, B. (2012). Benchmarking regression algorithms for loss given default modeling. International Journal of Forecasting, 28 (1), 161–170
13) Loterman, G., Debruyne, M., Vanden Branden, K., Van Gestel, T., Mues, C, (2014). A Proposed Framework for Backtesting Loss Given Default Models. The Journal of Risk Model Validation, 8.1: 69-90
14) Qi, M., Zhao, X. (2011). Comparison of modeling methods for Loss Given Default. Journal of Banking and Finance, 35, 2842-2855
15) Siddiqi, N. (2012). Credit risk scorecards: developing and implementing intelligent credit scoring. 3. John Wiley & Sons
16) Sundmacher, M., Ellis, C. (2011). Bank 'ratings arbitrage': Is LGD a blind spot in economic capital calculations?. International Review of Financial Analysis, 20, 6-11
17) Thomas, L.C., Matuszyk, A., Moore, A. (2012). Comparing debt characteristics and LGD models for different collections policies. International Journal of Forecasting, 28, 196-203
18) Yashkir, O., Yashkir, Y. (2013). Loss Given Default Modelling: Comparative Analysis. Journal of Risk Model Validation, 7 (1)

Paper sent to revision: 09.08.2016.

Paper ready for publication: 05.12.2016.