Loss Given Default Estimations in Emerging Capital Markets

Mikhail Pomazanov

Abstract This paper proposes an approach to decompose the RR/LGD model development process with two stages, specifically, for the RR/LGD rating model, and to calibrate the model using a linear form that minimizes residual risk. The residual risk in the recovery of defaulted debts is determined by the high uncertainty of the recovery level according to its average expected level. Such residual risk should be considered in the capital requirements for unexpected losses in the loan portfolio. This paper considers a simple residual risk model defined by one parameter. By developing an optimal RR/LGD model, it is proposed to use a residual risk metric. This metric gives the final formula for calibrating the LGD model, which is proposed for the linear model. Residual risk parameters are calculated for RR/LGD models for several open data sources for developed and developing markets. An implied method for updating the RR/LGD model is constructed with a correction for incomplete recovery through the recovery curve, which is built on the training sets. Based on the recovery curve, a recovery indicator is proposed which is useful for monitoring and collecting payments. The given recommendations are important for validating the parameters of RR/LGD model.

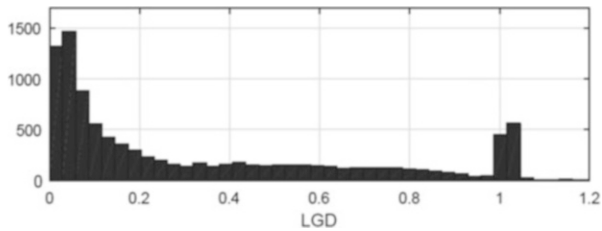
Keywords Credit risk · Residual risk · IFRS 9 standards · Unexpected losses · Loss given default · Recovery rate · Recovery curve · Capital requirements

JEL C58 · G17 · G28 · G32

M. Pomazanov (✉)
National Research University Higher School of Economics, Promsvyazbank, Russia

A. M. Karminsky et al. (eds.), *Risk Assessment and Financial Regulation in Emerging Markets' Banking*, Advanced Studies in Emerging Markets Finance,
https://doi.org/10.1007/978-3-030-69748-8_6

Fig. 1 Typical frequency distribution of the level of losses after LGD model



1 Introduction

LGD—Loss given default is one of the most important credit risk assessment parameters. Along with PD—Probability of default and EAD—Exposure at default, LGD contributes as a key parameter in calculating regulatory requirements, as well as economic capital requirements, as part of an approach based on internal IRB ratings (International Convergence 2006). The purpose of the LGD assessment is to accurately and efficiently quantify the level of recovery risk inherited as part of the default risk. The incentive to build LGD valuation models is the possibility of obtaining permission from the regulator to use the bank's approach based on internal ratings to calculate reserves and requirements for economic capital. The inverse of LGD is the RR (Recovery Rate), RR = 1 – LGD, so the RR simulation is identical to LGD. Recovery from default RR or its inverse value LGD = 1 – RR in practice demonstrates random dynamics and has a typical frequency profile, shown in Fig. 1. Many empirical studies have noted bimodality with a higher concentration of observations at zero and close to one and a higher LGD during periods of economic recession. This is evidenced by the results of a number of empirical works on mortgage lending (Araten et al. 2004; Karminsky et al. 2016) and corporate lending, including corporate bond market (Qi and Zhao 2011; Dermine and de Carvalho 2006; Schuermann 2004; Felsovalyi and Hurt 1998). Therefore, to calculate unexpected losses, it is necessary to take into account the volatility of LGD in addition to its expected estimate. The dispersion of LGD, reinforced by bimodality of distribution, contributes to unexpected losses, which are the basic component of residual credit risk.[1]

The typical model of LGD dispersion is not difficult to determine with the commonly used relation (Gordy and Lutkebohmert 2013):

[1]According to the definition given, for example, by the Bank of Russia (see Bank of Russia Ordinance No. 3624-U, dated April 15, 2015, "On Requirements for the Risk and Capital Management System of a Credit Organization and Banking Group"), residual risk is the risk remaining after the Bank's actions to reduce inherent risk. Suppose a bank takes measures (that is, requires collateral) to recover debt after default, based on which it statistically fairly expects a recovery share of RR = 1-LGD. And, let's say, on a statistically significant portfolio, this share of recovery will take place. However, due to the dispersion of LGD and the granularity of the default part of the portfolio, deviations from the expected value will be observed, including towards losses. This gives unexpected losses related to residual risk.

$$D(\mathrm{LGD}_i) = \gamma \cdot E(\mathrm{LGD}_i) \cdot (1 - E(\mathrm{LGD}_i)), \tag{1}$$

where $D(\cdot)$ is the variance (squared standard deviation), $E(\cdot)$ is the mathematical expectation, $i = 1\ldots N$ is the index of a model-homogeneous population for LGD,[2] γ is a RR/LGD dispersion parameter theoretically belonging to the interval of [0,1], its mean value $\gamma = 0.25$ is proposed, for example, in the CreditMetrics approach (CreditMetrics 1997). Assuming that, within the framework of the TAC, the LGD model corresponds to the average statistical observations of reconstructions, i.e. relatively medium, it does not overestimate or underestimate the calculations, we put $E(\mathrm{LGD}_i) = \mathrm{LGD}_i$. In practice, the parameter γ can be statistically refined at the stage of validation of the internal LGD model, for example, by the formula:

$$\gamma = \frac{\sum\limits_{d \in D} \left(\widehat{\mathrm{LGD}}_d - \mathrm{LGD}_d\right)^2}{\sum\limits_{d \in D} \mathrm{LGD}_d \cdot (1 - \mathrm{LGD}_d)}, \tag{2}$$

where LGD_d is the model estimate of the one default to the LGD before default, $\widehat{\mathrm{LGD}}_\mathrm{d}$ is the observed loss after the completion of the default debt recovery process.

The study (Antonova 2012) presents the result of the LGD assessment of Russian default issuers according to the information-analytical agency Cbonds. During the observation period from December 31, 2002 to December 31, 2011, 124 Russian corporate issuers made a real default on ruble corporate bonds that were traded on the MICEX. A real default is understood as failure to fulfill an obligation by the issuer before the expiration of the grace period. Based on the calculation method chosen by the author, RR: RR = 1-LGD were calculated for defaults of corporate bonds issued by Russian issuers in 59 cases, which formed a statistical sample. The overall outcome of the assessment was the average rate RR = 48.8% (LGD = 51.2%) with a standard deviation of σRR = σLGD = 29.2%. For the case of an LGD-insensitive assessment model, formula (2) takes a simple form:

$$\gamma = \frac{n-1}{n} \cdot \frac{\sigma \mathrm{RR}^2}{(1 - \mathrm{RR}) \cdot \mathrm{RR}} = 0.34. \tag{3}$$

The numerical estimate of γ is based on the result of the evaluation of LGD model as the average LGD, without constructing a refinement model. This estimate given by issuers can be considered a conservative estimation of uncertainty parameter γ of the LGD for the Russian bond market. It is useful to estimate the statistical error of

[2] A model-homogeneous population should be understood, for example, such industry segments of borrowers as "Banks", "Individuals, consumer loans", "Mass segment of small business", "Large corporate business" including credited to a particular bank, etc. It is reasonable to classify LGD segments of credit assets by business model or financial instrument. For each segment, various parameters γ are possible.

Table 1 Parameters γ for various industry segments of the default bonds of Russia

Industry	Average, in %	Standard deviation, in %	Number of observations	γ
Light Industry	19.4	10	4	0.05
Heavy Industry	63.3	25	11	0.24
Trade	48.5	29	15	0.31
Construction	57.2	27	6	0.25
Agriculture and food processing	50.6	30	18	0.34
Other services	24.4	28.2	5	0.34
Total	48.8	29.2	59	0.34

Table 2 Parameters γ for various industry segments of US default bonds

Industry	Average recovery, in %	Standard deviation, in %	Number of observations	γ
Real estate	41.97	16.05	71	0.10
Transportation	38.17	18.85	70	0.15
Electricity	48.03	22.67	39	0.20
Oil&Gas	44.37	23.68	21	0.22
Manufacturing	38.93	28.55	573	0.34
Service&Leisure	38.65	30.37	190	0.39
Retail	33.4	34.19	33	0.51
Media&Communications	34.7	34.56	163	0.52
Total	38.68	28.22	1160	0.34

the parameter γ, since, when developing the LGD model, statistics are often not enough. The estimate of $\sigma\gamma$ is as follows:

$$\frac{\sigma\gamma}{\gamma} \cong \frac{1}{\sqrt{n}}\left(\sqrt{2} + \sigma\mathrm{LGD}\frac{|2\mathrm{LGD} - 1|}{\mathrm{LGD}(1 - \mathrm{LGD})}\right). \tag{4}$$

Formula (4) gives the standard deviation of the statistical error γ, provided that the model LGD is equal to the average. The statistical error (estimation of the standard deviation of the error) for the above sample of 59 issuers was $\sigma\gamma = 0.06$.

The study of (Antonova 2012) indicators of average RR and standard deviations for several industry segments was also evaluated separately. The results of the evaluation of individual parameters γ are presented in Table 1.

The work of (Jankowitscha et al. 2014) presents the calculation of recovery levels for defaulted US bonds for the period July 2002 to October 2010, as well as standard deviations. A similar calculation of γ for non-financial sector companies is shown in Table 2 by industry and in general.

Figure 2 shows the ranges of γ taking into account standard deviations due to statistical error. It can be seen from Fig. 2 that, taking into account the statistical error

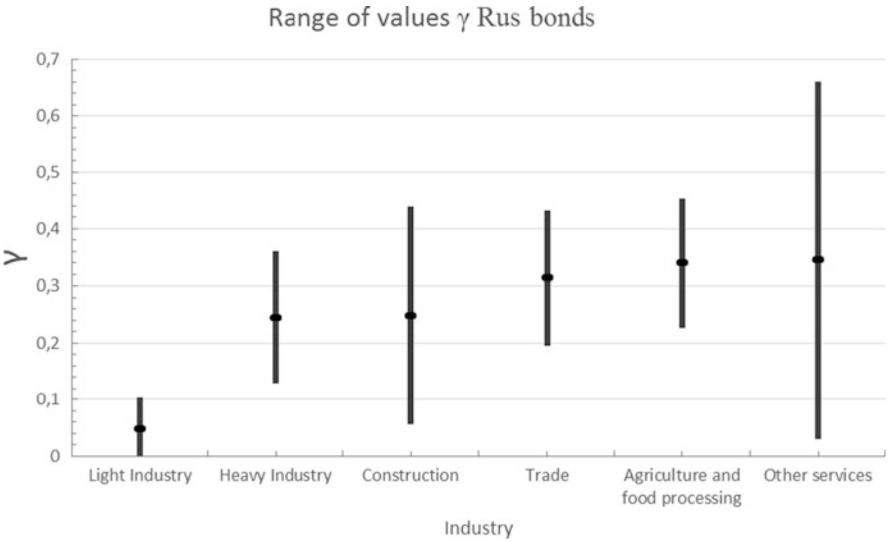


Fig. 2 Ranges γ for different industry segments of Russia, taking into account standard deviations due to statistical error

for different industry segments, the ranges of possible values of γ substantially intersect.

An exception is only for the light industry. But in this segment there are very few measurements and, perhaps, this is just an extreme result, which is usually discarded in statistical measurements (see Fig. 2). Comparing the results of recoveries of default bonds of the US and Russia obtained at the same observation periods, it is obvious that the average recovery level in the US was 10% lower than the Russian ones, however, the average volatility parameter γ practically coincided with the Russian one at the level $\gamma = 0.34$.

Figure 3 shows the ranges of γ according to the standard deviations due to statistical error.

However, a clear stratification of the values of γ by industry segments is revealed, in particular, the real estate differs in the minimum level of the volatility parameter, $\gamma = 0.1$, the sectors Retail and Media & Communications, $\gamma = 0.5$, have the maximum. The inclusion of statistical error, obviously, rejects the hypothesis of independence of γ, in particular, from the industry segment.

Therefore, it makes sense when building the LGD model to a model for the volatility parameter γ, too. With a lack of observations, it is possible to assume that $\gamma =$ const for all measurements within a model-homogeneous population, but this will fix the model error.

In the next part of the work, it is necessary to answer these questions: how to take into account the results of recoveries of default borrowers, if the provided the recovery process is incomplete? How to use statistically implemented recovery dynamics to build recovery indices for early defaults? What functionality should

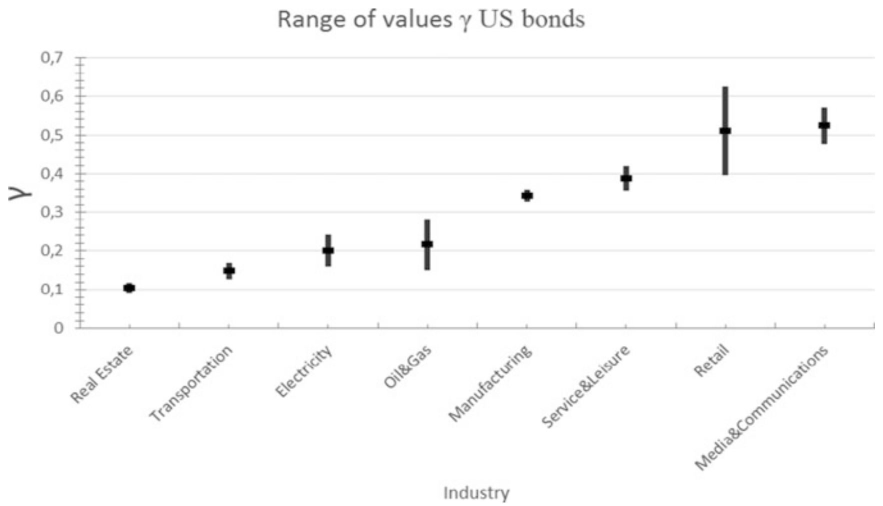


Fig. 3 Ranges γ for different US industry segments, taking into account standard deviations due to statistical error

Table 3 LGD assessment method

	Default count averaging	Exposure weighted averaging
Default weighted averaging	$\mathrm{LGD} = \frac{\sum_{y=1}^{m}\sum_{i=1}^{n_y} LR_{i,y}}{\sum_{y=1}^{m} n_y}$ (5)	$\mathrm{LGD} = \frac{\sum_{y=1}^{m}\sum_{i=1}^{n} EAD_{i,y}*LR_{i,y}}{\sum_{y=1}^{m}\sum_{i=1}^{n_y} EAD_{i,y}}$ (6)
Time weighted averaging	$\mathrm{LGD} = \frac{\sum_{i=1}^{n_y}\left(\frac{\sum_{i=1}^{n_y} LR_{i,y}}{n_y}\right)}{m}$ (7)	$\mathrm{LGD} = \frac{\sum_{i=1}^{n_y}\left(\frac{\sum_{i=1}^{n_y} EAD_{i,y}*LR_{i,y}}{\sum_{i=1}^{n_y} EAD_{i,y}}\right)}{m}$ (8)

be optimized to build an LGD model while minimizing residual risk? How does residual risk affect economic capital requirements? What is the model? A simple, but optimal from the point of view of residual risk, LGD model will be proposed, based on a positively discriminatory rating of LGD.

2 Recovery Curve

The start of identifying the types of RR (LGD) that can be considered as measures of LGD. In the extensive literature on LGD, for example (Vujnović et al. 2016), four are represented (Table 3).

Where i is the observation of default, y is the year of default, n_y is the number of defaults in each year, m is the years of observation, LR is the loss coefficient or LGD for each observation.

For practical purposes, it suffices to contrast on two approaches for calculating RR.

A. Simple recovery index (medium/median or frequency):

$$\mathrm{RR}_{\mathrm{avg}} = \frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{E_i}, \tag{9}$$

where R_i is the amount of funds received to repay the debt of borrower i, discounted to the default date (both direct and indirect recovery are taken into account), E_i is the exposure to default (EAD) of borrower i. EAD—the amount of the main debt, accrued interest, fines, and other charges to the reporting period before default. After the moment of default, fines, interest, and other accruals after default are not included in the EAD exposure, the off-balance part is not included, but the amounts issued after default are included. The net credit exposure is the adjusted (reduced) credit exposure for the amount of the discounted financial collateral. The simple recovery index (RR) is not oriented to amounts; it shows the average share of recovery among defaulting borrowers.

B. Weighted Average Recovery Index

$$RRw = \frac{\sum R_i}{\sum E_i}. \tag{10}$$

The weighted index is sensitive to the defaulted amounts (to losses). Thus, the indicators RRavg and RRw will differ if the share of recovery depends on the amount in default. If large loans recover heavier than small ones, then a simple recovery index exceeds a weighted one and vice versa. The recovery amount is calculated based on recovery payments discounted to the default date.

$$R = \sum_{t=0}^{\infty}\frac{P_t - C_t}{(1+q)^t}, \tag{11}$$

where P_t—recovery payments at time t from the date of default, C_t—costs of bank recovery costs $\frac{1}{(1+q)^t}$ —discount factor with the rate q, the sign "∞" means that theoretically wait for the a completed collection can indefinite (in practice, of course, the wait is limited and will be seen later). The repayment history for the sample of default loans (at least $\widehat{\tau}$) is presented in Table 4. The sample is taken for a sufficiently wide period of "observing" $\widehat{\tau}$ >3–5 years. Those. on the interval of $[t - \widehat{\tau}, t]$, where t is the current moment of observation of defaults (reporting date—90 days). The list of repayment history parameters:

1. ID (number) of the borrower;
2. Exposure in default (EAD, taking into account possible loans issued after default,

Table 4 Parameters of repayment history

ID	EAD	Discount rate, q in %	Default date	Recovery period after default (year)						
				1	2	3	...	S	...	P
1	E1	10	01.05.2008	R11	R12	R13	...	R1S	...	R1P
2	E2	9	01.08.2008	R21	R22	...	...	...	R2...	
...				...	...	...	...	R...S		
k	Ek	11	01.08.2011	RK1	...	...	...			
...				...	...	...				
...				...	...	...				
...				...	...					
...				...	R..2					
N	EN	6	01.01.2020	RN1						

discounted by default date);

3. Discount rate (q, in practice, the average rate for the lending period is often used in a model-uniform sample of all loans);
4. Date of default, (month of default);
5. Repayment payments discounted with the rate (q) on the maturity date, counted from the date of default (exposure period after default).

For the ease of calculation, repayments are sorted in descending order of exposure after default. The applied formulas for calculating the recovery curve are selected from two possible formats:

1. Simple format (medium/frequency)

$$\mathrm{RR_{Avg}}(\tau) = \frac{1}{\mathrm{n}(\tau)} \sum_{\mathrm{i}:\exists \mathrm{V_i}(\tau)} \frac{\sum\limits_{\mathrm{s}\leq\tau} \mathrm{V_i}(\mathrm{s})}{\mathrm{E_i}}, \tag{12}$$

where $n(\tau)$ is the number of default loans that "survived" until the payment of $V_i(\tau)$ in the period τ, i.e. only those loans i are taken into account for which there may be a payment $V\ V_i(\tau)$, $i:\ \exists\ V_i(\tau)$ (obviously, if $\tau = 0$, $n(0) =$ all default loans in the database). $V_i(s)$ is discounted payments in the period s from the moment of default (discount), E_i is amount in the default.

Moreover, the square of the standard deviation (the square of the error $RR_{\mathrm{Avg}}(\tau)$) is substantially heterogeneous due to the different dimension $n(\tau)$ for each period τ. $\delta RR^2(\tau)$ is calculated by the formula:

$$\delta RR_{\mathrm{Avg}}{}^2(\tau) = \frac{1}{n(\tau)^2} \sum_{i:\exists V_i(\tau)} \left(\frac{\sum\limits_{s\leq\tau} V_i(s)}{E_i} - RR_{\mathrm{Avg}}(\tau) \right)^2. \tag{13}$$

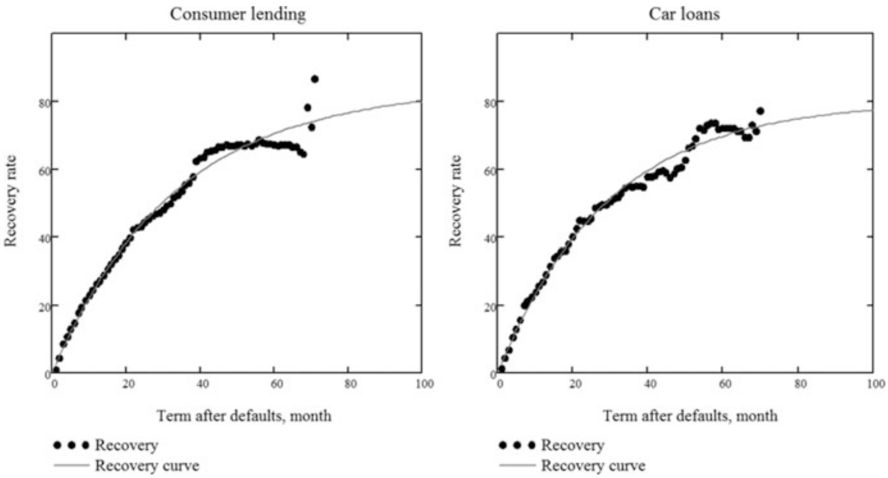


Fig. 4 Examples of constructing recovery curves

2. Weighted average format (taking into account default amounts):

$$\mathrm{RR}_w(\tau) = \frac{\sum\limits_{i:\exists V_i(\tau)} \sum_{s\le\tau} V_i(s)}{\sum\limits_{i:\exists V_i(\tau)} E_i}. \tag{14}$$

The square of the standard deviation can be estimated by the formula:

$$\delta RR_w{}^2(\tau) = \frac{HHI_\tau}{n(\tau)} \sum_{i:\exists V_i(\tau)} \left(\frac{\sum\limits_{s\le\tau} V_i(s)}{E_i} - RR_{Avg}(\tau) \right)^2, \tag{15}$$

where is the Herfindahl–Hirschman index is calculated as:

$$HHI_\tau = \frac{\sum\limits_{i:\exists V_i(\tau)} E_i^2}{\left(\sum\limits_{i:\exists V_i(\tau)} E_i \right)^2}. \tag{16}$$

An example of recovery curves is shown in Fig. 4.

The practice implication shows that the curve $RR(\tau)$ can be approximated with high accuracy by a function of the form:

Table 5 Statistical parameters of recovery curves

Product	Recovery period	Total size	R_∞	T, months	R-sq.	Error ΔR
Consumer lending	2011–2016	1309	83.8%	32.8	97.6%	17.6%
Car loans	2011–2016	228	80.0%	29.4	98.5%	12.3%

$$\rho_\tau(R_\infty, T) = R_\infty \cdot \left(1 - e^{-\frac{\tau}{T}}\right), \tag{17}$$

where T is the average recovery time.

The maturity curve limit $RR(\infty)$ is the recovery forecast for a non-default company, and $LGD(0) = 100\ \% \ - RR(\infty)$, term T is the average recovery period. In the work of (Benjelloun 2019) proposed a method for modeling LGD/RR through a random process, averaging of which gives dynamics close to the behavior of Fig. 4. To approximate $RR(\tau)$ of curve (17), the weighted least squares method is used (see, for example, Strutz 2016), in which the residual is calculated in the Euclidean metric with weights $\frac{1}{\delta RR^2(\tau)}$ and is minimized by the parameters $\boldsymbol{R}_\infty$ (limit recovery) and $\boldsymbol{T}$ ((average recovery period):

$$L(\mathrm{RR}_\infty, T) = \sum_\tau \frac{1}{\delta RR^2(\tau)} \cdot \left(\mathrm{RR}(\tau) - \rho_\tau(R_\infty, T)\right)^2 \rightarrow \min{}_{R_\infty, T}. \tag{18}$$

In this case, the error δR_∞ of the estimate R_∞ is estimated using linearized regression (18) at the optimal point $\boldsymbol{R}_\infty, \boldsymbol{T}$. The detailed formula for estimating δR_∞ is given in Appendix 1.

The output is a calculation of the "slow" values of $\boldsymbol{R}_\infty^\Omega$ and $\boldsymbol{T}^\Omega$ in the current long-term "viewing window" for interval $[t - \Omega, t]$. For example, for the data in Fig. 4 values of recovery parameters were calculated (see Table 5).

Numerous empirical calculations show a high level of fit of the recovery curve using the parametric formula (17), for example, for retail products and consumer lending R-sq. = 97–99%.

3 Recovery Indicators

For a company that has an exposure in default with a period of τ and a certain negative account balance, the loss forecast will be estimated using the conditional LGD (τ):

$$\mathrm{LGD}(\tau) = \frac{1 - R_\infty}{1 - RR(\tau)}, \tag{19}$$

or, using the parametric formula (17):

$$\mathrm{LGD}(\tau) = \frac{1 - R_\infty}{1 - R_\infty \cdot (1 - e^{-\frac{\tau}{T}})}. \tag{20}$$

Therefore, based on the current estimations, at the time $\tau > 0$, the recovery value RR^i_τ, we can construct an unbiased estimate of recovery "for infinity" as:

$$RR^i_\infty(\tau) = RR^i_\tau + \left(1 - RR^i_\tau\right)(1 - LGD(\tau)), \text{i.e.}$$

$$RR^i_\infty(\tau) = \begin{cases} RR^i_\tau + \begin{cases} \left(1 - RR^i_\tau\right) \dfrac{R_\infty \cdot e^{-\frac{\tau}{T}}}{1 - R_\infty \cdot (1 - e^{-\frac{\tau}{T}})}, & \text{recovery process is not completed} \\ 0, & \text{recovery process ended.} \end{cases} & , \tau > 0 \\ 0, & \tau \le 0 \end{cases}. \tag{21}$$

Obviously, for large waiting times τ after default, the correction to RR^i_τ, estimated by the second term in (6), tends to zero and $RR^i_\infty(\infty) = R^i_\infty$, which goes to the statistical base model LGD/RR.

Evaluation (21) should be used as a model estimate of the expected recovery of the debt of borrower in the case when the period after default has not passed, sufficient so that the issue of debt recovery is considered closed. Then it makes sense to determine the recovery indicator for the entire model-homogeneous segment of the population. Recovery indicator determines the forecast of recovery on loans that defaulted on a given "short" indicative moving horizon $[t - \omega, t]$. A simple (or a medium) recovery indicator is constructed as:

$$1. RR^\omega_{Avg}(t) = \frac{1}{N^\omega(t)} \cdot \sum_{i=1}^{N^\omega(t)} RR^i_\infty(t - t_i), \tag{22}$$

2. And, a weighted average indicator, taking into account the amounts of E_i at the time default, t_i, is constructed as:

$$RR^\omega_w(t) = \frac{\sum\limits_{i=1}^{N^\omega(t)} RR^i_\infty(t - t_i) \cdot E_i}{\sum\limits_{i=1}^{N^\omega(t)} E_i}, \tag{23}$$

where $N^\omega(t)$ is the number of borrowers defaulted on a given "short" indicative interval $[t - \omega, t]$.

The recovery indicator is of a great practical importance for monitoring the process of collecting defaulted debts, the strategy for securing loans, segmenting credit policy, etc. If the average recovery indicator exceeds the weighted average, then this means small loans (below average) are more easily repaid than large ones and vice versa.

4 Residual Risk at Loss Given Default Models

The question of residual risk LGD is associated with at least two risk drivers of unexpected losses, which can be underestimated when calculating the requirements for the own economic capital of the loan portfolio. The first driver is macroeconomic, this is a possible correlation of the default rate (i.e. PD) of the loan portfolio and the average LGD, associated with crisis phenomena in the economy, as well as the correlation of the average LGD with other macroeconomic factors. The second driver is local, it is associated with the LGD uncertainty (volatility), for which a "typical" model (1) with parameter γ has been selected. Historical data on the correspondence between the level of default and the level of recovery after default on the corporate bond market in America and Europe (Moody's data) gives the following dependence for the historical period 1982–2016 (Fig. 5).

According to historical data, the credit risk assessment methodology recommends applying a stress correction to the unperturbed value of losses after default LGD in the form $LGD_{stress} = LGD_0 + (1 - LGD_0)(1 - e^{-17.6 \cdot EDR})$, where EDR is the

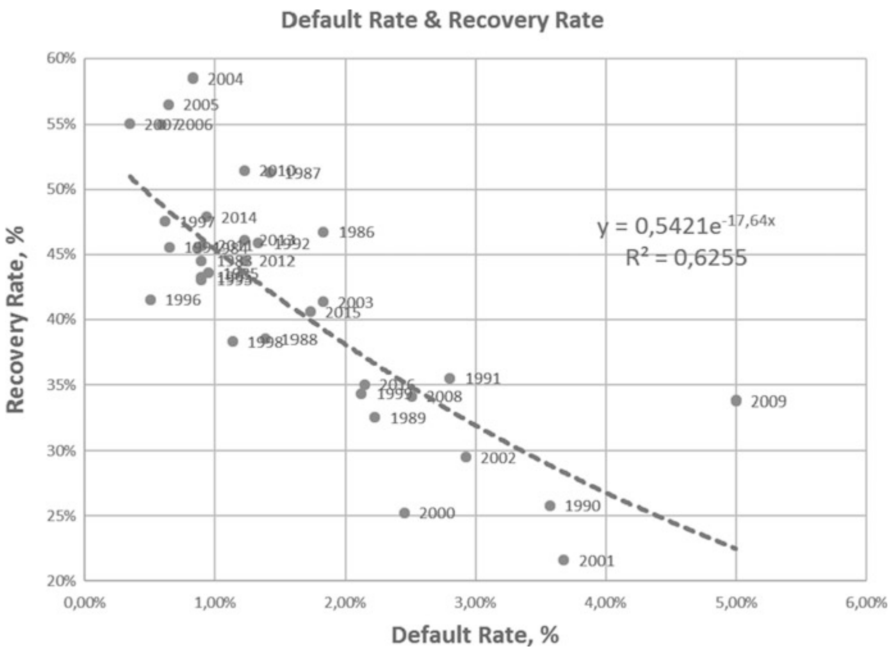


Fig. 5 Historical relationship between the default rate and the recovery rate for the period 1982–2016 according to US corporate bonds and EU (data Moody's 2017)

expected default rate (central tendency), and LGD_0 is the unperturbed LGD value in the stable period. For Moodys data, the $LGD_0 \cong 50\%$.

The correlation problem between PD and LGD (or RR) is one of the key issues in assessing credit risk. For example, a study of (Allen and Saunders 2005) demonstrates calculations according to which the interaction of PD and LGD increases expected losses and capital requirements by up to 30%. However, portfolio credit risk assessment models are often based on the assumption that LGD is fixed and independent of PD. The authors Miu and Ozdemir (2006). note that if PD and LGD correlations are ignored in the model, the LGD should be increased on average by 6% (from 35% to 41%) to compensate for the correlation effect of PD and LGD. At the same time, the results of study of (Ermolova and Penikas 2017) do not allow us to state that there is a relationship between these components of credit risk for the Russian corporate bond market. A generalization of risk metrics that takes into account the dependence of LGD on PD within the framework of the proposed approach can be represented as the dependence of LGD on a random, normally distributed variable, implying that the parameter γ is a constant. In this case, it is recommended to use one of the LGD models (PD (Y)) presented in (Frye and Jacobs 2012) but it should be borne in mind that the basic requirements for the economic capital of an infinitely granular portfolio within the framework of the adjusted one-factor model will differ from the calculation formula recommended by the Basel Committee. Within the framework of approach (1) simulating the dispersion of LGD, the simplest, continuous version of modeling the distribution of losses after default is possible—these are losses Loss = L × EAD with probability pL and losses (Loss = 0) with probability (1 – pL). The parameters L and pL can be determined from the following conditions:

$$\begin{cases} \mathrm{E(Loss)} = \mathrm{LGD} \cdot \mathrm{EAD} \\ \mathrm{D(Loss)} = \gamma \cdot \mathrm{LGD} \cdot (1 - \mathrm{LGD}) \cdot \mathrm{EAD}^2 \end{cases}; \tag{24}$$

These conditions give a unique solution for L and pL:

$$\begin{cases} \mathrm{L} = \gamma + (1 - \gamma) \cdot \mathrm{LGD} \\ \mathrm{pL} = \dfrac{\mathrm{LGD}}{\gamma + (1 - \gamma) \cdot \mathrm{LGD}} \end{cases}. \tag{25}$$

Then, the metrics in which the adjusted PD and EAD can be determined will be set in the form:

$$E_\gamma = \mathrm{EAD} \times (\gamma + (1 - \gamma) \times \mathrm{LGD}), \tag{26}$$

$$\mathrm{PD}_\gamma = \mathrm{PD} \cdot \frac{\mathrm{LGD}}{\gamma + (1 - \gamma) \times \mathrm{LGD}}.$$

The boundary values A: $\gamma = 0$ (the lack of LGD uncertainty) and B: $\gamma = 1$ (maximum LGD uncertainty) will mean, for case A:, $PD_0 = PD$, $E_0 = \mathrm{EAD} \cdot \mathrm{LGD}$; for case B: $PD_1 = PD \cdot LGD$, $E_1 = EAD$.

Obviously, case B implies a greater exposure to default and the capital requirement should be higher for it, despite the fact that the probability of losses will decrease. This issue was investigated in (Witzany 2009). The authors used the one-factor approach to calculating capital recommended by the Basel Committee, taking into account the LGD parameter, first introduced in (Vasicek 1987). Based on the extreme scenarios presented above, it was possible to evaluate VAR (Value at Risk) LGD as the difference between the capital requirement in case B and A. The difference turned out to be positive and monotonous with respect to the model parameters, including expected level of LGD.

In the current approach, we will act similarly in the paradigm of the recommended Basel-2 approach to assessing the requirements for economic capital, created on the basis of the Vasicek formula, under these conditions:

$$\mathrm{UL}_\gamma = E_\gamma \cdot \left(N\left(\frac{N^{-1}(PD_\gamma) + \sqrt{R} \cdot N^{-1}(0.999)}{\sqrt{1-R}} \right) - PD_\gamma \right), \tag{27}$$

where UL is for the estimate of unexpected losses at the recommended reliability level of 0.999 (can be changed), $N(\cdot)$ and $N^{-1}(\cdot)$ are the standard normal and inverse distributions, respectively, R is the correlation parameter, E_γ, PD_γ from equation (7). The UL_0 is the standard recommended form for evaluating the capital of the Basel-2 Advanced Approach. Define $ULGD_\gamma$ as a contribution to equity in relation to EAD:

$$\mathrm{ULGD}_\gamma = \frac{\mathrm{UL}_\gamma - \mathrm{UL}_0}{\mathrm{EAD}}, \tag{28}$$

which will be responsible for the influence of the dispersion parameter γ of LGD on capital requirements (i.e., unexpected losses).

$$\mathrm{ULGD}_\gamma = (\gamma + (1-\gamma) \cdot \mathrm{LGD}) \cdot \mathrm{N}\left(\frac{\mathrm{N}^{-1}(\mathrm{PD}_\gamma) + \sqrt{\mathrm{R}} \cdot \mathrm{N}^{-1}(0.999)}{\sqrt{1-\mathrm{R}}} \right) - \mathrm{LGD} \cdot \mathrm{N}\left(\frac{\mathrm{N}^{-1}(\mathrm{PD}) + \sqrt{\mathrm{R}} \cdot \mathrm{N}^{-1}(0.999)}{\sqrt{1-\mathrm{R}}} \right). \tag{29}$$

Obviously, $\mathrm{ULGD}_0 = 0$. Figure 6 shows graphs of ULGD_γ behavior over the entire range of values $\gamma \in [0, 1]$.

Figure 6 shows that, the values of the correlation R, reliability 0.999 and PD, the capital requirements monotonously increase with increasing uncertainty coefficient γ. Figure 7 shows the surfaces $\frac{d}{d\gamma}\mathrm{ULGD}_\gamma$ at the extreme points $\gamma = 0$ (upper surface) and $\gamma = 1$ (lower surface). In the entire “working” range PD, LGD $\in [0, 1]$, the surfaces are located above the zero plane.

The study shows that the parameter γ is monotonic with respect to unexpected losses and its growth leads to an increase in the additional capital requirement due to

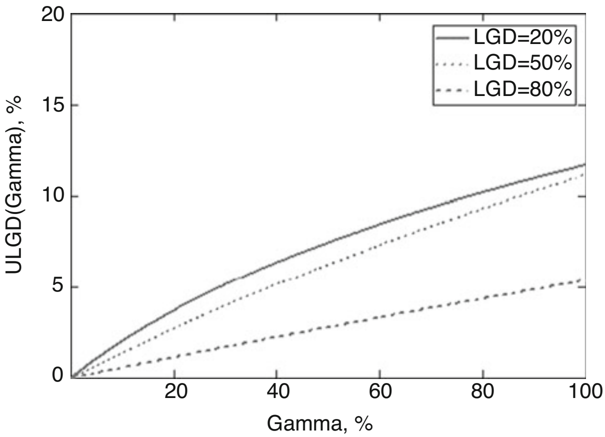


Fig. 6 A graph of the dependence of the additional ULGD requirement for capital on γ (in %) for PD = 10%, correlation $R = 0.2$, and significance level at 99.9%

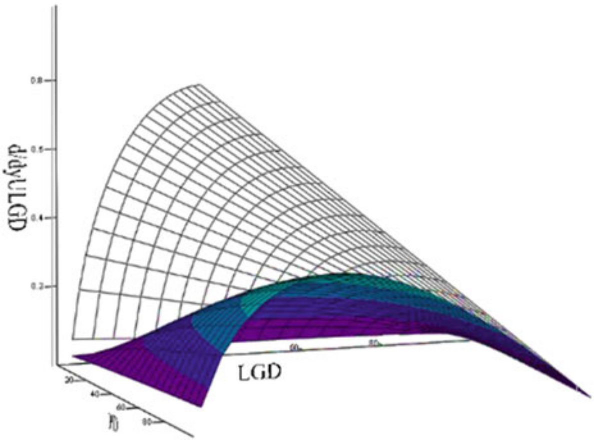


Fig. 7 The surfaces of the derivatives $\frac{\mathrm{d}}{\mathrm{d}\gamma}\mathrm{ULGD}_{\gamma}$ for the correlation value $K = 0.2$ and the reliability 0.999. Lower for $\gamma = 1$, upper for $\gamma = 0$ over the area space of PD, LGD $\in [0, 1]$

the dispersion of LGD. Therefore, when developing the LGD model, it is reasonable to minimize the uncertainty parameter γ.

The largest contribution to capital will be at $\gamma = 1$ and the probability of default PD = 1:

$$\mathrm{ULGD}_{\max}(\mathrm{LGD}) = N\left(\frac{N^{-1}(\mathrm{LGD}) + \sqrt{R}\cdot N^{-1}(0.999)}{\sqrt{1-R}}\right) - \mathrm{LGD}. \tag{30}$$

Figure 8 shows a graph of $\mathrm{ULGD}_{\max}(\mathrm{LGD})$ and the correlation $R = 0.2$.

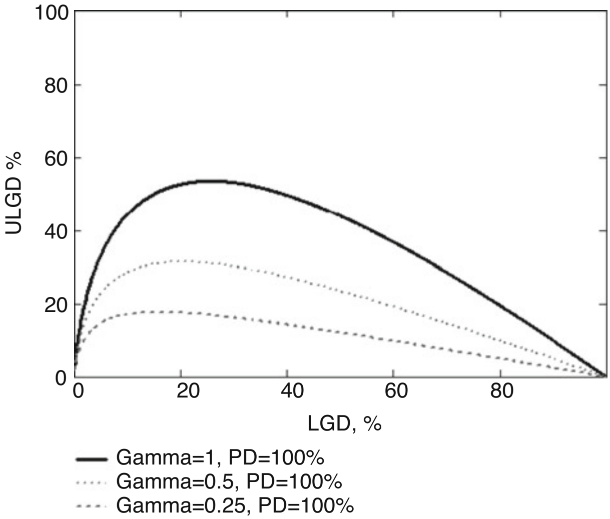


Fig. 8 The graph of the contribution to capital due to the dispersion of LGD for the values PD = 100%, maximum $\gamma = 1$ (black), $\gamma = 0.5$ (light gray), $\gamma = 0.25$ (dark gray)

The maximum function $\mathrm{ULGD}_{\max}(LGD)$ achieved when:

$$\mathrm{LGD}^* = N\left(\frac{\sqrt{(1-R)\left(N^{-1}(0.999)^2 - \ln(1-R)\right)} - N^{-1}(0.999)}{\sqrt{R}}\right). \quad (31)$$

For correlation parameters $R = 0.2$ and significance level (0.999) $\mathrm{LGD}^* = 25.5\%$. Obviously, the shift of the shift down of unexpected losses is towards LGD $<50\%$. This indicates increased responsibility for the model in the event of a model error in the direction of lowering LGD (increasing RR).

5 Optimal Loss Given Default Model from the Point of Residual Risk

Let introduce θ as the dimension LGD[3] (or RR) rating of an indifferent internal structure. The linear model $\widehat{R}_\theta$ of the recovery level RR relative to the rating θ can be estimated as:

[3]LGD rating means any specially developed function that depends on the risk-dominant parameters of LGD/RR, which correlates with the implemented LGD/RR.

$$\widehat{R}_\theta = \widehat{R} + \mu \cdot \frac{\theta - \widehat{\theta}}{\delta\theta} \cdot \delta R, \tag{32}$$

where $\widehat{R}$[4] is the mean value of n realized recoveries of level R, in other words, $\widehat{R} = \frac{1}{\mathrm{n}}\sum_\theta R$, δR is the standard deviation of R, measured by a biased estimation as $\delta R^2 = \frac{1}{\mathrm{n}}\sum_\theta \left(R - \widehat{R}\right)^2$.

Equally, $\widehat{\theta}$ is defined as the average value of θ over the entire set of reconstruction implementations on which the model is built, $\widehat{\theta} = \frac{1}{\mathrm{n}}\sum_\theta \theta$, $\delta\theta^2 = \frac{1}{\mathrm{n}}\sum_\theta \left(\theta - \widehat{\theta}\right)^2$. The most important parameter sought for model (32) is μ—multiplicator, which should depend on the risk-determinism of the LGD rating and minimize the LGD dispersion coefficient indicated by the γ RR/LGD dispersion parameter. The observed recovery of R will be determined by the random variable ε and the model $\widehat{R}_\theta$ in the form $R = \widehat{R}_\theta + \varepsilon$, where the variance ε is modeled, according to (32), by the relation as:

$$D\varepsilon = \gamma \cdot \widehat{R}_\theta\left(1 - \widehat{R}_\theta\right). \tag{33}$$

In this case, the mathematical expectation Mε = 0 by the definition of the model. Further, at the input of the model, it is necessary to determine the correlation ρ between the implemented restorations R and the LGD rating indicated by θ, the estimate of which will be given by the equation:

$$\rho = \frac{1}{N}\sum_\theta \frac{\left(\mathrm{R} - \widehat{\mathrm{R}}\right)\left(\theta - \widehat{\theta}\right)}{\delta\mathrm{R} \cdot \delta\theta}. \tag{34}$$

The more complex, non-linear LGD model in practice makes little sense. It will not provide a significant increase in the estimation accuracy due to the high volatility of LGD due to the two-mode distribution of Fig. 1. The proposed linear LGD model does not automatically guarantee natural restrictions on the simulated recovery level $\widehat{R}_\theta \in [0, 1]$ such as, the popular logistic representation of the type $\widehat{R}_\theta = \frac{1}{1+e^{A\theta+B}}$, but practice shows (see Sect. 6) that the LGD model cannot be created so powerful that the results of its forecast differ by multiples.

For example, if we turn to the recommendations on LGD of the Basel Committee [Basel II 2006], then the recommendations of the minimum LGD vary in the range of 35–45%. Below these values, LGD can be formally evaluated only if there is financial security, which, in fact, should adjust the exposure to default EAD, and not LGD. If this is not done, then LGD uncertainty model is formally destroyed, since financial security is a 100% realizable recovery.

[4]The mean is in the sense of RR_{avg} according to the app. A.2.

Below we will show the range of parameters $\widehat{R}, \rho$ for which the linear model does not go beyond the limits of natural restrictions. Passing to estimates of the observed quantities, it can be equated as[5]:

$$n \cdot MSE = \sum_{\theta} \left(R - \widehat{R}_{\theta}\right)^2 = \sum_{\theta} \varepsilon^2 = \sum_{\theta} \mathrm{D}\varepsilon = \gamma \cdot \sum_{\theta} \widehat{\mathrm{R}}_{\theta}\left(1 - \widehat{\mathrm{R}}_{\theta}\right)$$
$$= \gamma \cdot \left(\sum_{\theta} \widehat{R}\left(1 - \widehat{R}\right) - \sum_{\theta} \left(\frac{\theta - \widehat{\theta}}{\delta\theta}\right)^2 \delta R^2 \cdot \mu^2 \right)$$
$$= \gamma \cdot n \cdot \left(\widehat{R}\left(1 - \widehat{R}\right) - \delta R^2 \cdot \mu^2\right). \tag{35}$$

Otherwise, it can be written as:

$$\mathrm{n} \cdot \mathrm{MSE} = \sum_{\theta} \left(R - \widehat{R}_{\theta}\right)^2 = \sum_{\theta} \left(R - \widehat{R} - \mu \cdot \frac{\theta - \widehat{\theta}}{\delta\theta} \cdot \delta R\right)^2$$
$$= \sum_{\theta} \left(R - \widehat{R}\right)^2 - 2\mu \cdot \delta R^2 \sum_{\theta} \frac{\left(R - \widehat{R}\right)\left(\theta - \widehat{\theta}\right)}{\delta R \cdot \delta\theta} + \sum_{\theta} \left(\frac{\theta - \widehat{\theta}}{\delta\theta}\right)^2 \delta R^2 \cdot \mu^2$$
$$= n \cdot \delta R^2 \cdot \left(1 - 2\mu \cdot \rho + \mu^2\right). \tag{36}$$

Equating the expressions obtained above, the dependence $\gamma\ (\mu)$ is described as:

$$\gamma(\mu) = \gamma_0 \cdot \frac{1 - 2\mu \cdot \rho + \mu^2}{1 - \gamma_0 \cdot \mu^2}, \tag{37}$$

where $\frac{\delta R^2}{\widehat{R}\left(1 - \widehat{R}\right)} = \gamma_0$ is denoted is the value of the parameter γ for the case that is not sensitive to the LGD estimation model considered in Sect. 2.

To find the solution for the optimal value of μ, the problem can be solved with:

$$\mu^* = \mathrm{argMin}_{\mu} \gamma(\mu), \tag{38}$$

where the optimal point for solution is $\gamma^* = \gamma(\mu^*)$.

Problem (38) is solved by the standard method of finding the minimum of a function using the first derivative optimum condition $\gamma'(\mu^*) = 0$. Without bothering the reader with standard mathematical calculations, one can write out the solution to (38):

[5] MSE—Mean Square Error.

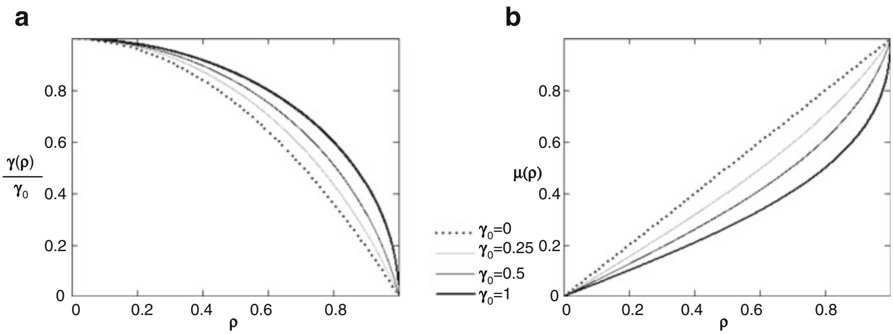


Fig. 9 The dependences of $\frac{\gamma(\rho)}{\gamma_0}$ LGD dispersion parameter (**a**) and μ (ρ) – multiplicator of model (**b**) from correlation ρ upon solution (33)

$$\mu^* = \rho \cdot \frac{2}{1 + \gamma_0 + \sqrt{(1+\gamma_0)^2 - 4\gamma_0\rho^2}},$$

$$\gamma^* = \gamma_0 \cdot \left[1 - \frac{2\rho^2}{1 + \gamma_0 + \sqrt{(1+\gamma_0)^2 - 4\gamma_0\rho^2}}\right], \tag{39}$$

$$\mathrm{MSE}^* = \delta R^2 \cdot \left[1 - 4\rho^2 \cdot \frac{\gamma_0 + \sqrt{(1+\gamma_0)^2 - 4\gamma_0\rho^2}}{\left(1 + \gamma_0 + \sqrt{(1+\gamma_0)^2 - 4\gamma_0\rho^2}\right)^2}\right].$$

For $\rho = 0$ (in the case when the LGD rating does not work properly), an obvious solution is obtained $\mu^* = 0$, $\gamma^* = \gamma_0$, $MSE^* = \delta R^2$.

Figure 9 shows the graphs of solutions (39) in the full range of non-negative correlation of the LGD rating with real measurements for different levels of LGD dispersion.

It can be seen from Fig. 9 that the effect of minimizing the dispersion of LGD becomes most significant as the risk-determinism of the LGD rating increases. However, for the optimal parameter μ of the LGD model, the effect appears immediately and μ becomes less than ρ as soon as the LGD volatility appears. The boundary parameters for the proposed linear model (32) are calculated from the condition: $0 \le \widehat{R}_\theta \le 1$. Assume, without loss of generality, that the rating θ is normally distributed over the interval [0; 1],[6] when $\widehat{\theta} = \frac{1}{2}$, $\delta\theta = \frac{1}{\sqrt{12}}$.

[6]A normal distribution of the random parameter ξ can be described using the substitution for $F(\xi)$, where F is the distribution function of ξ.

According to the model: $\delta R = \sqrt{\gamma_0 \cdot \widehat{R}\left(1 - \widehat{R}\right)}$, then the boundary values of recovery will be

$$\widehat{R_\theta}^{\pm} = \widehat{R} \pm \mu \cdot \sqrt{3 \cdot \gamma_0 \cdot \widehat{R}\left(1 - \widehat{R}\right)}. \tag{40}$$

It means that: $\mu_{\max} = \frac{\min\left(\widehat{R}, 1 - \widehat{R}\right)}{\sqrt{3 \cdot \gamma_0 \cdot \widehat{R}\left(1 - \widehat{R}\right)}}$.

Avoiding the analysis of the full variety of the three-dimensional parameter region $\widehat{R}, \gamma_0, \rho$, in which the restriction $0 \leq \widehat{R}_\theta \leq 1$ is satisfied, we will calculate μ_{max} for typical LGD parameters according to the recovery of US corporate bonds (see Sect. 2). For them, $\gamma_0 = 0.34$, $\widehat{R} = 38.7\%$ $\mu_{\max} = 0.79$, which corresponds to very high risk-determinism indices of the LGD model with a correlation $\rho > 0.8$, which is not achieved by any models.

In the practically significant range of possible models of LGD ratings and not "extreme" practical levels of average recovery $\widehat{R}$ (that is, not close to 0 and 1), the linear LGD model (32) will not give out a range of predictive recoveries $\widehat{R}_\theta$ beyond the limits of [0,1]. In practice, when constructing the LGD model, it is recommended to convert the LGD rating to a range of uniformly distributed values, evaluate μ^* (39) and check constraint (39).

In the next section, we will consider several public models for the LGD rating and their authors' assessments show the applicability of the approach described.

6 Practical Drivers of Loss Given Default Models

The level of recovery of the borrower after default is very specific and depends on many factors. In the literature (see, for example, (Grunert and Weber 2009) four categories of factors for corporate borrowers are defined (see Fig. 10), which correspond to:

- for the borrower, the company of the borrower, incl. creditworthiness (rating) above all;
- for macroeconomics, incl. default rate;
- for the condition of the loan, incl. collateral in the first place;
- for business relations of the borrower, incl. their intensity.

Factors are divided into quantitative and qualitative groups, involving expert assessment. A set of factors forms a long-list from which factors are selected that correlate with the level of implemented LGD results.

To build models for various asset classes, data sources, and measurement methods, which are classified in Table 6.

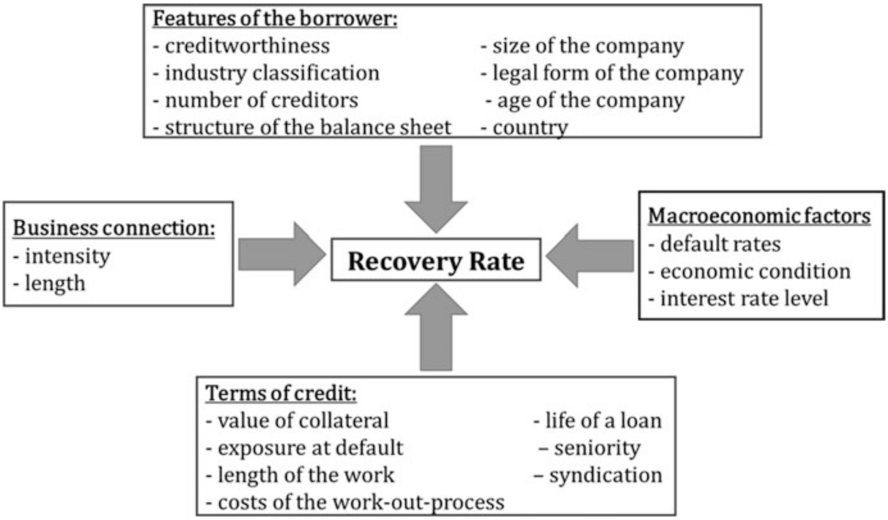


Fig. 10 Drivers for RR/LGD

Table 6 Classification of evaluation methods LGD

Source	Measure	Methods	Exposure
Market values	Price differences	Market LGD	Large corporate, sovereigns, banks
	Credit spreads	Implied market LGD	Large corporate, market LGD sovereigns, banks
Recovery and cost experience	Discounted cash flows	Workout LGD	Retail, SMEs, large corporate
	Historical losses and estimated PD	Implied historical LGD	Retail

Various linear and non-linear algorithms are used to train the LGD classification model. In the literature, (Loterman et al. 2012; Qi and Yang 2009; Bonini and Caivano 2014), a range of methods are analyzed:

– Ordinary Least Squares (OLS);
– Ridge Regression (RiR);
– Robust Regression (RoR);
– Ordinary Least Squares with Beta transformation (B-OLS);
– Beta Regression (BR);
– Ordinary Least Squares with Box-Cox transformation (BC-OLS);
– Regression trees (RT);
– Multivariate Adaptive Regression Splines (MARS);
– Least Squares Support Vector Machines (LSSVM);
– Artificial Neural Networks (ANN);
– Linear regression + non-linear regression (OLS+);
– Logistic regression + (non)linear regression (LOG+).

Table 7 Source data

Dataset	Type	Total size
BANK1	Personal loans	47,853
BANK2	Mortgage loans	119,211
BANK3	Mortgage loans	3351
BANK4	Revolving credit	7889
BANK5	Mortgage loans	4097
BANK6	Corporate loans	4276

Source: Loterman et al. (2012)

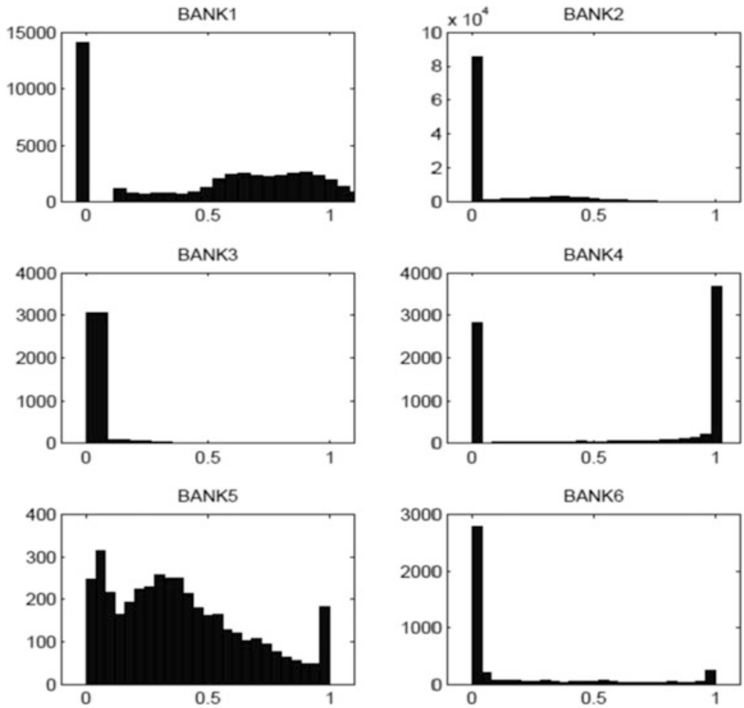


Fig. 11 Density of LGD distribution by Loterman G

Nevertheless, even on impressive empirical data (Table 7), with tens of thousands of measurements for corporate and consumer portfolios of banks, it was found that the obtained models have limited predictive characteristics regardless of which method is used, although non-linear methods give higher characteristics than traditional linear methods. The banks analyzed by the author have unique LGD distributions, which are shown in Fig. 11.

Table 8 shows the result of measuring the linear Pearson correlation predicted and implemented by LGD for different banks. Table 8 shows that significant differences in the results obtained by different methods are observed only for Bank N 3, and for the data of this Bank, even the best models show a weak result. In general, one can

Table 8 The result of measuring the linear Pearsons' correlation predicted and implemented for different LGD methods for different banks

Pearson's R (Cohen et al. 2002) measures the degree of linear relationship between predictions and observations.

Technique	BANK1	BANK2	BANK3	BANK4	BANK5	BANK6
OLS	0.311	0.485	0.117	0.664	0.474	0.350
B-OLS	0.295	0.477	0.077	0.651	0.507	0.305
BR	0.260	0.464	0.157	0.653	0.456	0.321
BC-OLS	0.240	0.472	0.137	0.573	0.501	0.286
RiR	0.306	0.492	0.146	0.666	0.478	0.354
RoR	0.306	0.477	0.173	0.653	0.454	0.349
RT	0.300	0.582	0.387	0.692	0.506	0.339
MARS	0.321	0.558	0.502	0.692	0.567	0.362
LSSVM	0.347	0.569	0.453	0.702	0.579	0.396
ANN	0.360	0.603	0.378	0.705	0.596	0.362
LOG+OLS	0.326	0.484	0.076	0.668	0.498	0.348
LOG+B-OLS	0.317	0.529	0.121	0.665	0.512	0.323
LOG+BR	0.280	0.453	0.074	0.668	0.457	0.335
LOG+BC-OLS	0.213	0.463	0.167	0.666	0.510	0.310
LOG+RiR	0.329	0.539	0.132	0.676	0.492	0.341
LOG+RoR	0.326	0.535	0.151	0.673	0.474	0.339
LOG+RT	0.330	0.555	0.455	0.666	0.500	0.335
LOG+MARS	0.332	0.553	0.488	0.675	0.569	0.329
LOG+LSSVM	0.340	0.559	0.415	0.677	0.580	0.365
LOG+ANN	0.350	0.559	0.538	0.670	0.585	0.369
OLS+RT	0.338	0.579	0.258	0.678	0.536	0.362
OLS+MARS	0.339	0.562	0.502	0.692	0.577	0.363
OLS+LSSVM	0.371	0.567	0.465	0.700	0.576	0.349
OLS+ANN	0.372	0.601	0.261	0.705	0.557	0.350
<r>	**0.32**	**0.53**	**0.28**	**0.67**	**0.52**	**0.34**
dr	**0.04**	**0.05**	**0.17**	**0.03**	**0.05**	**0.02**

Source: Cohen et al. (2002)

notice that the linear OLS model gives an average level result, for corporate bank N6 even above the average.

The study (Seidler et al. 2017) presented the LGD model, trained in the Czech consumer lending market. The aim of the study was to show that lag macrovariables involved in the delayed model are still strong risk factors. As a result, the authors agreed on a meaningful set of factors presented in Table 9.

The following informative LGD model is presented in (Košak and Poljšak 2010). The model has been trained in the rapidly developing small and medium business borrowing market (SME) of Eastern Europe. Table 10 shows the risk-dominant variables that were identified by the authors as defining the LGD model.

Table 12 also presents calculations of model parameters (32) for Košak and Poljšak (2010). The authors used a limited number (124 observations), which

Table 9 Variables included in the LGD model

Explanatory variable logit LDG	Macroeconomic variables, current values	Macroeconomic variables, lagged and lead values
Client- specific factors	Real GDP growth (y-o-y)	Real GDP growth (y-o-y) (t-1)
Exposure at default	Real GDP growth (y-o-y)	Real GDP growth (y-o-y) (t-2)
Relationship with bank	Real Consumption Growth (y-o-y)	Real investment growth (y-o-y) (t-2)
Age	Real Investment Growth (y-o-y)	Unemployment rate (t-8)
Children	Real Pribor3m	Real wage growth (y-o-y) (t-3)
Phone	Inflation rate (y-o-y)	Real wage growth (y-o-y) (t-4)
Employment	Property prices (y-o-y)	Real wage growth (y-o-y) (t-5)
Education	Default rate	
Female	Retail loan growth (y-o-y)	

Source: Seidler et al. (2017)

Table 10 Variables included in the LGD model

Collateral type	Industry	Period	Rating of the borrower before default	EAD
Assignment of receivables Financial collateral Personal guarantee Physical collateral Real Estate collateral Unsecured	Manufacturing Real Service Trade	Long-term loan Short-term loan	Last rating C Last rating D Last rating E.	Large Medium Small

Source: Košak and Poljšak (2010)

gives rise to a tangible statistical error in determining the parameters characterizing the uncertainty. For the parameter γ_0 and γ^* according to formula, the statistical error is at the level of 10%. A third example of the LGD model is proposed to consider a model prepared by linear regression based on 10 years of historical development of real data on corporate and retail loans from a group of European commercial banks under the control of the ECB [Bonini and Caivano 2016]. 26,000 cases were processed, including 7500 large and medium corporate defaults. The result is a recovery level model presented in Table 11.

Table 12 shows the calculations of the parameter γ_0 of the "LGD dispersion" without taking into account the LGD model, the optimal γ^* from the point of view of residual risk after applying model (8), the optimal sensitivity parameter μ^*, and also the range $\widehat{R}_\theta^{\,\pm}$ of possible values for the model RR as it applied in (8). The correlation ρ between the implemented LGD and the model was estimated by the formula $\rho = \sqrt{R_{\text{squared}}}$. The calculations were carried out for three sources in which the parameters of the models are indicated.

Table 12 shows that the model recovery level (8) does not go beyond the range (0.1). Judging by the relation γ^*/γ_0 and Fig. 6, the models presented in Table 12 can

Table 11 Model RR (recovery rate)

Variables	Grouping	Coefficient	*p*-value	Variable weight
Macro-geographical area	Intercept	0.1001	<.0001	
	Center	0.2145	<.0001	13.87%
	North East	0.1113		
	Sud & Island	0.0788		
	North West	0		
Exposure at Default	EAD	0.1567	<.0001	10.13%
Portfolio segmentation	Medium – Large Corporate	0.594	0.0033	38.40%
	Small Business (Retail)	0.377	0.0022	
	Individuals (Retail)	0	<.0001	
Type of product	Mortgages	0.1876	<.0001	12.13%
	Other products	0		
Presence of personal guarantee	Absence	0.1134	<.0001	7.33%
	Presence	0		
Presence of mortgages	Absence	0.1609	<.0001	10.40%
	Presence	0		
Type of recovery process	Out of court	0.1189	<.0001	7.69%
	In court	0.0533		
	No information	0		

Source: Bonini and Caivano (2016).

provide a 10–25% reduction in the residual risk of LGD relative to how if LGD were assessed in the zero-approximation by the average LGD.

Summing up the results of a sample study of the results of RR /LGD modeling performed by different authors on different statistical recovery databases, we can draw the following conclusions:

1. It is impossible to unequivocally give preference to a particular method that is optimal in terms of modeling accuracy. In many cases, for example, see Table 7, an increase in the complexity and accuracy of the methods does not lead to a noticeable improvement in the results of the RR/LGD model and, on the other hand, often to a deterioration;
2. The set of risk-dominant parameters of the RR/LGD model can vary significantly when analyzing the statistical bases of different banks and different economies or different model-homogeneous populations;
3. The average recovery parameters and their dispersion can fluctuate significantly with a narrowing of model-homogeneous populations, including lending segments including in different banks. The maximum accuracy achieved on certain optimal models is also significantly heterogeneous.

The general results of the maximum achieved accuracy of LGD modeling, measured in various metrics, such as the correlation of the realized and model LGD, show a rather modest result. Very rarely a correlation greater than 0.6 is achieved, the average achieved on the best models is about 0.45.

Table 12 Calculations of the parameter γ_0 for LGD dispersion without taking into account the LGD model, optimal γ^* in terms of residual risk after solution of problem (33), which are presented (39), optimal value multiplicator μ^*, and also the boundary values of recovery $\widehat{R}_\theta^{\pm}$, which are described in (40)

Source	Seidler (2017)	Košak and Poljšak (2010)	Bonini and Caivano (2016)
LGD model	GLM[a]	GLM	OLS
Type of asset	Retail, 2003q1-2010q2, 18 698 obs.	SME, 2002 – 2005, 124 obs.	Individuals (Retail), Small size Corporate (Retail), Medium—Large size Corporate, 2002q4-2012q4, 26 000 obs.
Mean value of realized recoveries $\widehat{R}$	0.42	0.73	0.51
Standard deviation of recoveries δR	0.40	0.35	0.46
Pseudo R-squared	0.152 (Adjusted)	0.363 (Nagelkerke)	0.31 (Adjusted)
Starting value dispersion parameter γ_0	0.657	0.622	0.847
Optimal value dispersion parameter γ^*	0.594	0.468	0.692
Optimal value multiplicator of model (9) μ^*	0.245	0.421	0.329
Boundary values of recovery $\widehat{R}_\theta^{\pm}$	0.25–0.59	0.48–0.98	0.25–0.77

[a]Generalized linear model/GLM

All this convincingly argues the practical expediency of using simple methods, such as (9), for which the optimal sensitivity setting is possible to minimize residual risk. The construction of the model is based on the maximum Pearson correlation. The results of other models can be compared with the results of model (9) to identify their effectiveness.

7 Conclusion

In this study, it is proposed an approach to divide the RR/LGD model development process into two stages, namely: the RR/LGD rating model and calibrate the latter using a linear form that minimizes residual risk. The RR/LGD rating model is constructed in such a way as to ensure the maximum Pearson correlation with the implemented RR/LGD on the training statistical sample. In preparing the RR statistical base, correction (4) for the incomplete recovery process for part of the sample is taken into account. To do this, the recovery curve parameters (4) should be estimated using the method (5) on the historical recovery base (see Table 4). At the same time, recovery payments, net of costs, must be cleared of non-payments and discounted at the time of default. Financial support should be included in the EAD model. The RR/LGD rating model is based on risk-dominant factors, examples of which are presented in Sect. 6. In the process of setting the optimum, from the point of view of correlation, RR/LGD rating model, it should be normalized so that the distribution of ratings is statistically (with an acceptable error) uniform.

At the next step, the optimal sensitivity parameter μ is calculated by formula (12) with allowance for the parameter γ_0 of the LGD dispersion and the correlation parameter ρ. When calculating these parameters, the correction for the incomplete recovery process should be taken into account. Including for the recovery sample $ID = 1\ldots N$ according to Table 4:

$$\gamma_0 = \frac{\sum_{ID}(R_{ID} - R(\tau_{ID}))^2}{\sum_{ID} R(\tau_{ID}) \cdot (1 - R(\tau_{ID}))},$$

$$\rho = \frac{\sum_{ID}(R_{ID} - R(\tau_{ID})) \cdot \left(\theta_{ID} - \widehat{\theta}\right)}{\sqrt{\sum_{ID}(R_{ID} - R(\tau_{ID}))^2 \cdot \sum_{ID}\left(\theta_{ID} - \widehat{\theta}\right)^2}}, \tag{41}$$

where R_{ID} is the share of the implemented borrower recovery ID, $R(\tau_{ID})$ is the recovery function (4) if recovery is not completed, or $R(\tau_{ID}) = R_\infty$ if it is completed by the time τ_{ID} after default, $\theta_{ID}, \widehat{\theta}$ is rating borrower's RR/LGD ID and average rating respectively.

The verification of the model is determined by formula (9). The validity of the model within the limits of the model RR restriction should be verified by formula (13). The value of the final adjustment and calibration of the LGD model can be estimated as a percentage of the EAD of economic capital savings on residual risk through the difference $\mathrm{ULGD}_{\gamma_0} - \mathrm{ULGD}_{\gamma^*}$ according to formula (8). For example, a capital saving of 1% EAD is tangible and comparable to the countercyclical capital premium (buffer) introduced by Basel—III (maximum 2.5% from Basel III, 2011). In addition, it is necessary to take into account the forecast/adjustment of the expected average RR (parameter $\widehat{R} = R_\infty$ in formula (9), taking into account the macroeconomic scenario and forecast. A reliable LGD driver, according to Moody's (see Fig. 5), is the central trend of PD.

To check and validate the already built "M" of RR/LGD model, it is necessary to compare it with the reference model (9), built on the data of the "M" model being tested. To do this, calculate the correlation ρ of the implemented LGD- construction with LGD_M, taking into account the possible incompleteness of recovery (all values for LGD_M are recommended to be consistent to a normal distribution). The second step will be the direct calculation of γ_M by the formula (2) for "M." Obviously, the average value of the realized LGD model should follows the rule, i.e. $\widehat{R} \cong R_\infty \pm \delta R_\infty$ (4), where δR_∞ is the error of the R_∞ estimate in problem (5), estimated (A2) in Appendix 1. One of the concepts of the recovery calculation format, or simple frequency, should be adhered to be weighted by means. It is generally accepted to adhere to the "simple" format, and balance on EAD should be taken into account in the LGD model, which depends on EAD. After calculating the optimal γ^* reference model (9) using formula (11), the obtained parameters of the LGD dispersion should be compared. If $\gamma_M > \gamma^* + \sigma\gamma$, where σγ is the statistical error (3), then the "M" model is not optimal and can be improved.

The next step is to check whether the LGD_M values goes beyond the lower limit of constraints (13). The values of LGD_M significantly (outside the statistical error) lower than the lower limit of the constraints $1 - R^+$ (13) are not permissible, since the conservative principle should be violated. In this case, the power of the "M" model is not enough to assign significantly lower values to the LGD model level. This can lead to a significant model risk, transformed into credit risk with the significant volumes for individual loans.

The Estimation Procedure of the Calculated Standard Error for the Average Marginal Share of Repayment

The solution of problem (5) gives the optimal values of the recovery period T and the limiting recovery R_∞. The error of the values depends on the quality statistics of the approximation of the cumulative recovery of the recovery curve (4). The linear problem of the parameter estimation question $\theta = \{R, T\}$ for the non-linear regression problem $(\tau) = \rho_\tau(\theta) + \delta_\tau \cdot \varepsilon_\tau$, near the optimal solution θ of problem (5) is given a linear regression relation for the error $\Delta\theta = \boldsymbol{\theta} - \theta$ in the standardized form:

$$\frac{RR(\tau) - \rho_\tau(\theta)}{\delta_\tau} = \frac{\partial_\theta \rho_\tau}{\delta_\tau}\ \Delta\theta + \varepsilon_\tau, \tag{42}$$

where $\partial_\theta \rho_\tau$ is composed by the $n \times 2$ partial derivatives matrix $\left[\frac{\partial}{\partial R}\rho_\tau(R, T), \frac{\partial}{\partial T}\rho_\tau(R, T)\right]$, ε_τ assumed to be normal uncorrelated random variable with unknown variance for each recovery period τ , of which there are n. Apparently, for an optimal solution in the sense of equation (5) for θ, the solution of problem (A1) for $\Delta\theta$ will be obvious $\Delta\theta = 0$. However, the error $\Delta\theta$ will be expressed through the covariance matrix according to the well-known formula (see, for example, Strutz 2016):

$$cov(\Delta\theta) = \left(\left[\frac{\partial_\theta \rho_\tau}{\delta_\tau} \right]^T \times \left[\frac{\partial_\theta \rho_\tau}{\delta_\tau} \right] \right)^{-1} \cdot \frac{RSS}{n-2}, \tag{43}$$

where for (A1):

$$RSS = \sum\nolimits_\tau \frac{1}{{\delta_\tau}^2} (RR(\tau) - \rho_\tau(\theta))^2.$$

Denoting the partial derivatives as:

$$\begin{aligned} \rho_\tau &= R \cdot \left(1 - e^{-\frac{\tau}{T}}\right); \\ \partial_R \rho_\tau &= 1 - e^{-\frac{\tau}{T}}; \\ \partial_T \rho_\tau &= -Re^{-\frac{\tau}{T}} \frac{\tau}{T^2}, \end{aligned} \tag{44}$$

and according for the estimation error R, the only the upper diagonal element of the matrix $cov(\Delta\theta)$, it is needed to obtain

$$\delta R^2 = \frac{1}{n-2} \cdot \frac{-\sum_\tau \frac{\partial_R \rho_\tau \cdot \partial_T \rho_\tau}{{\delta_\tau}^2} \cdot \sum_\tau \frac{(RR(\tau) - \rho_\tau)^2}{{\delta_\tau}^2}}{\sum_\tau \frac{{\partial_R \rho_\tau}^2}{{\delta_\tau}^2} \cdot \sum_\tau \frac{{\partial_T \rho_\tau}^2}{{\delta_\tau}^2} - \left(\sum_\tau \frac{\partial_R \rho_\tau \cdot \partial_T \rho_\tau}{{\delta_\tau}^2} \right)^2}. \tag{45}$$

To estimate the error R_∞ as the measure for the standard deviation δR_∞, it is necessary in formula (45) to substitute the solution of problem (5) as R—the limiting recovery R_∞, the time for recovery T, and ${\delta_\tau}^2 = \delta RR_{\text{Avg}}{}^2(\tau)$ or $\delta RR_w{}^2(\tau)$, these replacements depend on the calculation of the recovery curve.

References

Allen, L. Saunders, A. (2005). *A Survey of Cyclical Effects in Credit Risk Measurement Models*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=315561

Antonova, B. N. (2012). Ocenka stavki vosstanovleniya po rossijskim korporativnym obligaciyam. *Journal of Corporate Finance Research, 6*(4), 130–143. [in Russian].

Araten, M., Jacobs, M., & Varshney, P. (2004). Measuring LGD on commercial loans: An 18-year internal study. *RMA Journal, 86*(8), 96–103.

Basel II. (2006). *International convergence of capital measurement and capital standards: a revised framework - comprehensive version*. Part 2. pp. 295. https://www.bis.org/publ/bcbs128b.pdf

Basel III. (2011). *A global regulatory framework for more resilient banks and banking systems*. Part 1, pp. 139. https://www.bis.org/publ/bcbs189.pdf

Benjelloun, M. (2019). *Stochastic modelling of the loss given default (LGD) for non-defaulted assets*. This work was supported by the Global Research & Analytics Dept. of Chappuis Halder

& Co. https://www.chappuishalder.com/wp-ontent/uploads/2019/03/GRA_White_Paper_LGD_stochastic.pdf
Bonini, S., & Caivano, G. (2014). Development of a LGD model basel2 compliant: A case study. In *Mathematical and Statistical Methods for Actuarial Sciences and Finance, 2014* (pp. 45–48). Springer. https://doi.org/10.1007/978-3-319-05014-0_10. ISBN: 3319050141.
Bonini, S., & Caivano, G. (2016). *Econometric approach for Basel II Loss Given Default Estimation: From discount rate to final multivariate model*. White paper. https://www.semanticscholar.org/paper/Econometric-approach-for-Basel-II-Loss-Given-%3A-from/fc954a9ac25c2fd9f2de67b9f89b53aa36742e6c
Cohen, M., Nelson, R., & Walsh, P. (2002). Links and impacts: The influence of public research on industrial R&D. *Management Science, 48*(1), 1–23.
CreditMetrics. (1997). *Technical Document*. JP Morgan.
Dermine, J., & de Carvalho, C. N. (2006). Bank loan losses-given-default: A case study. *Journal of Banking and Finance, 30*(4), 1219–1243.
Ermolova, M. D., & Penikas, H. I. (2017). PD-LGD correlation study: Evidence from the Russian corporate bond market. *Model Assisted Statistics and Applications, 12*(4), 335–358.
Felsovalyi, A., & Hurt, L. (1998). Measuring loss on Latin American defaulted bank loans: A 27-year study of 27 countries. *Journal of Lending and Credit Risk Management, 80*, 41–46.
Frye, J., & Jacobs, M., Jr. (2012). Credit loss and systematic loss given default. *The Journal of Credit Risk, 8*(1), 109–140.
Gordy, M. B., & Lutkebohmert, E. (2013). Granularity adjustment for regulatory capital assessment. *International Journal ofCentral Banking, 9*(3), 38–77.
Grunert, J., & Weber, M. (2009). Recovery rates of commercial lending: Empirical evidence for German companies. *Journal of Banking & Finance, Elsevier, 33*(3), 505–513. https://doi.org/10.1016/j.jbankfin.2008.09.002.
Jankowitscha, R., Naglerb, F., & Subrahmanyam, M. G. (2014). The determinants of recovery rates in the US corporate bond market. *Journal of Financial Economics., 114*(1), 155–177. https://doi.org/10.1016/j.jfineco.2014.06.001.
Karminsky, A. M., Lozinskaia, A. M., & Ozhegov, E. M. (2016). Estimation methods of creditor's loss in residential mortgage lending. *HSE Economic Journal, 20*(1), 9–51. [in Russian].
Košak, M., & Poljšak, J (2010). *Loss given default determinants in a commercial bank lending: an emerging market case study*. Zbornik radova Ekonomskog Fakulteta u Rijeci: časopis za ekonomsku teoriju i praksu. - Rijeka, ISSN 0353-3689, ZDB-ID 12830379. - 28.2010, 1, p. 61–88.
Loterman, G., Brown, I., Martens, D., Mues, C., & Baesens, B. (2012). Benchmarking regression algorithms for loss given default modeling. *International Journal of Forecasting, 28*, 161–170.
Miu, P., & Ozdemir, B. (2006). Basel requirement of downturn LGD: Modeling and estimating PD & LGD correlations. *Journal of Credit Risk, 2*(2), 43–68.
Moody's Corporation. (2017). Annual Default Study: Corporate Default and Recovery Rates, 1920–2016 - Excel data.
Qi, M., & Yang, X. (2009). Loss given default of high loan-to-value residential Mortgages. *Journal of Banking and Finance, 33*(5), 788–799.
Qi, M., & Zhao, X. (2011). Comparison of modeling methods for loss given default. *Journal of Banking and Finance, 35*(11), 2842–2855.
Schuermann, T. (2004). *What do we know about loss given default?* (February 2004). Wharton Financial Institutions Center Working Paper No. 04-01. Available at SSRN: https://doi.org/10.2139/ssrn.525702
Seidler, J., Konečný, T., Belyaeva, A., Belyaev, K. (2017). *The time dimension of the links between loss given default and the macroeconomy*. Working Paper Series 2037, European Central Bank. https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp2037.en.pdf
Strutz, T. (2016). *Data Fitting and Uncertainty (A practical introduction to weighted least squares and beyond)*. Springer Vieweg. ISBN 978-3-658-11455-8., paper 3.

Vasicek, O. (1987). *Probability of Loss on a Loan Portfolio*. [on-line], San Francisco, KMV, c1987, [cit. 27th March, 2009]. http://www.moodyskmv.com/research/files/wp/Probability_of_Loss
Vujnović, M., Nikolić, N., & Vujnović, A. (2016). Validation of loss given default for corporate. *Journal of Applied Engineering Science, 14*(4), 465–476. https://doi.org/10.5937/jaes14-11752.
Witzany, J. (2009). Unexpected recovery risk and LGD discount rate determination. *European Financial and Accounting Journal*, ISSN 1805-4846, University of Economics, Faculty of Finance and Accounting Prague*, 4*(1), 61–84. https://doi.org/10.18267/j.efaj.63.