Loss Given Default Modelling: Comparative Analysis

Yashkir, Olga and Yashkir, Yuriy

Yashkir Consulting

27 March 2013

Loss Given Default Modelling: Comparative Analysis

(pre-typeset version)

(Final version is published in "Journal of Risk Model Validation", v.7,No.1(2013))

Dr. Olga Yashkir * Dr. Yuri Yashkir †

Abstract

In this study we investigated several most popular Loss Given Default (LGD) models (LSM, Tobit, Three-Tiered Tobit, Beta Regression, Inflated Beta Regression, Censored Gamma Regression) in order to compare their performance. We show that for a given input data set, the quality of the model calibration depends mainly on the proper choice (and availability) of explanatory variables (model factors), but not on the fitting model. Model factors were chosen based on the amplitude of their correlation with historical LGDs of the calibration data set. Numerical values of non-quantitative parameters (industry, ranking, type of collateral) were introduced as their LGD average. We show that different debt instruments depend on different sets of model factors (from three factors for Revolving Credit or for Subordinated Bonds to eight factors for Senior Secured Bonds). Calibration of LGD models using distressed business cycle periods provide better fit than data from total available time span. Calibration algorithms and details of their realization using the R statistical package are presented. We demonstrate how LGD models can be used for stress testing. The results of this study can be of use to risk managers concerned with the Basel accord compliance.

*olga.yashkir@gmail.com; YASHKIR CONSULTING LTD.
†yuri.yashkir@gmail.com; YASHKIR CONSULTING LTD.

Contents

1 Introduction

The goal of Loss Given Default (LGD) modelling is to produce simulated LGDs close to and as correlated with historical LGDs. Difficulties with modelling depend directly on the specifics of the data used and on the limitations of the models.

In recent years, the importance of modelling LGD has increased significantly. The LGD model development, calibration, and implementation strategies have been analysed and summarized in several publications (Gupton, 2005), (Schuermann, 2004).

The predictive power of any LGD model depends, first of all, on proper choice (and availability) of the model input parameters obtained from obligor's information. These (predictive) variables were analyzed and used for LGD model calibration in many publications. For example, the nine-factor model was analysed in (Gupton and Stein, 2002), the survey of LGD model factors is presented in (Friedman and Sandow, 2003). A case study of the modelling of bank loan LGDs where the primary factors (the period of loan origination, quality of the collateral, the loan size, and the length of the relationship with the obligor) were identified (Chalupka and Kopecsni, 2009). The link between default and recovery rates was investigated in (Altman et al., 2003), (Altman et al., 2004), (Altman, 2006a), (Altman, 2006b). The incorporation of the dependence between probabilities of default and recovery rates investigated by (Bade et al., 2011) demonstrated some improvement of the LGD model. A significant impact of the uncertainty of model parameters on estimated LGDs was demonstrated in (Luo and Shevchenko, 2010). The influence of the length of the LGD workout process on the level of estimated LGD can be significant, as shown in (Gurtler and Hibbeln, 2011).

The LGD models based on the linear regression approach can be found in (McDonald and Moffitt, 1980) (the Tobit model), (Huang and Oosterlee, 2012) (Beta regression model), (Pereira and Cribari-Neto, 2010) (Inflated beta regression), (Sigrist and Stahel, 2010), (Sigrist and Stahel, 2011) (censored Gamma regression). Authors of (Altman and Kalotay, 2010) used mixture of distributions to model LGD. The beta-component mixture for modelling LGD and CDS rates as model variables was successfully used in (Baixauli and Alvarez, 2010). The bimodal structure of the LGD distribution was modelled by a mixture of two beta distributions (Hlawatsch and Ostrowski, 2011). The LGD model in the merton-structured credit risk framework was also investigated in (Jacobs Jr., 2011).

The portfolio credit risk model dependent on LGD was developed in (Hillebrand, 2006) and compared with several alternative LGD models. Calibration methods for LGD models applied to mortgage markets can be found in (van der Weija and den Hollandera, 2009). The results in (Bellotti and Crook, 2012) contain comparison of several models (Tobit, decision-tree model, beta transformation, fractional Logit, and the Least Square method). They demonstrated the importance of the inclusion of macroeconomic conditions (interest rates, unemployment levels, and earning index) for the LGD model stress testing. The paper (Yang and Tkachenko, 2012) proposes some empirical approaches for EAD/LGD modeling and provides technical insights into their implementation. Validation techniques and performance metrics for loss given default models were introduced by (Li et al., 2009).
An attempt to develop analytic formulas for downturn LGD estimation was done by (Barco, 2007). The downturn LGDs were considered as a 1/1000-year event with account of correlated PD and LGD. The paper by (Rosch and Scheule, 2007) developed a framework to stress sensitivities of risk drivers, and therefore a credit portfolio losses.
Given results of all the above research publications, the main question for a practitioner remains: what is the best model for LGD estimation? The goal of our research is to provide comparative tests of popular LGD estimation models, to analyze their performance, to calibrate the models on different data subsets, and, in addition, to provide recommendations on how test results can be used for the stress testing of LGDs.
We do not include data manipulation techniques. Based on some examples, we show how models can become sensitive to the choice of data.

2 Methodologies

The LGD models, analyzed and compared in this paper, are based on several different linear regression algorithms. A short description of the models is summarized in this section.

Censored Least Square Method

Given known historical LGD values $\vec{LGD}^*$, coefficients x_k are derived using the Least Square Method (LSM) by minimizing the following object function:

$$\min_{\vec{x}} \sum_i \left(y_i(\vec{x}, \mathbf{r}) - LGD_i^*\right)^2 \tag{1}$$

with

$$y_i(\vec{x}, \mathbf{r}) = x_0 + \sum_{k=1}^{n} x_k \cdot r_{ik} \tag{2}$$

where

r_{ik} is the k^{th} predictive parameter for i^{th} counterparty ,

x_k is the coefficient for k^{th} predictive parameter ,

x_0 is a constant ("intercept").

The LGD for the debt facility i is estimated as:

$$LGD_i(\vec{x}, \mathbf{r}) = \max\left[0, \min\left[1, y_i(\vec{x}, \mathbf{r})\right]\right] \tag{3}$$

Censored Linear Regression (Tobit) Model

The Tobit LGD model is based on the latent "loss" parameter z_i for each debt facility i (see(McDonald and Moffitt, 1980)):

$$z_i = y_i + \epsilon(\sigma) \tag{4}$$

Here: $\epsilon(\sigma)$ is a normally distributed random driver with standard deviation σ, $y_i(\vec{x}, \mathbf{r})$ is the linear combination of explanatory variables r_{ik} as in (2).
The latent loss variable z_i is a normally distributed random value with expected value of y_i and standard deviation of σ. Therefore, the probability of realization of $LGD_i = s$ can be expressed through the standard Gaussian function $g(\cdot)$ as follows:

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{s-y_i}{\sigma}\right)^2\right] \equiv \frac{1}{\sigma} \cdot g\left(\frac{s-y_i}{\sigma}\right) \tag{5}$$

Assuming that all $s \leq 0$ correspond to $LGD = 0$ and that all $s \geq 1$ correspond to $LGD = 1$ we can define the probability function $P_i(x)$ for LGD as follows:

$$P_i(x) = \begin{cases} P_i^{(0)} & \text{if } x \leq 0 \\ \frac{1}{\sigma} \cdot g\left(\frac{x-y_i}{\sigma}\right) & \text{if } 0 < x < 1 \\ P_i^{(1)} & \text{if } x \geq 1 \end{cases} \tag{6}$$

Where

$$P_i^{(0)} = \frac{1}{\sigma} \cdot \int_{-\infty}^{0} g\left(\frac{z-y_i}{\sigma}\right) \mathrm{d}z = N\left(-\frac{y_i}{\sigma}\right) \tag{7}$$

$$P_i^{(1)} = \frac{1}{\sigma} \cdot \int_{1}^{\infty} g\left(\frac{z-y_i}{\sigma}\right) \mathrm{d}z = N\left(-\frac{1-y_i}{\sigma}\right) \tag{8}$$

Where $N()$ is a standard normal cumulative distribution function.

The probability function (6) can be also presented in a more convenient form

$$P_i(z) = P_i^{(0)} \cdot \delta(z) + P_i^{(1)} \cdot \delta(z-1) + \frac{1}{\sigma} \cdot g\left(\frac{z-y_i}{\sigma}\right) \cdot (1 - \delta(z) - \delta(z-1)) \tag{9}$$

where $\delta()$ is the standard delta-function.

Note that the probability (6) is also the function of model coefficients x_k ($k = 0 : n$) and of the LGD volatility σ which are the subjects of the model calibration. Expected LGD for i^{th} debt facility is calculated as $E[LGD_i] = \int_0^1 z \cdot P_i(z) \cdot \mathrm{d}z$. The result is

$$E[LGD_i] = P_i^{(1)} + (1 - P_i^{(0)} - P_i^{(1)}) \cdot y_i + \sigma\left[g\left(\frac{y_i}{\sigma}\right) - g\left(-\frac{1-y_i}{\sigma}\right)\right]. \tag{10}$$

The cumulative LGD probability can be calculated using (6) as $Q(LGD_i) = \int_0^{LGD_i} P_i(z)\, \mathrm{d}z$:

$$Q_i(LGD) = \begin{cases} P_i^{(0)} & \text{if } LGD = 0 \\ N\left(\frac{LGD-y_i}{\sigma}\right) & \text{if } 0 < LGD < 1 \\ 1 & \text{if } LGD = 1 \end{cases} \tag{11}$$

The η^{th} percentile of the modelled LGDs ($LGD_i^{(\eta)}$) is the solution of the equation $Q_i(LGD) = \eta$.

$$LGD_i^{(\eta)} = \begin{cases} 0 & \text{if } \eta < P_i^{(0)} \\ y_i + \sigma \cdot N^{-1}(\eta) & \text{if } P_i^{(0)} < \eta < 1 - P_i^{(1)} \\ 1 & \text{if } \eta > 1 - P_i^{(1)} \end{cases} \tag{12}$$

The calibration of the Tobit model consists in finding model coefficients x_k ($k = 0 : n$) and LGD volatility σ by best fit of the model (with predictive parameters r_{ik}) to historical data LGD_i ($k = 1 : n$, $i = 1 : J$). If we consider the input data sample as a set of independent "measurements" then the best model fit is obtained by maximizing the total probability of getting this input data set:

$$\hat{P}(\vec{x}, \sigma) = \prod_{i=1}^{J} P_i(LGD_i) \tag{13}$$

which is equivalent to minimization of the following objective function

$$\Phi(\vec{x}, \sigma) = -\sum_{i=1}^{J} \log P_i(LGD_i) \tag{14}$$

For numerical optimization we employ the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method (for solving non-linear optimization problems without constraints).
The linear dependence of the function (2) on explanatory variables r_{kj} may not be sufficient to describe the cause-effect link of LGD_j to r_{kj}. It is possible to increase flexibility of the model by including a quadratic term, such that

$$y_i(\vec{x}, \mathbf{r}) = x_0 + \sum_{k=1}^{n} \left(x_{2k-1} \cdot r_{ik} + x_{2k} \cdot r_{ik}^2 \right) \tag{15}$$

Note that number of model coefficients for the nonlinear Tobit model is $2n + 1$.

Censored Linear Regression Three-Tiered Tobit Model

Since processes causing LGD to be zeroes or ones may have a different nature compared to processes where $0 < LGD < 1$, we introduce in this section a three-tiered model. The LGD estimator is introduced in this case in the following linear form:

$$y_i^{(0)}(\vec{x}, \mathbf{r}) = \sum_{k=1}^{n} x_k \cdot r_{ik} + x_{n+2} \tag{16}$$

$$y_i^{(c)}(\vec{x}, \mathbf{r}) = \sum_{k=1}^{n} x_{k+n+2} \cdot r_{ik} + x_{2n+4} \tag{17}$$

$$y_i^{(1)}(\vec{x}, \mathbf{r}) = \sum_{k=1}^{n} x_{k+2n+4} \cdot r_{ik} + x_{3n+6} \tag{18}$$

where components of the model coefficient vector $\vec{x}$ have the following meaning:

$$\vec{x} = \begin{cases} x_{1\dots n} & \text{coefficients for the } LGD = 0 \text{ model} \\ x_{n+1} & \sigma_0 \text{ for the } LGD = 0 \text{ model} \\ x_{n+2} & \text{intercept for the } LGD = 0 \text{ model} \\ x_{(n+3)\dots(2n+2)} & \text{coefficients for modeling } 0 < LGD < 1 \\ x_{2n+3} & \sigma \text{ for the } 0 < LGD < 1 \text{ model} \\ x_{2n+4} & \text{intercept for the } 0 < LGD < 1 \text{ model} \\ x_{(2n+5)\dots(3n+4)} & \text{coefficients for the } LGD = 1 \text{ model} \\ x_{3n+5} & \sigma_1 \text{ for the } LGD = 1 \text{ model} \\ x_{3n+6} & \text{intercept for the } LGD = 1 \text{ model} \end{cases} \tag{19}$$

This LGD model is based on the following probability function for an i^{th} facility:

$$P_i(z) = \begin{cases} p_i^{(0)} & \text{if } z = 0 \\ \frac{\kappa}{\sigma} g\left(\frac{z - y_i^{(c)}}{\sigma}\right) & \text{if } 0 < z < 1 \\ p_i^{(1)} & \text{if } z = 1 \end{cases} \tag{20}$$

where

$$\begin{cases} p_i^{(0)} & = N\left(-\frac{y_i^{(0)}}{\sigma_0}\right) \\ p_i^{(1)} & = \kappa \cdot N\left(-\frac{1-y_i^{(1)}}{\sigma_1}\right) \\ \kappa & = N\left(\frac{y_j^{(0)}}{\sigma_0}\right)\left[N\left(-\frac{1-y_i^{(1)}}{\sigma_1}\right) + N\left(\frac{1-y_i^{(c)}}{\sigma}\right) - N\left(-\frac{y_i^{(c)}}{\sigma}\right)\right]^{-1} \end{cases} \tag{21}$$

Here κ is a normalization factor.

The model coefficients $\vec{x}$ can be found as a result of the maximization of the following log-likelihood function:

$$H(\vec{x}) = \sum_i \log P^i(LGD_i) \tag{22}$$

Using calibrated model coefficients one can estimate expected LGDs:

$$LGD_i = p_i^{(1)} + y_i^{(c)}\kappa\left[N\left(\frac{1-y_i^{(c)}}{\sigma}\right) - N\left(-\frac{y_i^{(c)}}{\sigma}\right)\right] + \sigma\kappa\left[N\left(\frac{y_i^{(c)}}{\sigma}\right) - N\left(\frac{1-y_i^{(c)}}{\sigma}\right)\right] \tag{23}$$

Inflated Beta Regression Model

The Inflated Beta Regression LGD model (Pereira and Cribari-Neto, 2010) is based on the following probability function for an i^{th} facility:

$$P^i(z) = \begin{cases} P_0^i & \text{if } z = 0 \\ (1 - P_0^i - P_1^i) \cdot f(z; \mu^j, \phi^i) & \text{if } 0 < z < 1 \\ P_1^i & \text{if } z = 1 \end{cases} \tag{24}$$

where

$$f(z; \mu^j, \phi^i) = \frac{\Gamma(\phi^i)}{\Gamma(\mu^i\phi^i)\Gamma((1-\mu^i)\phi^i)} z^{\mu^i\phi^i - 1}(1-z)^{(1-\mu^i)\phi^i - 1} \tag{25}$$

with $0 < \mu^i < 1$ and $\phi^i > 0$ (μ^i being the mean value).

Given P_0^i, P_1^i, μ^i, and ϕ^i, one can calculate (using (24)) probabilities $P^i(LGD_i)$ of getting LGD_i values. In order to establish a connection between explanatory variables r_k^i of debt facilities and expected facility LGDs, the following four linear predictors are introduced:

$$\log \frac{P_0^i}{1-P_0^i} = a_0 + \sum_{k=1}^{n} x_k^0 r_k^i \tag{26}$$

$$\log \frac{P_1^i}{1-P_1^i} = a_1 + \sum_{k=1}^{n} x_k^1 r_k^i \tag{27}$$

$$\log \frac{\mu^j}{1-\mu^j} = a_\mu + \sum_{k=1}^{n} x_k^\mu r_k^j \tag{28}$$

$$\log \frac{\phi^i}{1-\phi^i} = a_\phi + \sum_{k=1}^{n} x_k^\phi r_k^i \tag{29}$$

Here vectors $\vec{x}^{(0,1,\mu,\phi)}$ and $a_{(0,1,\mu,\phi)}$ are model coefficients and intercepts, respectively. The model calibration consists of finding coefficients and intercepts by maximizing the following log-likelihood (objective) function:

$$H = \sum_i \log(P^i(LGD_i)) \tag{30}$$

Using calibrated model one can estimate expected LGD for an i^{th} debt facility:

$$E[LGD]^i = \int_0^1 zP^i(z)dz = P_1^i + \frac{1-P_0^i-P_1^i}{(1+\exp(-(a_\mu + \sum_{k=1}^n x_k^\mu \cdot r_k^i)))} \tag{31}$$

The log-likelihood function (30) can be split as follows:

$$H = H_{01} + H_\beta \tag{32}$$

where $H_{01}(\vec{x}^{(0)}, \vec{x}^{(1)})$ and $H_\beta(\vec{x}^{(\mu)}, \vec{x}^{(\phi)})$ can be optimized independently:

$$\begin{aligned} H_{01} = & \sum_i^{LGD_i=0} \log(P_0(\vec{x}^{(0)})) + \sum_i^{LGD_j=1} \log(P_1(\vec{x}^{(1)})) \\ & + \sum_i^{0<LGD_i<1} \log(1-P_0(\vec{x}^{(0)}) - P_1(\vec{x}^{(1)})) \end{aligned} \tag{33}$$

$$H_\beta = \sum_i^{0<LGD_i<1} \log f(LGD_i; \vec{x}^{(\mu)}, \vec{x}^{(\phi)}) \tag{34}$$

The Inflated Beta Regression model was tested using the R-coded function developed by Yashkir Consulting.

Beta Linear Regression Model

If $0 < LGD < 1$ then $P_0 = P_1 = 0$, reducing the problem to a general Beta Regression Model (34)[1]. This model (34) can be also used if all LGDs are scaled as $LGD^* = LGD \cdot (\beta - \alpha) + \alpha$, calibration performed using LGD^* , and values of LGD^*_{est} (estimated on the basis of this calibration) are scaled back as $LGD_{est} = (LGD^*_{est} - \alpha)/(\beta - \alpha)$. The Beta Regression Model was tested using the BetaReg library function of the R statistical package.

Censored Gamma Linear Regression Model

The Censored Gamma LGD model (Sigrist and Stahel, 2010) is based on the following probability function for an i^{th} facility:

$$P_j(z;\xi,\alpha,\theta_i) = \begin{cases} \Psi(\xi,\alpha,\theta_i) & \text{if } z = 0 \\ \gamma(z+\xi,\alpha,\theta_i) & \text{if } 0 < z < 1 \\ 1 - \Psi(1+\xi,\alpha,\theta_i) & \text{if } z = 1 \end{cases} \tag{35}$$

where

$$\begin{cases} \gamma(u;\alpha,\theta_i) = \frac{1}{\theta_i \Gamma(\alpha)} u^{\alpha-1} e^{-\frac{u}{\theta_i}} & \text{(gamma distribution)} \\ \Psi(u;\alpha,\theta_i) = \int_0^u \gamma(x;\alpha,\theta_i)dx & \text{(cumulative gamma distribution)} \end{cases} \tag{36}$$

with $u > 0$, $\alpha > 0$, and $\theta_i > 0$.

Given ξ, α, and θ_i, using (35), one can calculate probabilities $P_i(LGD_i)$ of getting LGD_i values. In order to establish a connection between explanatory variables r^i_k of debt facilities and expected facility LGDs, the following linear predictors are introduced:

$$\begin{cases} \log\alpha = \alpha^* \\ \log\xi = \xi^* \\ \log\theta_i = x_0 + \sum_{k=1}^{n} x_k r^i_k \end{cases} \tag{37}$$

Here x_k are model coefficients (including the intercept x_0). The model calibration involves finding coefficients and parameters α^*, ξ^* by maximizing the following log-likelihood (objective) function:

$$H(\xi^*,\alpha^*,\vec{x}) = \sum_i \log P_i(LGD_i;\xi,\alpha,\theta_i) \tag{38}$$

Using the calibrated model one can estimate expected LGD for an i^{th} debt facility:

$$\begin{aligned} E[LGD]^i = & \int_0^1 zP_j(z;\xi,\alpha,\theta_i)dz \qquad (39) \\ = & \alpha \cdot \theta_i \left(\Psi(1+\xi,1+\alpha,\theta_i) - \Psi(\xi,1+\alpha,\theta_i)\right) \\ & + (1+\xi)\left(1 - \Psi(1+\xi,\alpha,\theta_i)\right) - \xi\left(1 - \Psi(\xi,\alpha,\theta_j)\right) \end{aligned}$$

The Censored Gamma Regression model was tested using the R-coded function developed by Yashkir Consulting.

[1] One can replace, for example, $LGDs = 0$ with $LGDs = \epsilon$ and $LGDs = 1$ with $LGDs = 1 - \epsilon$ where $\epsilon \ll 1$

3 Data, Explanatory Variables and Correlation Analysis

Data used for LGD model calibration

The data set **All Data** represents all available data in an internal or an external database used for the LGD model development and calibration. In our analysis, **All Data** is the S&P LossStats data (2011 update, 4275 cases) of defaulted facilities. Only results of analysis based on this data are presented in this paper.

The **Peaks Data** is the LGD data related to the time periods of the business cycle when the number of defaults and losses is significantly higher than the average default and losses values. We chose years 1990−1991 as the **Peak 1**, years 2001−2002 as the **Peak 2**, and years 2008−2009 as **Peak 3**. All three peaks have distinctly high levels of defaults and losses (shown in Fig.1, based on the recent report from S&P (Standard&Poor's, 2012)). During peaks of the cycle, global market and credit conditions are different from the quiet periods of the cycle, therefore, the most important predictive factors are correlated at a higher level with the historical LGD data collected for these time periods.

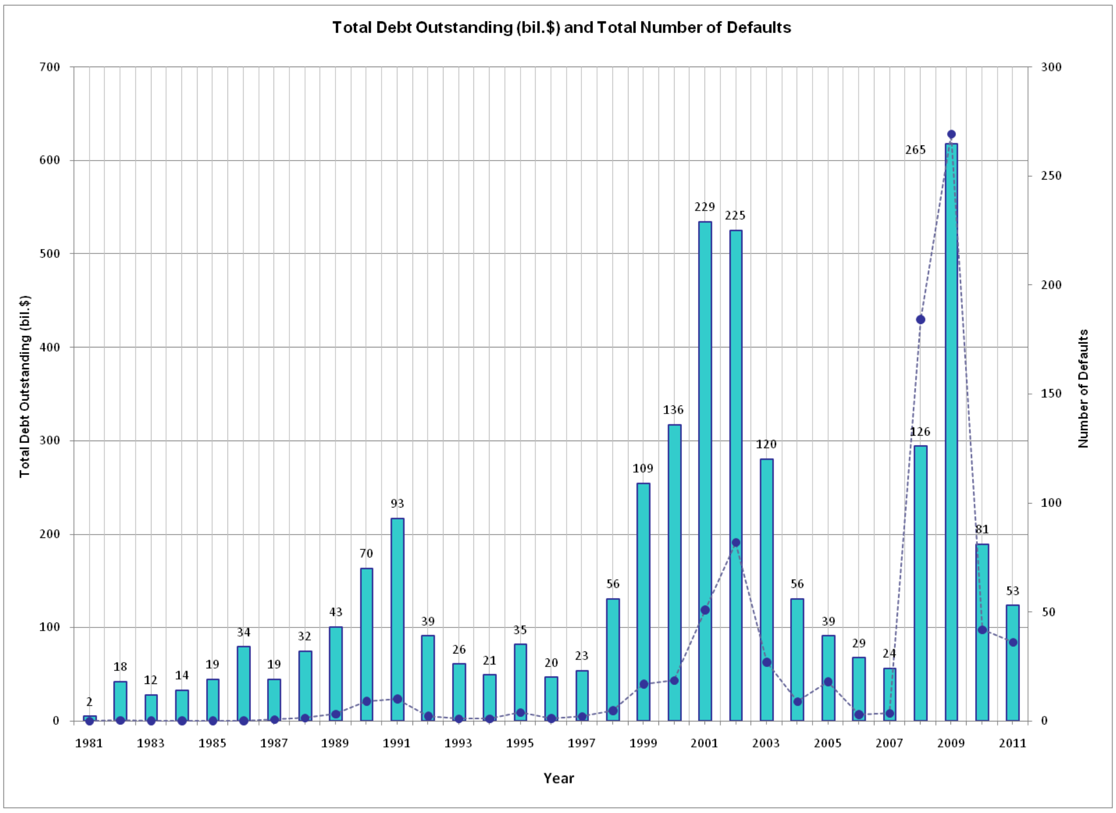

Figure 1: Total Debt Outstanding and Total Number of Defaults

Methods discussed in this paper were tested on four data sets: **All Data**, **Bankruptcy Data**, **Peaks Data**, and **Bankruptcy and Peaks Data**. **All Data** represents all available LGD data in the data set. The **Bankruptcy Data** includes only bankruptcy data cases from **All Data**. We have considered and have used separately **All Data** and **Bankruptcy Data**. The

Bankruptcy and Peaks Data represents the bankruptcy data from the peak periods.

Explanatory Variables/Factors

The explanatory variables were chosen based on how they are correlated with the LGDs based on the collected historical data. The proper choice of data and instrument types are very important for good performance of the models, therefore, the model calibration was tested for several groups based on instrument types.
The following main five factors (explanatory variables) were used: **Ranking** (defines rank in the capital structure, the more senior the instrument, the higher the recovery rate), **Debt Cushion** (amount/percentage of debt below a defaulted instrument), **Principal Above** (amount of debt above a defaulted instrument), **Effective Interest rate** (prepetition rate at the time the last coupon was paid), and **Spread**.
We introduce also three additional factors (dependence on industry, on the type of collateral, and on the facility ranking): **Industry mean LGD**, **Collateral mean LGD**, **Ranking mean LGD**.
The choice of **additional factors** makes the model dependent on industry, collateral type, and ranking, for which no numerical predictive parameters are available. These additional factors were calculated as the mean of all LGD values for a given industry, for a given collateral type, and for a given ranking. For example, **Industry Mean LGD** is the mean value of all LGDs for the cases related to a specified industry. This value is added as an additional factor to all cases belonging to the specified industry. The same was done for **Collateral Mean LGD** and **Ranking Mean LGD**. The **Collateral Mean LGD** depends on a type of the collateral and it defines the mean of all cases for this type of collateral.

Correlation Analysis

To identify factors that affect LGDs the most, their correlations were investigated and the results are presented in Table 1 and Table 2.

Table 1: Correlation between factors and LGD (**All Data**, for different instruments)

	Ranking	Debt Cushion	Principal Above	Effective Interest Rate	Spread	Industry meanLGD
Senior Unsecured Bonds	-0.056	-0.071	-0.023	**0.243**	**0.267**	**0.399**
Revolving Credit	0.061	**-0.299**	0.027	-0.016	0.035	**0.193**
Term Loan	**0.180**	**-0.347**	**0.191**	0.016	0.059	**0.208**
Sr. Subordinated Bonds	**-0.043**	**-0.143**	0.087	-0.024	0.038	**0.269**
Subordinated Bonds	-0.004	0.023	0.019	**-0.133**	**-0.144**	**0.205**
Senior Secured Bonds	**0.146**	**-0.403**	**0.202**	**0.146**	**0.173**	**0.281**
Jr. Subordinated Bonds	**0.268**	0.043	**-0.273**	**0.215**	-0.007	**0.516**
Other	**0.713**	**-0.652**	**0.589**	**0.280**	**0.545**	**0.260**
All Instruments	0.348	-0.442	0.367	0.227	0.359	0.272
	Collateral mean LGD	Ranking mean LGD	Original Amount	Principal Default Amount	Acclaimed Amount	Total Debt
Senior Unsecured Bonds	0.000	-0.078	**0.132**	0.001	**0.147**	**-0.102**
Revolving Credit	**0.188**	0.060	0.096	0.092	n/a	0.076
Term Loan	**0.187**	**0.176**	0.032	-0.009	n/a	-0.065
Sr. Subordinated Bonds	**0.106**	-0.019	0.001	0.054	0.062	-0.053
Subordinated Bonds	-0.010	0.008	**0.118**	-0.022	0.048	0.091
Senior Secured Bonds	**0.209**	**0.158**	0.027	0.017	0.066	**-0.133**
Jr. Subordinated Bonds	0.000	**0.173**	**0.337**	-0.058	**0.135**	**0.224**
Other	**0.761**	**0.617**	**0.755**	**0.604**	n/a	**0.840**
All Instruments	0.507	0.458	0.364	0.065	0.020	0.132

If all instrument types are considered together, the correlation between Instrument type mean LGDs with all LGDs is equal to 0.5103.

Table 2: **Peak Data** (years $1990 - 1991; 2001 - 2002; 2008 - 2009$), correlation between factors and LGD

	Ranking	Debt Cushion	Principal Above	Effective Int.Rate	Spread	Industry meanLGD
Senior Unsecured Bonds	-0.076	-0.006	-0.016	**0.301**	**0.315**	**0.484**
Revolving Credit	-0.032	**-0.345**	**0.105**	**-0.153**	0.071	**0.262**
Term Loan	0.050	**-0.278**	-0.044	-0.065	**0.118**	**0.326**
Sr. Subordinated Bonds	0.092	**-0.221**	-0.034	**-0.164**	-0.022	**0.141**
Subordinated Bonds	0.040	-0.038	**0.127**	**-0.119**	**-0.135**	**0.162**
Senior Secured Bonds	**0.328**	**-0.463**	-0.011	**0.211**	**0.211**	**0.255**
Jr. Subordinated Bonds	**0.235**	-0.088	**0.286**	**-0.194**	**-0.225**	**0.310**
Other	**0.664**	**-0.699**	**0.596**	**0.815**	**0.919**	**0.676**
All Instruments	0.292	-0.421	0.321	0.228	0.386	0.322
	Collateral mean LGD	Ranking mean LGD	Original Amount	Principal Df. Amount	Acclaimed Amount	Total Debt
Senior Unsecured Bonds	0.000	-0.096	0.083	0.062	**0.166**	-0.058
Revolving Credit	**0.181**	-0.035	**0.234**	**0.214**	n/a	**0.148**
Term Loan	**0.187**	0.035	0.062	0.031	n/a	-0.052
Sr. Subordinated Bonds	0.035	**0.108**	0.062	0.004	0.047	-0.034
Subordinated Bonds	0.091	0.025	-0.051	0.023	0.065	**0.140**
Senior Secured Bonds	**0.312**	**0.323**	**0.133**	0.100	**0.193**	**-0.197**
Jr.Subordinated Bonds	0.000	**0.105**	-0.070	-0.043	**0.179**	**0.142**
Other	**0.935**	**0.638**	**0.472**	**0.568**	n/a	**0.586**
All Instruments	0.441	0.298	0.059	0.053	0.140	0.012

In the case of **Peak Data**, if all instrument types are considered together, the correlation between Instrument type mean LGDs with all LGDs is equal to 0.470. The comparison of the correlation level when using **All Data** and when using **Peak Data**, demonstrates the following:

1) Correlations of historical LGDs with Industry Mean LGD (mean of historical LGDs for a specified industry) are high for all instruments (from 14% to 68%),
2) Changes in correlation level are clearly seen when comparing **All Data** correlation results and **Peak Data** correlation results. The absolute values of correlation are higher for the **Peak Data** results. For example, Revolving Credit correlation with Debt Cushion is equal to -0.299 when using **All Data**, and it is equal to -0.345 when using **Peak Data**,
3) The significance of Spread, Effective Interest rate and Total Debt factors increases during cycle peaks (as expected) due to the influence of macroeconomic conditions increasing in cycle peaks.

The correlation level analysis demonstrates that sets of significant factors (explanatory variables) are different for different instruments (Table 3). In this table, the most significant explanatory variables are marked for each instrument. They were chosen based on the criteria that absolute values of correlations exceed 10%.

Table 3: Marked cells: absolute values of correlations(factor,LGD) exceed 10%

	Ranking	Debt Cushion	Principal Above	Effective Int. Rate	Spread	Industry mean LGD	Collateral mean LGD	Ranking mean LGD
Revolving Credit								
Term Loan								
Sr.Unsec. Bonds								
Sr.Sec. Bonds								
Sr.Sub. Bonds								
Sub. Bonds								
Jr.Sub. Bonds								

The Spread Data was not always available, therefore, we did not include Spread into the calibration of models. Based on the similarity of the factor sets, there are three groups of instruments that should be calibrated together:
Group A: Term Loan and Revolving Credit,
Group B: Senior Unsecured,
Group C: Senior Secured Bonds, Senior Subordinated Bonds, Subordinated Bonds, and Junior Subordinated Bonds.
Note, that for testing purposes, we considered Senior Secured Bonds separately, but the obtained results did not show visible improvement in calibration criteria.

4 Comparative Model Analysis

Criteria for the Methodology Analysis

The Goodness-of-Fit and the model LGD Correlation were chosen as the criteria for the methodology performance analysis. As a measure of Goodness-of-Fit (G) we use the following parameter (often called "the coefficient of determination")

$$G = 1 - \frac{MSE}{varLGD} \tag{40}$$

where MSE is the mean square error (model versus historical LGD), and $varLGD$ is the variance of the input data. The interpretation of the Goodness-of-Fit parameter G is as follows. For a "naïve" model, where predicted values of the model LGD are equal to the mean historical LGD, we would have $MSE = varLGD$ and $G = 0$ (the model "fit" is not better than "naïve" model). On the other hand, if a model provides prediction such that $MSE \ll varLGD$ (ideal case) then $G \sim 1$ (a very good fit). Using MSE as a criterion of the fit quality might be misleading. In the following sections we will use the G parameter as a criterion for comparison of different models (values of MSE and/or values of mean absolute error (MAE) are presented for convenience).
The model LGD Correlation (ρ), defined as the correlation between historical and simulated LGDs, is also used for model comparison.

Calibration Details

This section describes the procedures, the R-functions used, and specific calibration approaches. Codes for Tobit, Inflated Beta, and Gamma Reg models were developed by Yashkir Consulting using the statistical R package. Additional applications were developed in Python.

Least Square method: The library function in R provides solving of the problem (1)

$$\mathbf{Q} = \mathbf{lsfit}\left(\mathbf{r}, \vec{L}, ...\right) \tag{41}$$

where $\mathbf{r}$ is the matrix of explanatory variables for a given set of defaulted cases, and $\vec{L}$ is the vector of observed LGDs. From the output object (list) $\mathbf{Q}$ we find the following: the coefficient vector $\vec{x} = Q_1$ (including intercept x_0), the array of residuals $\vec{\delta} = Q_2$, and the modelled LGD $L_i - \delta_i$ for the i^{th} case.

Object function minimization: The library function in R provides solving of the problem (14) (Tobit model):

$$\mathbf{Q} = \mathbf{optim}\left(\vec{z}, \Phi(\vec{z}), ...\right) \tag{42}$$

From the output object (list) we find the coefficient vector $\vec{z} = (\vec{x}, \sigma) = \mathbf{Q}_1$.

Maximization of the log-likelihood function: The library function in R provides solving of the problem (22) (Three-Tiered Tobit model):

$$\mathbf{Q} = \mathbf{optim}\left(\vec{x}, -H(\vec{x}), ...\right) \tag{43}$$

From the output object (list) we find the coefficient vector $\vec{x} = \mathbf{Q}_1$.

Maximization of the log-likelihood function for Beta Regression: The library function in R provides solving of the problem (34):

$$\mathbf{Q} = \mathbf{betareg}(FORMULA, link = "logit", data = DATA) \tag{44}$$

where $FORMULA$ is the following string: "$LGD \sim N_1 + N_2 + ...$" and $DATA$ is the table containing input data in the following format:

Table 4: Input data

$\mathbf{N_1}$	$\mathbf{N_2}$	...	**LGD**
V_{11}	V_{12}	...	LGD_1^*
V_{21}	V_{22}	...	LGD_2^*
...	...	...	...

where $\mathbf{N_i}$ (predictive variable names) and **LGD** are column headers; V_{ji} and LGD_j^* are corresponding numerical values for every j^{th} transaction [2]. From the output object (list) **Q** we find the following: the coefficient vector $\vec{x} = Q_1$ (including intercept x_0), and the array of residuals $\vec{\delta} = Q_2$.

The modelled LGD for i^{th} case is $LGD_i^{mod} = \frac{LGD_i^* - \delta_i - \alpha}{\beta - \alpha}$. In general, an LGD of any j^{th} transaction is estimated as

$$LGD_j = \frac{1}{1 + e^{-y_j}} \tag{45}$$

with the predictor

$$y_j = x_0 + \sum_k x_k \cdot r_{jk} \tag{46}$$

Term Loan and Revolving Credit (Group A)

Results of the methods performance are presented below for the Instrument Group A (Term Loan and Revolving Credit) [3].

Table 5: Instrument Group A (Factors: Debt Cushion , Industry mean LGD, Collateral mean LGD)

Data Model	All Data	Peak Data	Bankruptcy data	Bankruptcy peaks data
Tobit				
$G =$	0.1538	0.2242	0.1768	0.2167
$MAE =$	0.1977	0.2246	0.2123	0.2282
$\rho =$	0.3932	0.4745	0.4215	0.4666
Least Square				
$G =$	0.1658	0.2341	0.1893	0.2280
$MAE =$	0.2021	0.2262	0.2157	0.2294
$\rho =$	0.4072	0.4838	0.4349	0.4772
Inflated Beta				
$G =$	0.1568	0.2254	0.1817	0.2157
$MAE =$	0.2036	0.2325	0.2201	0.2352
$\rho =$	0.3955	0.4798	0.4274	0.4691
BetaReg				
$G =$	0.1615	0.2311	0.1857	0.2251
$MAE =$	0.2085	0.2305	0.2214	0.2337
$\rho =$	0.4062	0.4836	0.4340	0.4771
GammaReg				
$G =$	0.1537	0.2239	0.1767	0.2164
$MAE =$	0.1976	0.2245	0.2122	0.2281
$\rho =$	0.3932	0.4743	0.4215	0.4664

The best fit for these instruments was obtained using LSM and BetaReg. Two important observations can be made for the case of Term Loans and Revolving Credits:

1) The best fit for all sets of data was achieved by using Least Square Method (LSM),

[2] Values LGD_j^* are scaled values of real LGDs as follows: $LGD_j^* = LGD_j * (\beta - \alpha) + \alpha$ to ensure that $0 < \alpha \leq LGD_j^* \leq \beta < 1$.

[3] Revolving line of credit is an agreement by a bank to lend a specific amount to a borrower, and to allow that amount to be borrowed again once it has been repaid. Term loan is a bank loan to a company, with a fixed maturity and often featuring amortization of principal. If this loan is in the form of a line of credit, the funds are drawn down shortly after the agreement is signed. Otherwise, the borrower usually uses the funds from the loan soon after they become available.

2) The best fit was achieved on **Peak Data** using LSM and Beta Reg (Goodness-of-Fit is approximately 0.23 and Correlation is approximately 0.47).

In peak conditions of the cycle, the predictive power of the chosen facility parameters increases, which results in higher values of Goodness-of-Fit and Correlations for **Peak Data**. This outcome of comparative tests for different calibration models clearly indicates that success of LGD modelling depends mainly on availability (and proper choice) of explanatory variables and on data quality, but not on fitting techniques.

Senior Unsecured (Group B)

Senior Unsecured transactions (Instrument Group B) do not have any collateral. According to the correlation matrices (Table 1 and Table 2), main parameters that are highly correlated with the historical LGDs are Industry Mean LGD and the Effective Interest Rate[4]. Therefore, in case of Senior Unsecured the main factor defining the LGD level at default is the industry cluster to which the facility belongs.

[4]Effective Interest Rate, as by the S&P definition is the prepetition rate at the time the last coupon was paid; fixed rate for fixed-coupon instruments, and the floating rate used at the time of default for floating-rate instruments.

Table 6: Instrument Group B (Factors: Industry mean LGD, Effective Interest Rate (EIR))

Data Model	All Data	Peak Data	Bankruptcy Data	Bankruptcy Peaks Data
Tobit				
Industry.meanLGD				
$G =$	0.1460	0.2220	0.2146	0.2692
$\rho =$	0.3894	0.4783	0.4710	0.5297
Industry.meanLGD and EIR				
$G =$	0.1691	0.2479	0.2423	0.3021
$\rho =$	0.4180	0.5048	0.4994	0.5601
Least Square				
Industry.meanLGD				
$G =$	0.1595	0.2350	0.2295	0.2856
$\rho =$	0.3985	0.4836	0.4782	0.5332
Industry.meanLGD and EIR				
$G =$	0.1825	0.2627	0.2562	0.3208
$\rho =$	0.4264	0.5115	0.5054	0.5653
Inflated Beta				
Industry.meanLGD				
$G =$	0.1562	0.2199	0.2206	0.2658
$\rho =$	0.4000	0.4871	0.4816	0.5379
Industry.meanLGD and EIR				
$G =$	0.1775	0.2463	0.2471	0.2963
$\rho =$	0.4279	0.5161	0.5086	0.5681
BetaReg				
Industry.meanLGD				
$G =$	0.1574	0.2335	0.2275	0.2843
$\rho =$	0.3972	0.4831	0.4774	0.5334
Industry.meanLGD and EIR				
$G =$	0.1808	0.2614	0.2548	0.3193
$\rho =$	0.4255	0.5112	0.5052	0.5653
GammaReg				
Industry.meanLGD				
$G =$	0.1457	0.2216	0.2142	0.2686
$\rho =$	0.3893	0.4783	0.4709	0.5296
Industry.meanLGD and EIR				
$G =$	0.1688	0.2474	0.2418	0.3013
$\rho =$	0.4179	0.5047	0.4994	0.5600

The Senior Unsecured case is sufficiently difficult to model due to the fact that the strongest dependency is only on Industry Mean LGDs. Two important observations can be made for the case of Senior Unsecured:

1) The best fit for all sets of data is done again using Least Square Method (LSM),
2) The best fit was achieved on **Bankruptcy Peak Data** using LSM and Beta Reg (Goodness-of-Fit is approximately 0.32 and Correlation is approximately is 0.57).

The dependency on **Bankruptcy Peak Data** shows that for bankruptcy cases in peak conditions of the cycle, the industry becomes even more important. It should be noted that for this group of instruments, calibrated on **Bankruptcy Peak Data**, the Tobit and GammaReg models also provide sufficiently good fit (Goodness-of-Fit is approximately 0.30, and Correlation is approximately 0.56). For contracts with fixed interest rates (if default data contains this rate) the Effective Interest Rate can also be used for calibration.

Senior Secured, Senior Subordinated, Subordinated, and Junior Subordinated bonds (Group C)

For the Instrument Group C (Senior Secured, Senior Subordinated, Subordinated, and Junior Subordinated bonds), according to the correlation matrices (Table 1 and Table 2), the main parameters that are highly correlated with the historical LGDs are: Debt Cushion, Principal Above, Effective Interest Rate, Industry Mean LGD, Collateral Mean LGD, Ranking Mean LGD.

Table 7: Instrument Group C (Factors: Debt Cushion, Principal Above, Effective Interest Rate, Industry Mean LGD, Collateral Mean LGD, Ranking Mean LGD)

Data Model	All Data	Peak Data	Bankruptcy Data	Bankruptcy Peaks Data
Tobit				
$G =$	0.2420	0.2086	0.3818	0.3361
$\rho =$	0.4998	0.4633	0.6235	0.5858
Least Square				
$G =$	0.2520	0.2161	0.3904	0.3429
$\rho =$	0.5017	0.4633	0.6245	0.5843
Inflated Beta				
$G =$	0.2406	0.2012	0.3660	0.3192
$\rho =$	0.4995	0.4623	0.6223	0.5872
BetaReg				
$G =$	0.2500	0.2123	0.3877	0.3358
$\rho =$	0.5013	0.4605	0.6235	0.5794
GammaReg				
$G =$	0.1998	0.2141	0.2955	0.2772
$\rho =$	0.4551	0.4714	0.5515	0.5383

The Instrument Group C model strongly depends on Debt Cushion, Principal Above, Industry mean LGD, Collateral mean LGD, Ranking mean LGD. The Effective Interest Rate also can be used if available. Two important observations can be made for this case:

1) The best fit for all sets of data is done again using Least Square Method (LSM),
2) The best fit was achieved on **Bankruptcy Data** using LSM and Beta Reg (Goodness-of-Fit is approximately 0.39, and Correlation is approximately 0.63).

The dependency on chosen factors and **Bankruptcy Data** shows the importance of the proper choice of main factors. It should be noted that for this group of instruments, the Tobit model also provides sufficiently good fit (Goodness-of-Fit is approximately 0.38, and Correlation is approximately 0.62) on **Bankruptcy Data**. The results from Inflated Beta provide also sufficiently good fit (Goodness-of-Fit is approximately 0.37, and Correlation is approximately 0.62). The results show that the factors were well chosen. The inclusion of the Effective Interest Rate does not change the Goodness-of-Fit and Correlation.

5 Calibration Examples with the Best Fitting Results

The results presented in this section are the best fit as described above. The marker (***) in Tables 8, 9, 10, indicates the most important factors. Results show that chosen factors for all three groups were properly chosen and are the important factors for the simulation.

Term Loans plus Revolving Credit (Group A)

The best fit was obtained with **Peak Data** using LSM and Beta Reg. It should be noted (in addition to the criteria used) that the mean of the historical LGDs (equal to 0.247) and the mean of the simulated LGDs (equal to 0.250 for LSM, and equal to 0.258 for BetaReg) are very close. This is a supporting factor for the models' results.

Table 8: Results for the best fit for cases of Term Loans and Revolving Credit

Term Loans and Revolving Credits				
BetaReg, Peaks Data			Least Square, Peaks Data	
Parameter	Coefficients	Pr(>|z|)	Parameter	Coefficients
(Intercept)	-1.40	$<$ 2e-16 ***	(Intercept)	-0.19
Debt Cushion	-0.70	4.22e-16 ***	Debt Cushion	-0.35
Industry meanLGD	2.57	$<$ 2e-16 ***	Industry meanLGD	1.25
Collateral meanLGD	0.78	7.21e-04***	Collateral meanLGD	0.39
$G =$	0.2311		0.2341	
$MAE =$	0.2305		0.2262	

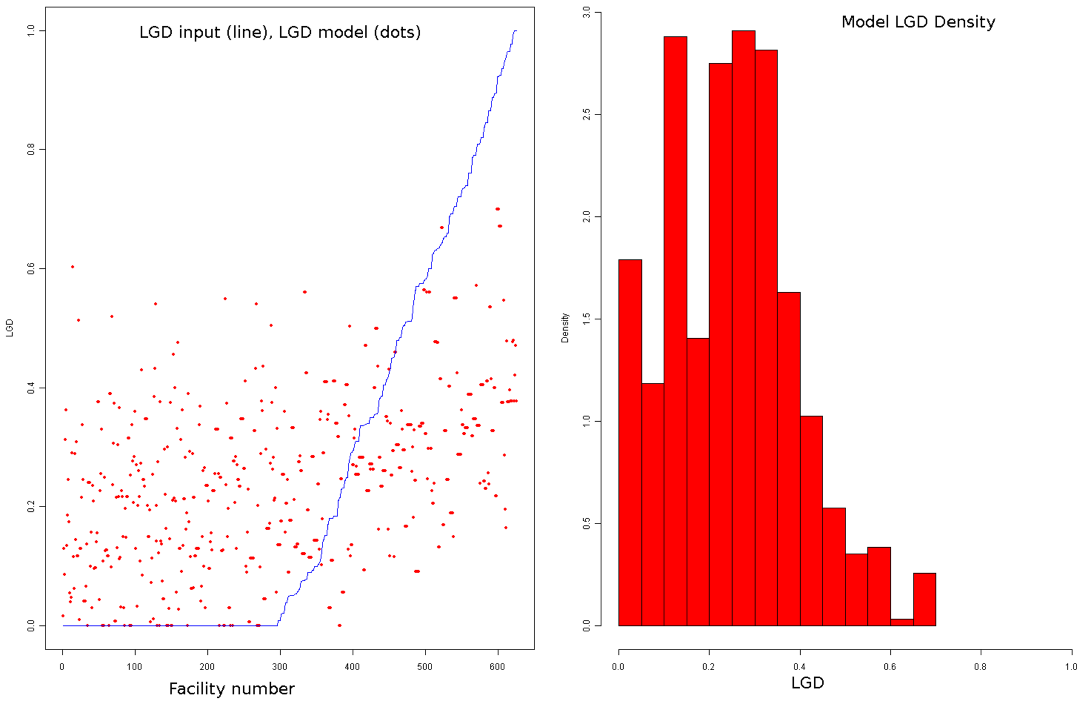

Figure 2: Modelling results for calibration on the Term Loans plus Revolving Credit data ($G = 0.2341$, $MAE = 0.2262$)

The historical LGDs and the model simulated LGDs (LSM model) are presented in Figure 2, left. In this graph the historical LGDs (solid line) and simulated LGDs (dots) are plotted as function of case numbers (all cases are sorted by their historical LGD values). In case of a perfect model the estimated LGDs (dots) would follow the historical LGDs (line). In the reality, the linear regression model used provides a lower fitting quality as seen in Figure 2, left. The advantage of this presentation of results is that one can clearly see what are limitations of chosen set of explanatory variables: LGD is overestimated for low historical LGDs and is underestimated for high historical LGDs.
Another method of visual representation of calibration results is shown in Figure 3 (estimated LGDs versus historical LGDs). The better model fitting would correspond to model LGDs concentrated

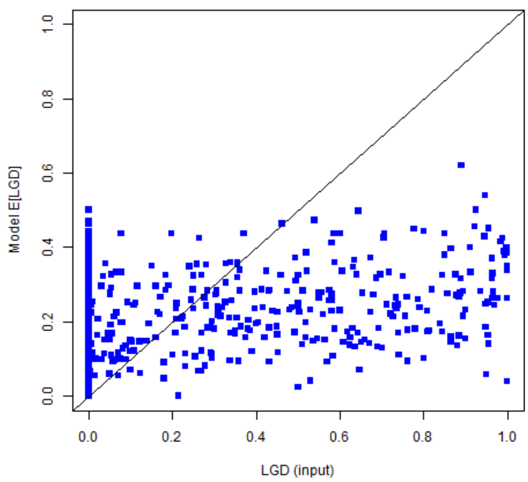

Figure 3: Model LGD versus historical LGDs (**All Data**, LSM, $G = 0.1658$, $\rho = 0.4072$)

around the diagonal line.
The histogram of the simulated LGDs is presented on Figure 2, right. The simulated LGD values are concentrated around the mean LGD value.

Senior Unsecured (Group B)

The best fit was obtained for **Bankruptcy Peak Data** using LSM and Beta Reg. In addition to the fitting quality criteria used, it is worth mentioning that the mean of the historical LGDs (equal to 0.5617) and the mean of the simulated LGDs (equal to 0.5616 for LSM, and is equal to 0.5583 for BetaReg) are very close. This is a supporting factor for the models' results.

Table 9: Results for the best fit for Senior Unsecured cases

Senior Unsecured, Bankruptcy Peak Data				
BetaReg			Least Square	
Parameter	Coefficients	Pr(>|z|)	Parameter	Coefficients
(Intercept)	-1.38	< 2e-16 ***	Intercept	-0.21
Effective.Interest.Rate	5.85	1.67e-07 ***	Effective.Interest.Rate	2.94
Industry.meanLGD	2.34	< 2e-16 ***	Industry.meanLGD	1.22
$G =$	0.3193		0.3208	
$MAE =$	0.2564		0.2507	

Limited flexibility of modelling of Senior Unsecured LGDs is due to the fact that only Industry Mean LGD has significant importance for these instruments. The chosen factors are all shown as important.

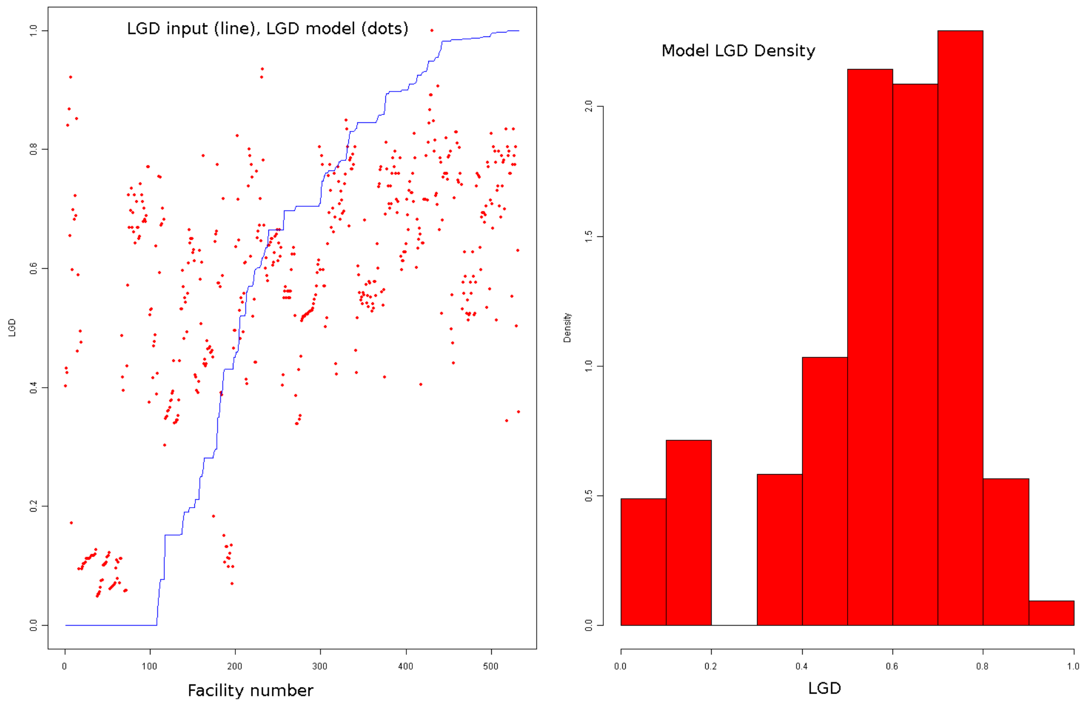

Figure 4: Fitting results for Senior Unsecured data set ($G = 0.3208$, $MAE = 0.2507$)

The historical LGDs and the model simulated LGDs (model LSM) are presented in Figure 4, left. Simulated LGD histogram is presented in Figure 4, right. There is concentration of the simulated values around the mean value as expected. Simulated LGDs (dots) reflect the general trend of historical LGDs (solid line).

Senior Secured, Senior Subordinated, Subordinated, and Junior Subordinated Bonds (Group C)

The best fit for **Bankruptcy Data** was obtained using LSM and Beta Reg models. Note that the mean of the historical LGDs (equal to 0.6178) and the mean of the simulated LGDs (equal to 0.6190 for LSM, and is equal to 0.6123 for BetaReg) are very close.

Table 10: Results for the best fit for the Instrument Group C

Bankruptcy Data				
BetaReg			Least Square	
Parameter	Coefficients	Pr(>|z|)	Parameter	Coefficients
(Intercept)	-0.98	< 2e-16 ***	Intercept	0.06
Principal Above	0.23	0.00129 **	Principal Above	0.16
Debt Cushion	-0.72	8.0e-14 ***	Debt Cushion	-0.36
Effective Interest Rate	-0.51	0.43069	Effective Interest Rate	-0.24
Industry mean LGD	1.55	3.9e-16 ***	Industry mean LGD	0.76
Collateral mean LGD	1.47	< 2e-16 ***	Collateral mean LGD	0.81
			Ranking.meanLGD	-0.19
$G =$	0.3877		0.3904	
$MAE =$	0.2381		0.2335	

All chosen factors, except EIR, are shown as important.

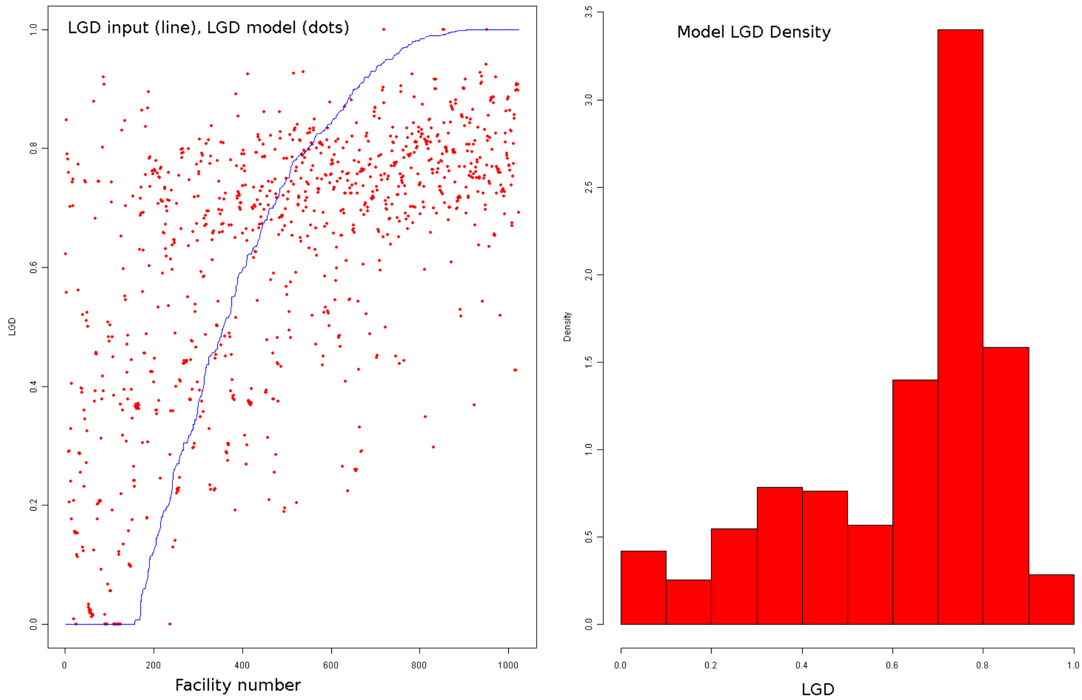

Figure 5: Fitting results for the Instrument Group C ($G = 0.3904$, $MAE = 0.2335$)

The historical LGDs (solid line) and the model simulated LGDs (dots, LSM model) are presented in Figure 5, left. The histogram of simulated LGDs is shown in Figure 5, right. The results show good agreement between the simulated and the historical LGDs. The "cloud" of simulated values follows the historical LGDs.

Summary of Calibration Examples on All Data set
The results of our tests for the instrument groups A, B and C based on the **All Data** set are presented in Table 11:

Table 11: Summary of Calibration Examples on the **All Data** set

	Root Mean Square Error	Mean Absolute Error	Goodness-of-Fit	Sample size	Number of variables
Group A	0.2633	0.2021	0.1658	1605	3
Group B	0.3265	0.2785	0.1825	1175	2
Group C	0.3381	0.2852	0.2520	1275	6
(Yang and Tkachenko, 2012) Mixture	0.3273	0.2355	n/a	500	8

In the last row of Table 11, the results of the LGD "Mixture" model test by Yang and Tkachenko (Yang and Tkachenko, 2012) are presented for comparison. In their tests, the authors explored several models (Logit raw, HL logit, Logit, Least-squares logit, Naive Bayes, Mixture, and Neural Network) and found that the "Mixture" model has the lowest fitting error. The results of our tests provide a similar or better level of fitting errors.
Finally, as an example, we compare calibration results for groups A and C with a "naïve" model where the "model" LGDs are defined as historical average LGD.

Table 12: Comparison of the linear regression model results with "naïve" model

	Mean Absolute Error	G
Group A		
Linear Regression	0.2021	0.1658
"Naïve" model	0.2365	0.0006
Group C		
Linear Regression	0.2852	0.2520
"Naïve" model	0.3549	0.0008

Results presented in Table 12 demonstrate clearly advantages of using the Goodness-of-Fit parameter G as a criterion for model comparison. For example, in case of the group A the MAE changes from 23.65% to 20.21% only, but the Goodness-of Fit changes dramatically from 0.06% to 16.58%.

6 Data Sensitivity and Stress Testing

Data Sensitivity Test

The example for the data set of Senior Unsecured cases was tested to demonstrate the sensitivity of results to adjustments of the initial data. If the initial historical data was adjusted, for example, by excluding all cases where LGDs are lower that 15% (taken into account that LGDs for this instrument, in general, could not be low). After performing the model calibration (**betaReg**) and the simulation of the LGDs we obtained the following results (see Figure 6). The fit is better than for the case described for the general Senior Unsecured case (Correlation increased from 0.52 to 0.54). This example shows that even small filtering of the data made based on the reasonable business assumptions could improve calibration results even for difficult cases such as Senior Unsecured.

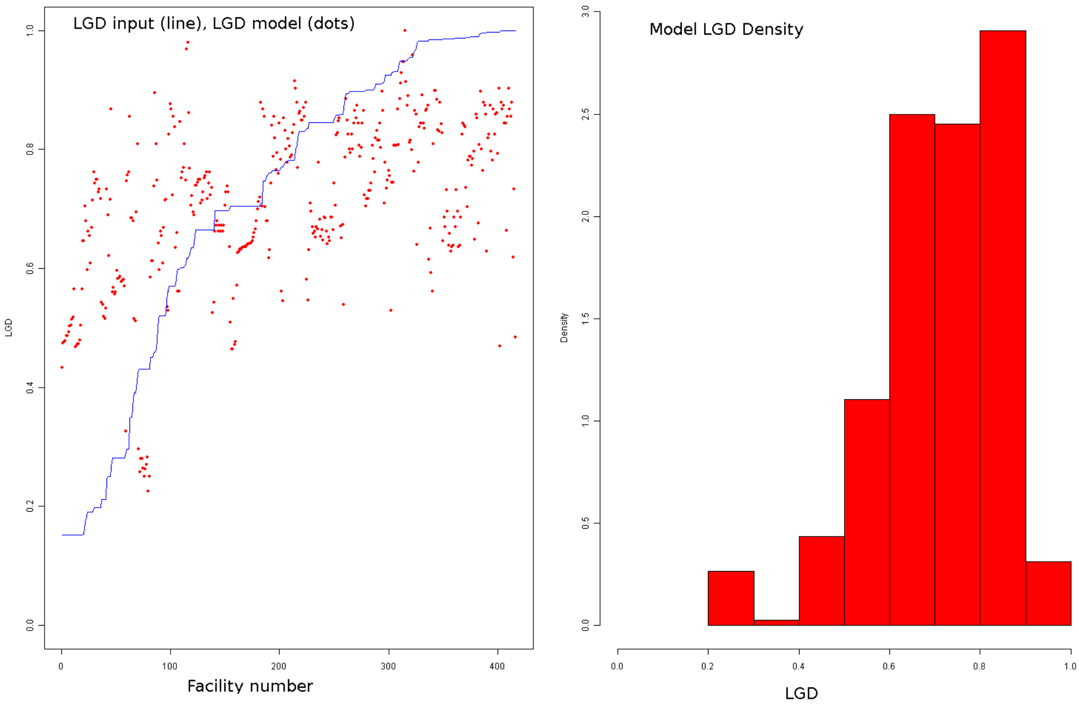

Figure 6: Results for Senior Unsecured with adjusted data, Bankruptcy Peak initial data ($G = 0.2963$, $MAE = 0.1796$)

Stress Testing for LGDs

The approach of the LGD stress testing comes naturally from the results of our analysis of models and data. The stress test procedure is as follows:

1) Derive model coefficients for peaks periods using the **Peaks Data** and/or data for each peak separately. These coefficients emphasize the peak of crisis period in the the business cycle conditions.
2) Run simulations of **All Data** LGDs using these peak related coefficients.

The resulting simulated LGDs provide predictions of the LGD levels in crisis (stress) conditions. The simulation results with **Bankruptcy Data** coefficients are presented on Figure 7.

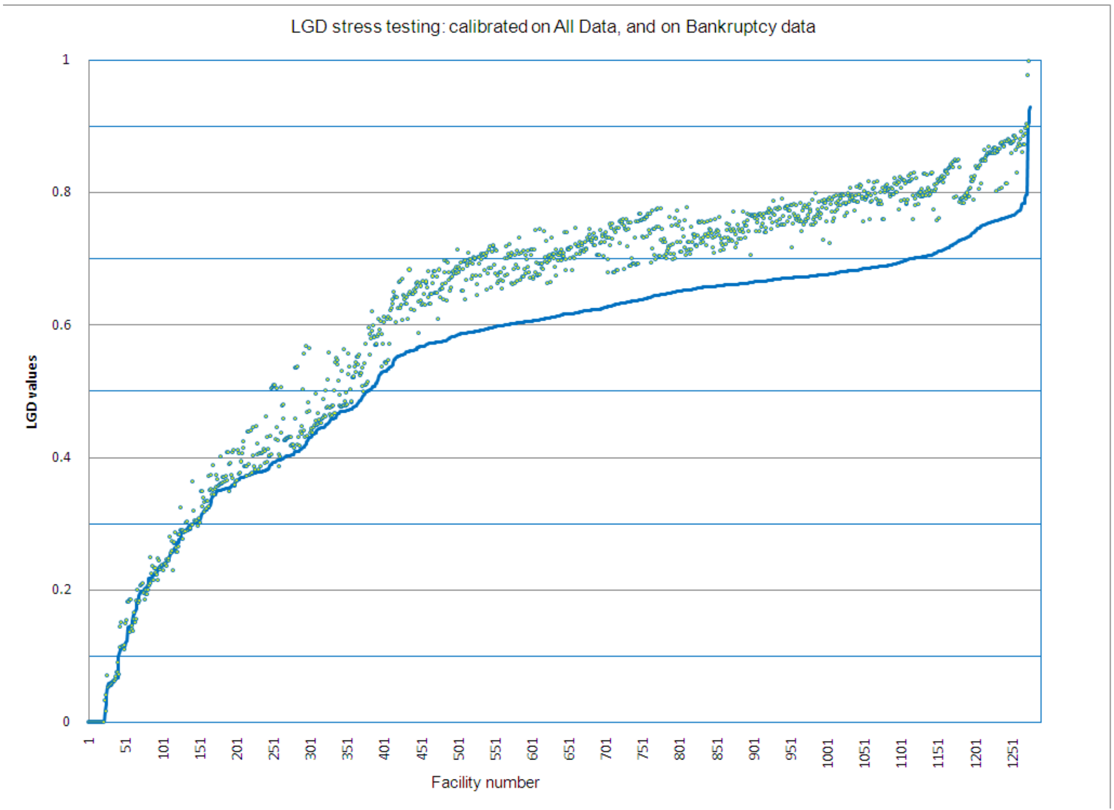

Figure 7: Stresstesting of LGDs: **All Data** calibration (solid line), **Bankruptcy Data** calibration (dots)

The simulation results for **Bankruptcy Data** coefficients and **Bankruptcy Peaks Data** coefficients are presented on Figure 8. The mean of simulated LGDs, based on **All Data** coefficients, is equal to 0.55; mean of simulated LGDs, based on **Bankruptcy Data** coefficients, is equal to 0.62; mean of simulated LGDs, based on **Bankruptcy Peaks Data** coefficients, is equal to approximately 0.66. Therefore, the average LGD increase, compared with the **All Data** LGD level, is equal to 7% (calibration on **Bankruptcy Data**), and it is equal to 11% (calibration on **Bankruptcy Peak Data**). The **Bankruptcy Data** results and **Bankruptcy Peak Data** results are shown on Figure 8.

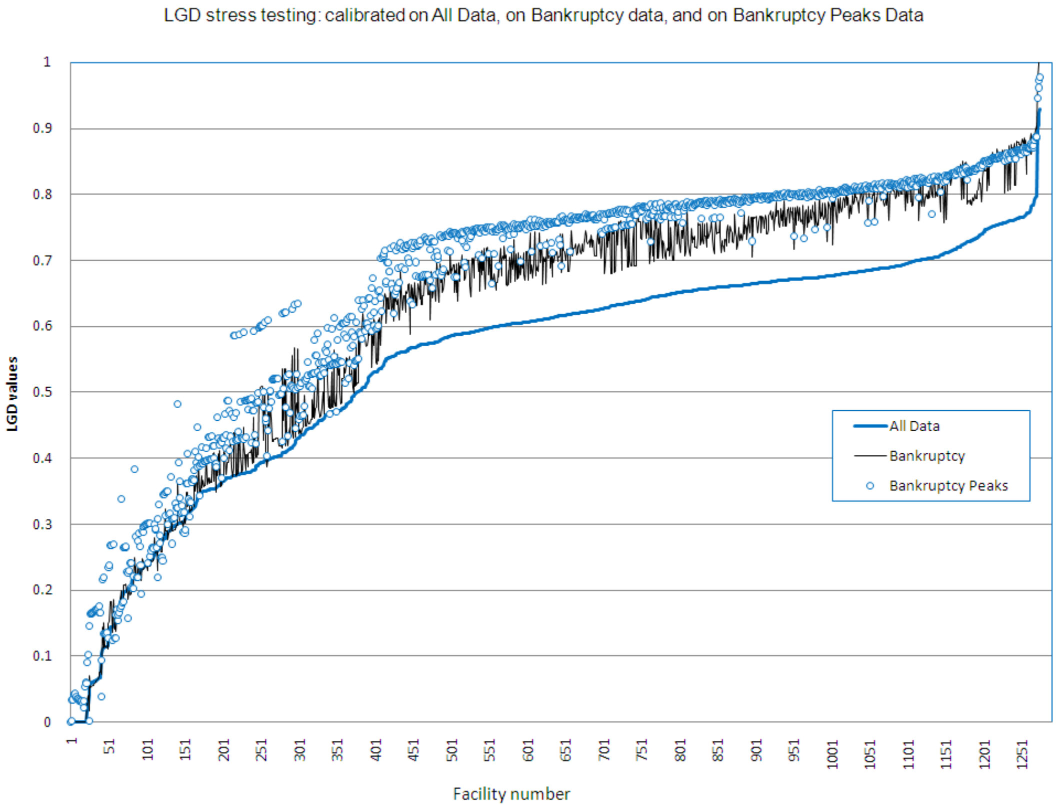

Figure 8: Stresstesting of LGDs. The **All Data** calibration (solid line), **Bankruptcy Data** calibration (solid thin line), **Bankruptcy Peaks Data** calibration (open dots).

Our approach for the estimation of downturn LGDs does not require any additional model assumptions such as analytic approach by (Barco, 2007) or the parameter sensitivity approach by (Rosch and Scheule, 2007). The approach naturally follows the chosen model calibration procedure and the data choice.The downturn LGDs are estimated based on the chosen data subset consistent with the downturn conditions in the business cycle. If a financial institution does not have enough data for **Peaks Data** set, then the external data for peaks periods can be used (following specific Basel II regulations). The external data that contains all available cases provide the data for the peak/stress LGD calibration.

7 Conclusions

Several most popular LGD models (LSM, Tobit, Three-Tiered Tobit, Beta Regression, Inflated Beta Regression, Censored Gamma Regression) were tested on real data in order to compare their performance. Model factors were chosen based on the amplitude of their correlation with historical LGDs of the calibration data set. Numerical values of non-quantitative parameters (industry, ranking, type of collateral) were introduced as their LGD averages. It is shown that:

- 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

- Different debt instruments depend on different sets of model factors (from three factors for Revolving Credit or for Subordinated Bonds to eight factors for Senior Secured Bonds)
- Calibration of LGD models using distressed business cycle periods provide better fit than the data from total available time span
- Calibration parameters obtained using distress business cycle periods can be productively used for stress testing.

Calibration algorithms and details of their realization using the R statistical package are presented. The results of this study can be of use to risk managers concerned with the Basel accord compliance.

8 References

Edward Altman. Are historically based default and recovery models in the high-yield and distressed debt markets still relevant in today's credit environment? *Stern School of Business*, 2006a. Special Report.

Edward Altman. Default recovery rates and lgd in credit risk modeling and practice. an updated review of the literature and empirical evidence. *Working paper*, 2006b.

Edward Altman and Egon Kalotay. A flexible approach to modeling ultimate recoveries on defaulted loans and bonds. 2010.

Edward Altman, Brooks Brady, Andrea Resti, and Andrea Sironi. The link between default and recovery rates: Theory, empirical evidence, implications. *Working Paper S-CDM-04-07*, 2003. Series Credit and Debt Markets Research Group.

Edward Altman, Andrea Resti, and Andrea Sironi. Default recovery rates in credit risk modeling: A review of the literature, empirical evidence. *Economic Notes*, 33:183–208, 2004.

Benjamin Bade, Daniel Rosch, and Harald Scheule. Empirical performance of loss given default prediction models. *The Journal of Risk Model Validation*, 5(2):25–44, 2011.

J. Samuel Baixauli and Susanna Alvarez. The role of market-implied severity modeling for credit var. *ANNALS OF ECONOMICS AND FINANCE*, 11(2):337–353, 2010.

Michael Barco. Going downturn. *Risk*, 20(9):39–44, 2007.

Tony Bellotti and Jonathan Crook. Loss given default models incorporating macroeconomic variables for credit cards. *International Journal of Forecasting, Special Section 2: Credit Risk Modelling and Forecasting*, 28(1):171–182, January-March 2012.

Radovan Chalupka and Juraj Kopecsni. Modeling bank loan LGD of corporate, sme segments. a case study. *Czech Journal of Economics and Finance*, 59(4):360–382, 2009.

Craig Friedman and Sven Sandow. Ultimate recoveries. *RISK*, 16(8):69–73, 2003.

Greg M. Gupton. Advancing loss given default prediction models: How the quiet have quickened. *Economic Notes by Banca Monte dei Paschi di Siena SpA*, 34(2):185–230, 2005.

Greg M. Gupton and Roger M. Stein. Losscalc: Model for predicting loss given default (LGD). *Federal Reserve Bank of New York (working paper)*, 2002.

Marc Gurtler and Martin Hibbeln. Pitfalls in modeling loss given default of bank loans. *Working Paper, Technische Universität Braunschweig*, 2011.

Martin Hillebrand. Modeling and estimating dependent loss given default. *CRC working paper*, 2006.

Stefan Hlawatsch and Sebastian Ostrowski. Simulation and estimation of loss given default. *The Journal of Credit Risk*, 7(3):39–73, 2011.

Xinzheng Huang and Cornelis W. Oosterlee. Generalized beta regression models for random loss given default. *The Journal of Credit Risk*, 7(4):45–70, 2012.

Michael Jacobs Jr. A two-factor structural model of ultimate loss-given-default: Capital structure and calibration to corporate recovery data. *The Journal of Financial Transformation*, 31(4): 31–43, 2011.

David Li, Ruchi Bhariok, Sean Keenan, and Stefano Santilli. Validation techniques and performance metrics for loss given default models. *The Journal of Risk Model Validation*, 3(3):3–26, 2009.

Xiaolin Luo and Pavel V. Shevchenko. LGD credit risk model: estimation of capital with parameter uncertainty using mcmc. *Working Paper, CSIRO Mathematics, Informatics and Statistics*, 2010.

John F. McDonald and Robert A. Moffitt. The uses of tobit analysis. *Review of Economics and Statistics*, 62(2):318–321, 1980.

Tarciana L. Pereira and Francisko Cribari-Neto. A test for a correct model specification in inflated beta regressions. *Working Paper, Instituto de Matemática, Estatística e Computação Científica Universidade Estadual de Campinas*, 2010.

Daniel Rosch and Harald Scheule. Stress-testing credit risk parameters: an application to retail loan portfolios. *Journal of Risk Model Validation*, 1(1):55–75, 2007.

Til Schuermann. What do we know about loss given default? *Federal Reserve Bank of New York (working paper)*, 2004.

Fabio Sigrist and Werner A. Stahel. Censored gamma regression models for limited dependent variables with an application to loss given default. *In 28th European Meeting of Statisticians, August 17–22, 2010, Piraeus, Greece*, 2010.

Fabio Sigrist and Werner A. Stahel. Using the censored gamma distribution for modeling fractional response variables with an application to loss given default. *ASTIN Bulletin*, 41(2), 2011.

Standard&Poor's. Default, transition, and recovery: 2011 annual global corporate default study and rating transitions. 2012. Report.

Wemke van der Weija and Marcel den Hollandera. Improving PD and LGD models: following the changes in the market. *SNS Reaal*, 2009. Utrecht.

Bill Huajian Yang and Mykola Tkachenko. Modeling exposure at default and loss given default: empirical approaches and technical implementation. *The Journal of Credit Risk*, 8(2):81–102, 2012.