Application LGD Model Development

A Case Study for a Leading CEE Bank

Stefan Stoyanov, Solutions Manager, Experian Decision Analytics

Credit Scoring and Credit Control XI Conference

Edinburgh, 26th-28th of August 2009

Introduction

- The presentation discusses three alternative LGD application scorecard development methodologies applied to a Private Individuals Car Leasing portfolio of a leading bank operating in Central and Eastern Europe
- The presentation is divided in two sections:
  - The first section is focused on the sampling methodology, LGD calculations and treatment of censored observations (incomplete workouts)
  - The second part of the presentation describes the LGD modelling and validation methodologies
- The presentation concludes with a comparison between the three alternative LGD models discussing the pros and cons of each approach, including a recommendation for the final model selection

Advantages of developing an application LGD model

- The combination of PD and LGD scorecards helps banks optimise their decision making process with regards to lending
- With the traditional scoring systems, the clients are graded based on their probability of default. A LGD application scorecard allows the estimation of the second parameter that characterises loan risk - the loss given default or LGD at origination stage
- Hence, the LGD scoring model can add a new dimension to the risk rating, by classifying the clients in terms of the expected final loss rate in case of default
- For example, for two applicants with the same PD but different LGDs the final decision should therefore be based on the Expected Loss Rate:

$$ELR = PD * LGD$$

- Hence, a better assessment of the credit risk at the time of application can be achieved by combining two statistical models (PD and LGD)

Agenda

- Development Sample Selection
- LGD Calculation
- LGD Transformations
- Calibration of the Logistic Regression Scores to Estimated LGD
- LGD Model Validation Tests
- LGD Models Comparison

Default window for LGD application model development
Methodological approach

▪For sample definition 12 months of fixed default observation period from application date were used in order to link the PD and LGD model estimates since they will be combined in the capital calculation

Risk-Based Capital Standards: US Advanced Capital Adequacy Framework—Basel II; Final Rule (December 2007)
Par. Economic Loss and Post-Default Extensions of Credit
"LGD is an estimate of the economic loss that would be incurred on an exposure, relative to the exposure's EAD, if the obligor were to default within a one-year horizon during economic downturn conditions"

Development sample
Definitions and selection criteria

Default observation period

- It is the period during which the defaults are observed
- A fixed 12 months observation period was used for this development (12 months from application date)
- Only the defaults which occurred during this period were included in the model development sample

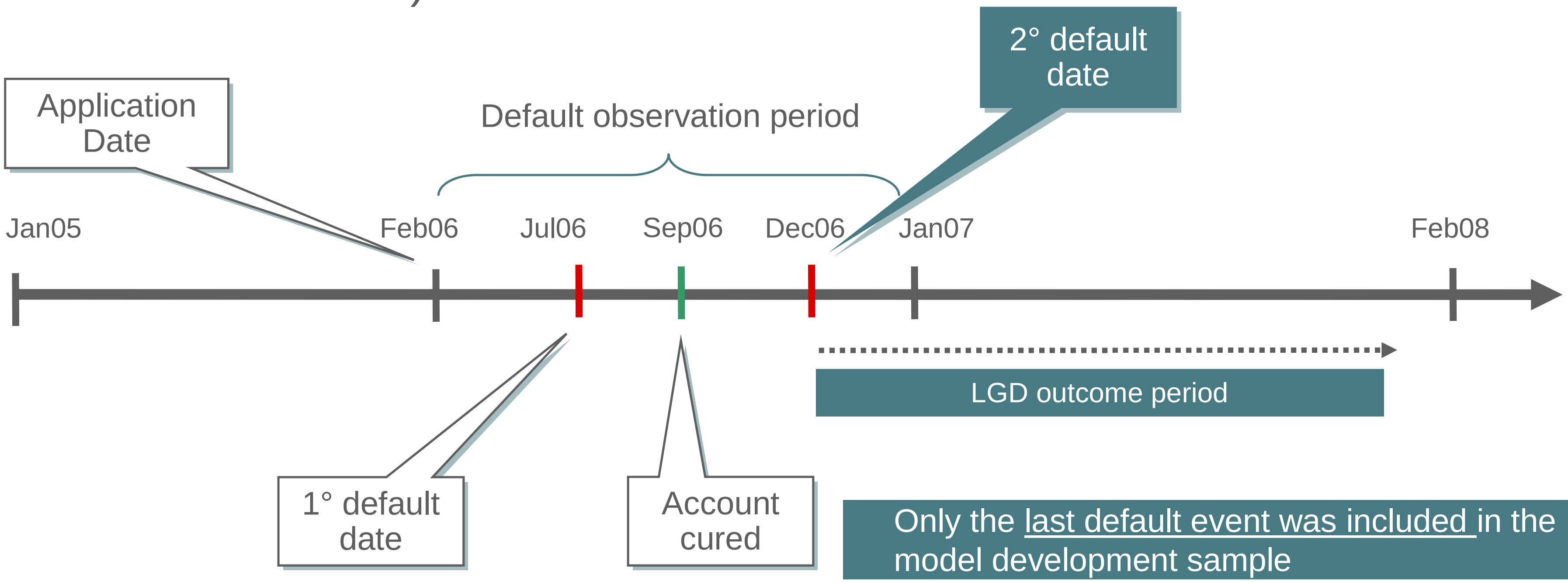
Outcome period

- It is the period during which the LGD is observed (LGD outcome period)
- A maximum of 12 months of outcome period were used depending on whether the account was cured, closed or still in collections

Development sample
Accounts with multiple defaults

- If an account defaulted more than once during the default observation period only the last default event was included in the development sample

- Consequently, the LGD was measured at such event (EAD, costs and recoveries related to it)



Only the last default event was included in the model development sample

Agenda

- Development Sample Selection
- LGD Calculation
- LGD Transformations
- Calibration of the Logistic Regression Scores to Estimated LGD
- LGD Model Validation Tests
- LGD Models Comparison

LGD components – car leasing portfolio

Exposure at default (EAD)

- Exposure at Default = Outstanding balance at time of default, corresponding to:

  Principal Amount + Arrears Amount

Recoveries

- Cash recoveries
- Non-cash recoveries (repossession value)

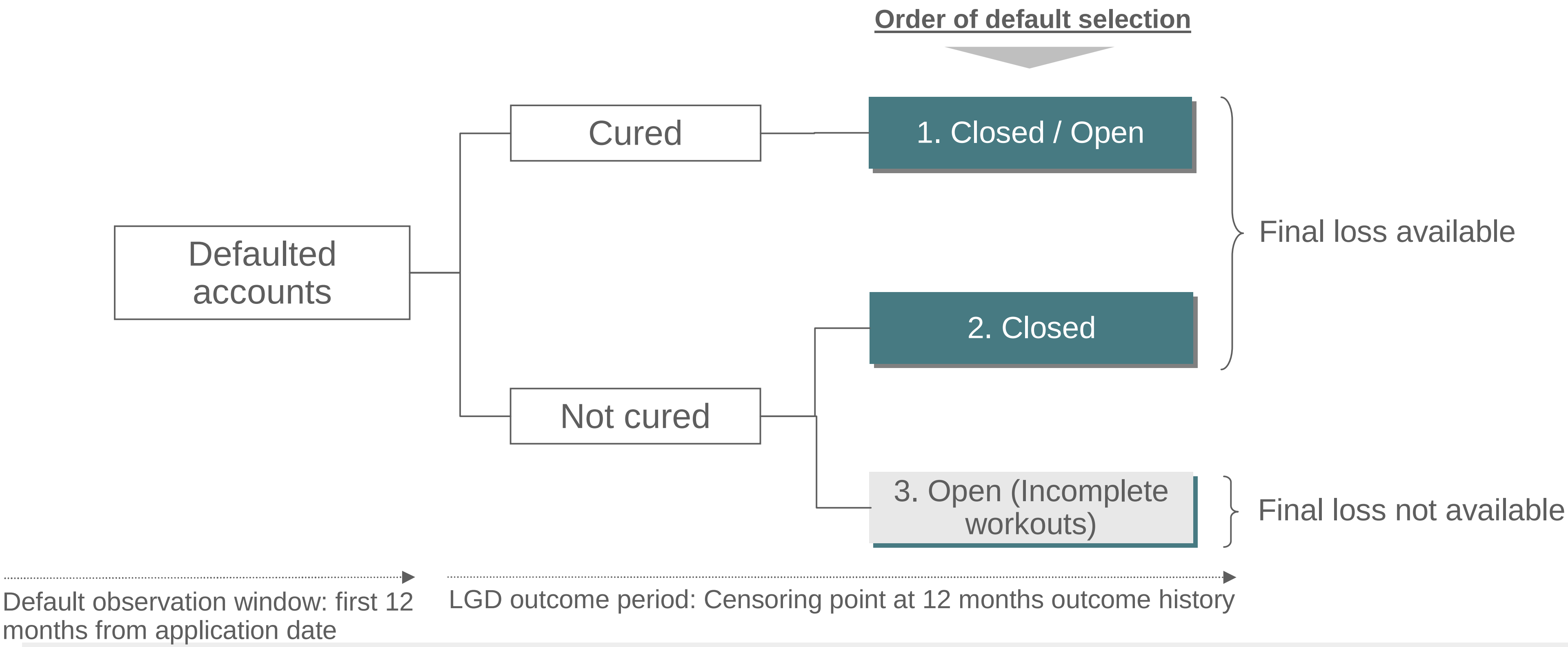
Internal and external collection costs

- Internal costs = direct and indirect costs
- External collection costs = costs for external collection agencies

Annual Percentage Rate was used as a discount factor in the calculation of the net present values of the recoveries and costs

LGD calculation
Default categories

Three categories of defaulted accounts were selected based on the criteria for LGD calculation :



LGD calculation
Modelling approach in case of incomplete recovery history

- For the open not cured accounts only incomplete recovery history is available. However, if such accounts are excluded from the model development sample it would introduce bias in the LGD estimates
- The firms can consider the incomplete workouts in their LGD estimates but estimations about future costs and recoveries are not allowed (*BKI – Supervisory regulations. Title II – Chapter 1 Par. 2.3*)
- It would be wrong to ignore the incomplete workouts in the LGD estimates *(FSA Staff Paper: Own estimates of Loss Given Default (Nov 2005) - Cures and Incomplete Workouts)*

Modelling approach

Considering the high level of sophistication and complexity of the models that could be used to estimate the final LGD for incomplete workouts compared to the provided benefits in terms of estimation accuracy, it was decided to use a censoring point at 12 months for such accounts without attempting to forecast their final LGD using statistical models

Summary of the LGD calculation approach

LGD calculation: The censoring point at 12 months of recovery history was used for all types of accounts to avoid bias in the development sample, since accounts with more than 12 months of recovery history would have more time to recover

Cured accounts

- Accounts in default during the observation period and then cured
- LGD observation period is between default date and date out of default or the censoring point at 12 months (whichever occurs first)
- The <u>current balance at time out of default</u> has to be included in the recoveries

Not cured and closed accounts

- Accounts always in default from default date to closure date (written-off or repossessed or sold accounts)
- LGD was calculated using collections information from default date to closure date or the censoring point at 12 months (whichever occurs first)

Not cured and still open accounts

- Accounts that go in default and remain in collections until the end of the 12-month LGD outcome period
- LGD was calculated using collections information from default date to the censoring point at 12 months (12 months of recovery history)

Agenda

- Development Sample Selection
- LGD Calculation
- LGD Transformations
- Calibration of the Logistic Regression Scores to Estimated LGD
- LGD Model Validation Tests
- LGD Models Comparison

LGD transformations

Both linear and logistic regression models were built in order to determine which of them provides the best estimate of the observed LGD.

- **Linear regression** - continuous dependent variable
  - In order to meet the statistical requirements of the OLS regression an attempt was made to transform the LGD to normality or at least symmetry by using Box – Cox type of transformation function
- **Logistic regression** – the LGD was transformed to a binary dependent variable using two methods:
  - **Uniform random number**: if LGD > random number then LGD Binary = 1 (Bads) ; else LGD Binary = 0 (Goods)

An example of using uniform random numbers for assigning binary outcome

LGD	RN	LGD Binary
0.024092	0.888345209	0
0.024264	0.007115099	1
0.243758	0.272714317	0

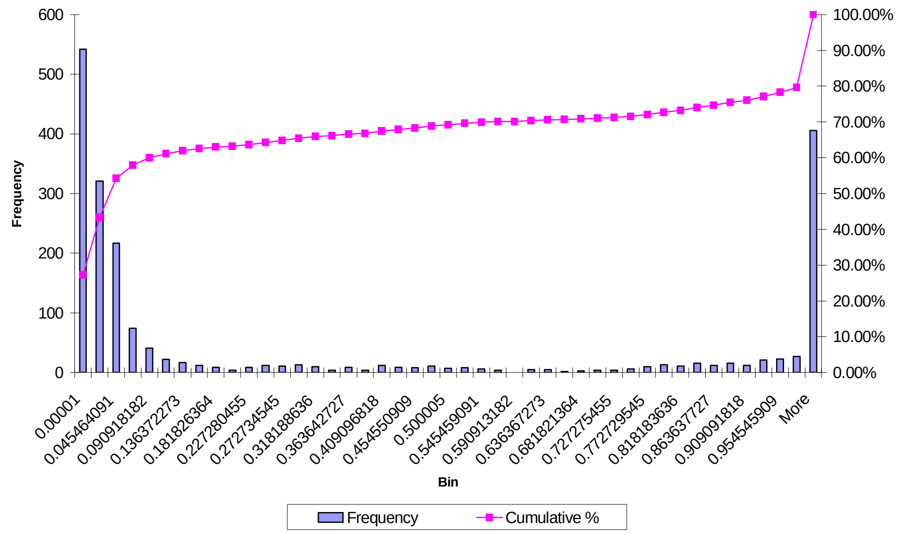
  - **Manual Cut-Off**: if LGD > 0.2 then LGD Binary = 1 (Bads) ; else LGD Binary = 0 (Goods)

An example of using manual cut-off for assigning binary outcome

LGD	Manual Cut-Off	LGD Binary
0.024092	0.2	0
0.024264	0.2	0
0.243758	0.2	1

Power transformation of the LGD distribution

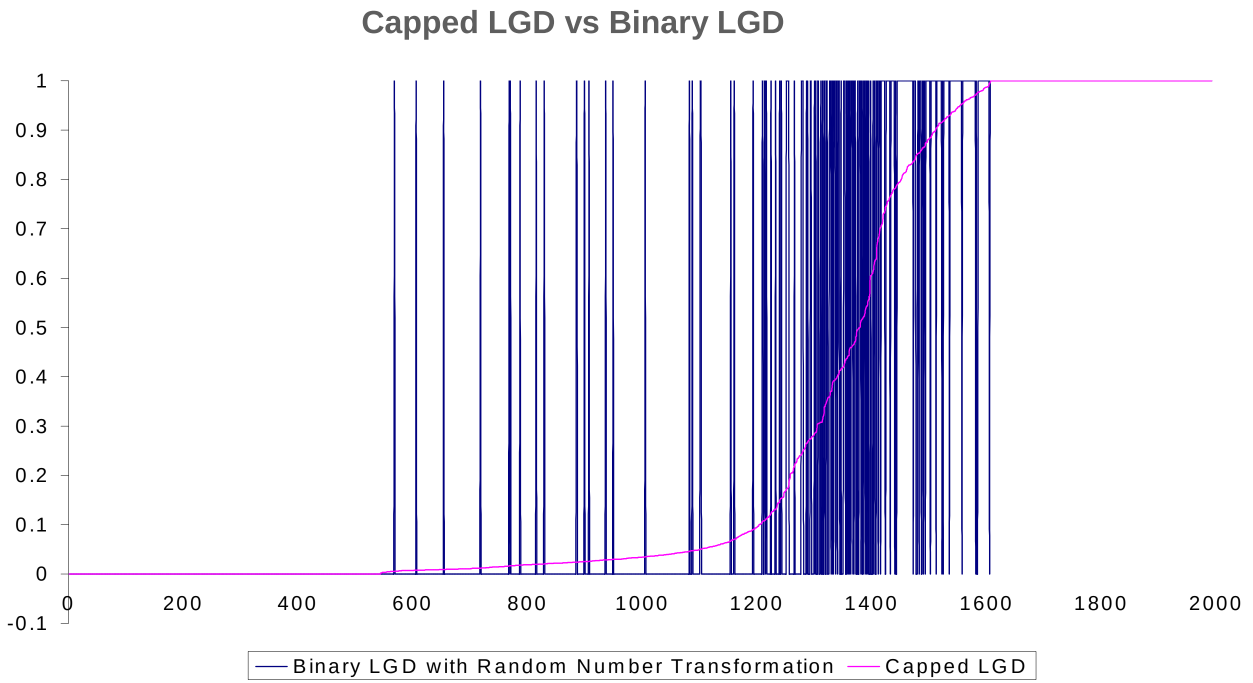
- Usually the LGD has highly non-normal distribution, often with an U shape and spikes at the two tails of the distribution
- However, the basic properties of the least squares regression do not require normality
- Non-normality does not affect the estimation of the regression parameters. The least squares estimates are still BLUE (best linear unbiased estimates) if the other regression assumptions are met
- Non-normality affects the tests of significance and the confidence interval estimates of the regression parameters



A Box-Cox type of transformation towards normality could not improve the symmetry of the distribution of the LGD, as there were large spikes at 0 and 100 while the middle range of the distribution was almost flat

Binary transformation of the LGD using uniform random numbers

- As expected, the majority of the records where the LGD was transformed to Binary =1(Bads) were at the higher end of the LGD distribution
- The random number procedure resulted in a sufficient number of Bads in the development sample in order to build robust statistical model using logistic regression



Binary transformation of the LGD using uniform random numbers - Simulations(1/4)

Simulations Step 1: Fit actual LGD distribution to theoretical Beta distribution

Theoretical Beta distribution parameters:

The estimation of the alpha (a) and beta (b) parameters of the theoretical beta distribution was based on the first two moments of the observed LGD distribution:

$$a = \frac{\mathrm{E}[Y]}{\mathrm{var}(Y)}\left(\mathrm{E}[Y](1-\mathrm{E}[Y]) - \mathrm{var}[Y]\right)$$

$$b = \frac{1-\mathrm{E}[Y]}{\mathrm{var}(Y)}\left(\mathrm{E}[Y](1-\mathrm{E}[Y]) - \mathrm{var}[Y]\right)$$

a = 0.070

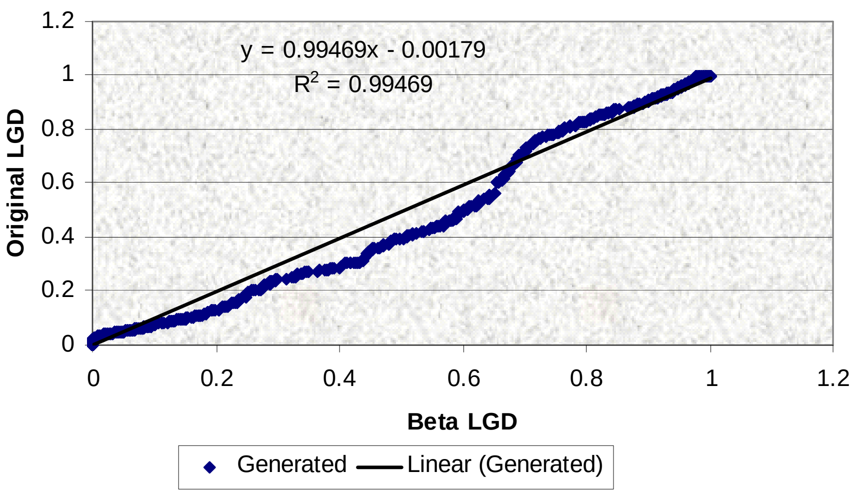
b = 0.147

Goodness of fit:

Actual LGD is plotted against the theoretical Beta distribution

$R^2 = 0.99469$

Original v. Beta distribution LGD



Binary transformation of the LGD using uniform random numbers - Simulations(2/4)

Simulations Step 2: Test the binary transformation bias

Estimate and test the binary transformation bias:

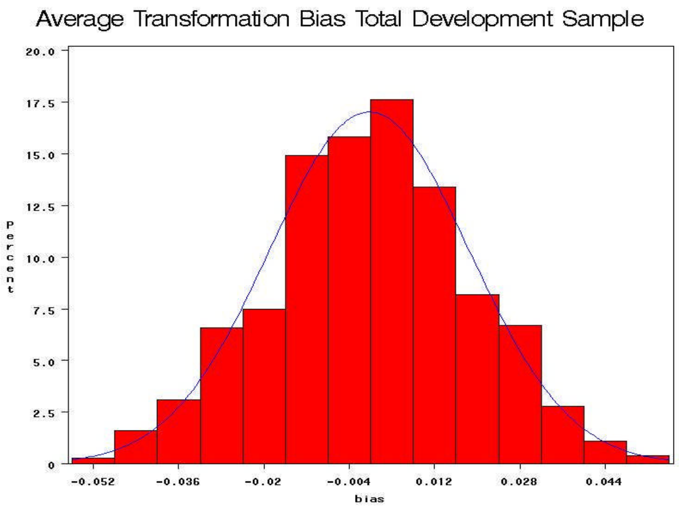
- Generate large number of samples with beta distributions. The samples are with size equal to the development sample size. The beta distributions have the parameters estimated in the previous step
- For each sample transform the theoretical LGD to binary variable using uniform random numbers
- For each sample calculate the transformation bias as:

$$Bi = \text{Mean (Beta) / P(Binary = 1)} - 1$$

- Calculate the mean value Mean (B) and the variance Var (B) for the transformation bias across all samples
- Test the H_0 (B=0) against H_1(B<>0) with a t-test

Binary transformation of the LGD using uniform random numbers - Simulations(3/4)

▪For the total model development sample the t-test results indicated that the null hypothesis that the average transformation bias is zero could not be rejected at the usual confidence levels



t-test results: t-value -0.40, p-value 0.6861

Based on the evidence from the simulation results it could be concluded that the average transformation bias for the total development sample caused by the transformation of the actual LGD to binary variable using uniform random numbers was negligibly small.

Binary transformation of the LGD using uniform random numbers - Simulations(4/4)

- Simulations were also used to investigate the average transformation bias per LGD band
- The t-tests results indicated that for the majority of the LGD bands the null hypothesis that the average transformation bias is zero could not be rejected at the usual confidence levels

LGD Scoreband	LGD Range	t-value	p-value	t-test results
1	0- 0.0199	11.89	0	reject Ho
2	0.02-0.069	1.92	0.055	can not reject Ho
3	0.07-0.119	-1.19	0.237	can not reject Ho
4	0.120-0.269	1.16	0.249	can not reject Ho
5	0.270-0.319	-0.19	0.849	can not reject Ho
6	0.320-0.369	-0.41	0.679	can not reject Ho
7	0.370-0.419	0.17	0.864	can not reject Ho
8	0.420-0.769	-0.29	0.768	can not reject Ho
9	0.770-0.819	-1.52	0.128	can not reject Ho
10	0.820-0.869	-0.65	0.518	can not reject Ho
11	0.870-0.919	0.7	0.482	can not reject Ho
12	0.920-High	1.04	0.298	can not reject Ho

Based on the evidence from the simulation results it could be concluded that for the majority of the LGD bands the average transformation bias caused by the transformation of the actual LGD to binary variable using uniform random numbers was negligibly small.

Binary transformation of the LGD using manual cut-off

- The manual cut-off which was used was 20% LGD. Hence, if LGD > 0.2 then LGD Binary = 1 (Bads) ; else LGD Binary = 0 (Goods)

- The transformation of the LGD to binary variable using manual cut-off resulted in a higher number of Bads compared to the uniform random numbers procedure

- After the manual cut-off transformation the average LGD in the model development sample was 36.90% which was higher than the observed average LGD of 32.37%

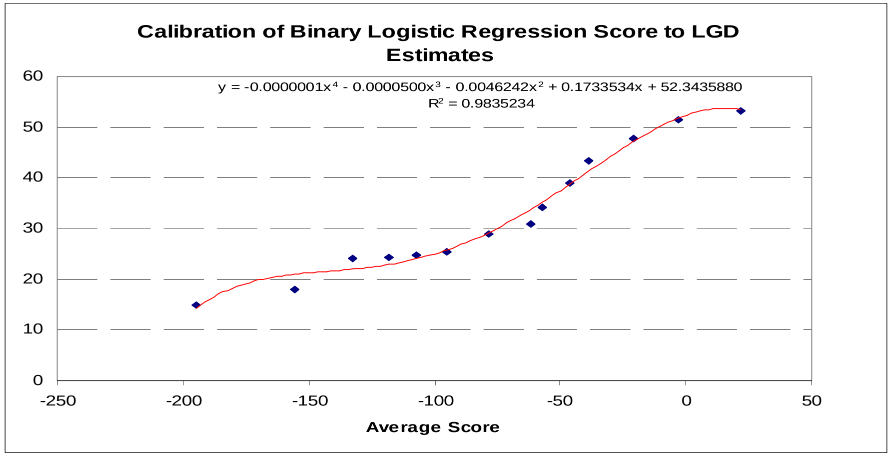
- After the uniform random numbers transformation the average LGD in the model development sample was 32.93% which was very close to the observed average LGD of 32.37%

Agenda

- Development Sample Selection
- LGD Calculation
- LGD Transformations
- Calibration of the Logistic Regression Scores to Estimated LGD
- LGD Model Validation Tests
- LGD Models Comparison

Functional calibration of the logistic regression scores to estimated LGD

- The OLS regression model provides direct LGD estimates whereas the logistic regression models provide indirect LGD estimates. Hence, it is necessary to calibrate the logistic regression scores to direct LGD estimates in order to be able to compare the two types of models
- After obtaining the functional relationship between the logistic regression scores and LGD it is possible to assign an estimated LGD to each individual score



Absolute average calibration bias: 0.11%

Agenda

- Development Sample Selection
- LGD Calculation
- LGD Transformations
- Calibration of the Logistic Regression Scores to Estimated LGD
- LGD Model Validation Tests
- LGD Models Comparison

LGD model validation tests

- Due to the small sample size all data were used for model development. The models were validated using bootstrap techniques
- The following calibration tests were used to validate the LGD models:
  - Spearman's Rank Correlation
  - Mean Squared Error
  - R-square
  - Wilcoxon Signed-Rank Test
  - Continuous Gini
- The results from the 5 statistical tests indicated similar performance of the logistic and linear LGD regression models in terms of LGD estimation accuracy
- The bootstrap validation results also indicated stable performance of the linear and logistic regression models

Agenda

- Development Sample Selection
- LGD Calculation
- LGD Transformations
- Calibration of the Logistic Regression Scores to Estimated LGD
- LGD Model Validation Tests
- LGD Models Comparison

LGD models comparison

LGD model	Pros	Cons
Continuous LGD after capping (0-100)	**General** • Direct estimation of the LGD in case no transformation towards normality of the LGD is required • Non-normality does not affect the estimation of the regression parameters. The least squares estimates are still BLUE (best linear unbiased estimates) if the other regression assumptions are met	**General** • Difficult to transform bimodal or skewed distributions towards normality before applying a linear regression model • Back transformation required in case of normality transformation and it can introduce bias • Non-normality affects the tests of significance and the confidence interval estimates of the regression parameters. In general, the probability levels associated with the tests of significance or the confidence coefficients will not be correct
Uniform random numbers LGD transformation	**General** • The bias after transformation of the LGD is small. Average transformed LGD similar to average continuous LGD • Objective criterion for LGD transformation	**General** • Back transformation/calibration necessary for direct LGD estimation • The calibration of logistic regression scores to direct LGD estimates could introduce calibration bias
Manual Cut-Off LGD transformation	**General** • Easy to implement and understand	**General** • Subjective criterion for LGD transformation • Larger bias after transformation (average transformed LGD different from average continuous LGD) • Back transformation/calibration necessary for direct LGD estimation • The calibration of logistic regression scores to direct LGD estimates could introduce calibration bias

Experian™

Summary

- The three scorecards included the same set of risk drivers of LGD. The similarity of the risk drivers indicated that little information was lost during the transformation of the LGD to binary variable
- EDA recommended the stepwise logistic regression model with uniform random numbers transformation of LGD to binary variable for the following reasons:
    - The bias introduced by the uniform random numbers transformation was negligible compared to the bias introduced by the manual cut-off transformation
    - The logistic regression model with uniform random numbers transformation had more characteristics and hence better granularity of the score especially when compared to the linear regression model
    - The logistic regression model was more robust compared to the linear regression model as the shape of the LGD distribution did not affect the statistical properties of the model parameters

Questions and Answers

Q&A

Application LGD Model Development

A Case Study for a Leading CEE Bank

Experian™
Decision Analytics