Modelling Bank Loan LGD of Corporate and SME Segment

Radovan Chalupka, Juraj Kopecsni
Charles University, Prague
1. introduction
2. key issues of LGD
3. discount rate
4. modelling LGD
5. data & risk drivers
6. methodology
7. results & conclusions
New Basel Accord

- better adjust regulatory capital with the underlying risk in a bank’s credit portfolio
- it allows banks to compute their regulatory capital in two ways:
  - using a standardized approach – regulatory ratings for risk weighting assets
  - using an own internal rating based (IRB) approach
  - IRB is based on three key parameters PD, LGD and EAD
  - LGD – the credit loss incurred if an obligor of the bank defaults
  - a move from the Foundation to the Advanced IRB approach
  - what a bank knows about LGD
motivation and contribution

• first contribution – proposition of a methodology for the advance IRB approach
• few studies have focused on the bank loans
• banks not yet developed LGD on the loan - costs, discount factors, downturn aspect, regulatory requirement
• analysis of cash flows recovery over time
• understand timing of distressed loans recoveries
• increase workout process efficiency – lower LGD

• second contribution – empirical study on a set of micro data
• propose three statistical modeling technique
• estimation of LGD based on own historical data
• identifying determinants of loan losses
• monitoring / analysis/ prevention
• several ways how to measure predictive performance
<table>
<thead>
<tr>
<th>introduction</th>
<th>key issues of LGD</th>
<th>discount rate</th>
<th>modelling LGD</th>
</tr>
</thead>
<tbody>
<tr>
<td>data &amp; risk drivers</td>
<td>methodology</td>
<td>results &amp; conclusions</td>
<td></td>
</tr>
</tbody>
</table>

1. introduction
2. key issues of LGD
3. discount rate
4. modelling LGD
5. data & risk drivers
6. methodology
7. results & conclusions
default definition (BIS)
- the obligor is unlikely to pay its credit obligations
- the obligor is past due more than 90 days on any material credit obligation

measurement of LGD
- LGD is the ratio of losses to exposure at default
- three type of losses
  - the loss of principal
  - the carrying costs of non-performing loans (interest income)
  - workout expenses (collections)
three ways of measuring LGD

1. **market LGD** – market prices of defaulted bonds
2. **workout LGD** – estimated cash flows resulting from the workout process, properly discounted, estimated exposure
3. **implied market LGD** – derived from risky but not defaulted bond prices using APM

**workout LGD**

- timing of the cash flows from the distressed asset
- cash flows should be discounted
- the correct rate would be for an asset of similar risk
- average LGD for a portfolio
  - price-weighting
  - default weighting
  - time-weighting
economic loss

- Material discount effects, direct and indirect costs associate with collection of the exposure

- Direct costs – external fees, cost of selling assets, cost of running a business
- Available for each default case

- Indirect costs – intensive care, workout department costs
- Related to the aggregate amount of exposure or aggregate recovery amount or to the number of defaults in a given period
1. introduction
2. key issues of LGD
3. **discount rate**
4. modelling LGD
5. data & risk drivers
6. methodology
7. results & conclusions
choice of a discount rate

- to calculate LGD for a particular client ex-post realized cash-flows have to be discounted back to the time of default

- a pre-default required rate \((k)\) (contract rate) can be split into
  - nominal risk-free rate \((r_f)\)
  - default premium \((\delta_{dp})\)
  - risk-premium \((\delta_{rp})\)

- assuming a loan with single cash-flow (full repayment) in one year, the present value equals

\[
PV = \frac{\$100}{1 + k} = \frac{\$100}{1 + r_f + \delta_{rp} + \delta_{dp}} = \frac{\$100}{1 + r_f + \delta_{rp}} \times (1 - \pi) + \frac{rr \times \$100}{1 + r_f + \delta_{rp}} \pi
\]

- where \(\pi\) is the probability of default, and \(rr\) is a recovery rate

\[
k = r_f + \delta_{dp} + \delta_{rp} \approx r_f + \pi(1 - rr) + \delta_{rp}
\]
choice of a discount rate (Maclachlan 2004)

- **original contractual rate**
  - it reflects the opportunity costs of losing future payments, but \( \delta_{rp} \) changes, inflation changes, and \( \delta_{dp} \) should not be used to discount already reduced cash flows

- **lender’s cost of equity**
  - sum of \( r_f \) and \( \delta_{rp} \), typically defined as one number averaging risk of all bank’s assets

- **ex-post defaulted bond and loan returns**
  - it reflects how market values defaulted bank loans, however, limited timeseries of data

- **systematic asset risk class**
  - loans are divided into groups based on the type of collateral and risk premium is assigned based on systematic risk of the asset as derived from CAPM model
risk premiums above $r_f$

- flat premiums 0–940 bps were applied, the 940 bps premium follows Brady et al. 2007
- increasing LGD by 100 bps lead approximately to the same increase in LGD
- relatively small effect compared to studies is due to relatively short sample and high losses in the beginning

asset classes – 5 levels of discount premiums

- 0 bps – cash collateral
- 240 bps – residential real estate and land
- 420 bps – movables and receivables
- 600 bps – commercial real estate, shares
- 990 bps – guaranties, promisory notes
- applying these premiums is equivalent to 5% flat premium
applying asset classes discount factor

The effect of discount factor on LGD

Without discount factor  With discount factor
1. introduction
2. key issues of LGD
3. discount rate
4. modelling LGD
5. data & risk drivers
6. methodology
7. results & conclusions
bimodality

- LGD tends to have a bimodal distribution instead of normal distribution
- bimodality means that at most of the cases there are recovery close to 100% (full repayment) or there are no recoveries at all (bankruptcy)
- makes parametric modeling of recovery difficult and proposes a non-parametric approach Renault and Scaillet (2004)

seniority and collateral

- bank loans are at the top of the capital structure
- recovery rate tend to be higher (and LGD tend to be lower) when the claim is secured by collateral with high rating than in the not secured case
business cycles

- there is strong evidence that recoveries in recessions are lower than during expansion according to Carey (1998) and Frye (2000) using US data

industry

- other studies by Grossman et al. (2001) and Acharya et al. (2003) show that industry also matters, Altman and Kishore (1996) received results that some industries such as utilities (70%) do better than other (for example manufacturing, 42%)

size of the loan

- Hurt and Felsovalyi (1998) show that large loan default exhibiting lower recovery rates.
1. introduction
2. key issues of LGD
3. discount rate
4. modelling LGD
5. data & risk drivers
6. methodology
7. results & conclusions
data sample

- based on all available files, historical closed files in 1995-2004 years and non-closed cases
- to enhance the dataset those non-closed files whose recovery period is longer than 12 quarters of a workout process are included
- Subsample A – with longer than 1 year workout period
- Subsample B - observations with very short workout period (less than a year), because these most likely represents special cases that are different from normal workout process cases (“technical defaults” or frauds)
bimodal distribution

- we use LGD grades proposed by Moody’s:
  - LGD1 \(0\% = \text{LGD} < 10\%\)
  - LGD2 \(10\% = \text{LGD} < 30\%\)
  - LGD3 \(30\% = \text{LGD} < 50\%\)
  - LGD4 \(50\% = \text{LGD} < 70\%\)
  - LGD5 \(70\% = \text{LGD} < 90\%\)
  - LGD6 \(90\% = \text{LGD} < 100\%\)
explanatory variables

• counterparty related factors
  o industry classification, age of the company, year of default, year of loan origination, length of business connection

• contract related factors
  o type of the contract, exposure at default, interest rate on the loan, number of different type of contracts

• collateral related factors
  o collateral type, collateral value by type and aggregate collateral value, collateral value relative to the EAD, collateral value as a percentage of aggregate collateral value, number of collaterals, diversification (number of different collaterals)
### Recovery rate determinants

<table>
<thead>
<tr>
<th>Counterparty related factors</th>
<th>Type</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of a counterparty</td>
<td>Continuous</td>
<td>Positive</td>
</tr>
<tr>
<td>Length of business relationship</td>
<td>Continuous</td>
<td>?</td>
</tr>
<tr>
<td>Year of default before 1995</td>
<td>Dummy</td>
<td>Negative</td>
</tr>
<tr>
<td>Year of loan origination before 1995</td>
<td>Dummy</td>
<td>Negative</td>
</tr>
<tr>
<td>New industries</td>
<td>Dummy</td>
<td>?</td>
</tr>
<tr>
<td>Industry not specified</td>
<td>Dummy</td>
<td>?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contract related factors</th>
<th>Type</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure at default</td>
<td>Continuous</td>
<td>Negative</td>
</tr>
<tr>
<td>Number of loans</td>
<td>Categorical</td>
<td>?</td>
</tr>
<tr>
<td>Investment type of loan</td>
<td>Dummy</td>
<td>?</td>
</tr>
<tr>
<td>Overdraft type of loan</td>
<td>Dummy</td>
<td>?</td>
</tr>
<tr>
<td>Revolving type of loan</td>
<td>Dummy</td>
<td>?</td>
</tr>
<tr>
<td>Purpose type of loan</td>
<td>Dummy</td>
<td>?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Collateral related factors</th>
<th>Type</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collateral value of A relative to EAD</td>
<td>Continuous</td>
<td>Positive</td>
</tr>
<tr>
<td>Collateral value of B relative to EAD</td>
<td>Continuous</td>
<td>Positive</td>
</tr>
<tr>
<td>Collateral value of C relative to EAD</td>
<td>Continuous</td>
<td>Positive</td>
</tr>
<tr>
<td>Collateral value of D relative to EAD</td>
<td>Continuous</td>
<td>Positive</td>
</tr>
<tr>
<td>Number of different collaterals</td>
<td>Categorical</td>
<td>Positive</td>
</tr>
</tbody>
</table>
explanatory variables

- we have used 4 collateral type classes based on the risk aspect of the collateral, the same classes as used in the calculation of discount rate
  - Class A: low risk – cash, land and residential real estate
  - Class B: lower average risk – movables and receivables
  - Class C: upper average risk – commercial real estate
  - Class D: high risk – securities and guarantees
explanatory variables

• we grouped industry groups into fewer categories based on these two classifications

<table>
<thead>
<tr>
<th>Standard Industry Codes (SIC)</th>
<th>Alternative industry classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Agriculture, Forestry, And Fishing</td>
<td>A Aviation and Transport Services</td>
</tr>
<tr>
<td>B Mining</td>
<td>B Business Services</td>
</tr>
<tr>
<td>C Construction</td>
<td>C Consumer Business</td>
</tr>
<tr>
<td>D Manufacturing</td>
<td>D Energy and Resources</td>
</tr>
<tr>
<td>E Transportation, Communications, Electric, Gas, And Sanitary Services</td>
<td>E Financial Services</td>
</tr>
<tr>
<td>F Wholesale Trade</td>
<td>F Life Sciences and Health Care</td>
</tr>
<tr>
<td>G Retail Trade</td>
<td>G Manufacturing</td>
</tr>
<tr>
<td>H Finance, Insurance, And Real Estate</td>
<td>H Public Sector</td>
</tr>
<tr>
<td>I Services</td>
<td>I Real Estate</td>
</tr>
<tr>
<td>J Administration</td>
<td>J Technology, Media and Telecommunications</td>
</tr>
</tbody>
</table>

• we “compressed” the alternative industry classification even further by having only two groups, the first one containing the “new industries” (Financial Services, Life Sciences and Health Care, Technology, Media and Telecommunications and Business and Consumer Services) and the rest being the “traditional industries”

• macroeconomic factors were not analyzed, because the dataset is relatively short
1. introduction
2. key issues of LGD
3. discount rate
4. modelling LGD
5. data & risk drivers
6. methodology
7. results & conclusions
multivariate analysis
Generalized linear models

- Models with fractional responses using quasi-maximum likelihood estimator
- Models with fractional responses using beta inflated distribution
- Models with ordinal responses

functions

- Symmetric Logit link
  \[ G(\alpha + \beta'x) = \frac{\exp(\alpha + \beta'x)}{1 + \exp(\alpha + \beta'x)} \]

- Asymmetric Log-log link
  \[ G(\alpha + \beta'x) = e^{-e^{-\alpha - \beta'x}} \]
Beta inflated distribution

\[
f_y(y | \mu, \sigma, \nu, \tau) = \begin{cases} 
  p_0 & \text{if } y = 0 \\
  (1 - p_0 - p_1) \dfrac{1}{B(\alpha, \beta)} y^{a-1}(1 - y)^{b-1} & 0 < y < 1 \\
  p_1 & \text{if } y = 1 
\end{cases}
\]

for \(0 \leq y \leq 1\), where \(\alpha = \mu(1 - \sigma^2) / \sigma^2\), \(\beta = (1 - \mu)(1 - \sigma^2) / \sigma^2\), \(p_0 = \nu(1 + \nu + \tau)^{-1}\), \(p_1 = \tau(1 + \nu + \tau)^{-1}\) so \(\alpha > 0, \beta > 0\), \(0 < p_0 < 1, 0 < p_1 < 1 - p_0\).
**ordinal responses - cumulative logit model**

\[
\text{logit}[P(Y \leq j | x)] = \log \frac{P(Y \leq j | x)}{1 - P(Y \leq j | x)} = \log \frac{\pi_1(x) + \ldots + \pi_j(x)}{\pi_{j+1}(x) + \ldots + \pi_J(x)}, \quad j = 1, \ldots, J - 1
\]

- each cumulative logit uses all \( J \) response categories, a model for \( \text{logit}[P(Y \leq j)] \) alone is an ordinary logit model for a binary response in which categories 1 to \( j \) form one outcome and categories \( j + 1 \) to \( J \) form the second, a model that simultaneously uses all cumulative logits is

\[
\text{logit}[P(Y \leq j | x)] = \alpha_j + \beta'x, \quad j = 1, \ldots, J - 1
\]

- each cumulative logit has its own intercept, the \( \{\alpha_j\} \) are increasing in \( j \), since \( P(Y \leq j | x) \) increases in \( j \) for fixed \( x \), and the logit is an increasing function of this probability, this model has the same effects \( \beta \) for each logit
ordinal responses – compl. log-log link models

- cumulative logit models use the logit link, as in univariate GLMs, other link functions are possible, an underlying extreme value distribution for Y implies a model of the form

\[ \log\{-\log[1 - P(Y \leq j | x)]\} = \alpha_j + \beta'x, \quad j = 1, \ldots, J - 1 \]

- this complementary log-log link has the property

\[ P(Y \leq j | x_1) = [P(Y \leq j | x_2)]^{\exp[\beta'(x_1 - x_2)]} \]

- with this link, \( P(Y \leq j) \) approaches 1 at a faster rate than it approaches 0, the related log/log link \( \log\{-\log[P(Y \leq j)]\} \) is appropriate when the complementary log-log link holds for the categories listed in reverse order

- these models are useful when we expect variables to have asymmetric effect on a response variable
selecting the appropriate model

• in order to select the most appropriate model, some commonly used procedures were followed
  o continuous variables were plotted for each LGD grade against the value to get “a feel” of the underlying relationship
  o categorical variables were tabulated to form an expectation of a potential relationship
  o the frequency table gives information whether there are enough counts for each cell to estimate reliably the effect
  o univariate regression using cumulative logit model was performed to see the effect of each variable independent of the other effects
  o then all potentially plausible variables were put together in the regression model
  o afterwards non-significant variables were gradually eliminated from the model based on the lowest t-statistic
selecting the appropriate model

- In order to select the most appropriate model, some commonly used procedures were followed
  - Univariate regression using cumulative logit model was performed to see the effect of each variable independent of the other effects
  - Then all potentially plausible variables were put together in the regression model
  - Afterwards non-significant variables were gradually eliminated from the model based on (backward elimination) on Akaike (AIC) and Schwarz information criteria (SIC)
  - Worm plot for residuals and QQ-plots were utilised to have a visual indication of normality of residuals
  - Normality of residuals was tested by Shapiro-Wilk normality test
model evaluation: back-testing

• process of assessing the model predictive power by using historical data:
  
  ➢ In sample back-testing
  ➢ Out of sample back-testing
  ➢ Out of time back-testing

• significant differences between the observed and predicted values indicate that the model is not robust (over fitting) with regard to changes over time or sample

• back-testing is a subject of data availability
1. introduction
2. key issues of LGD
3. discount rate
4. modelling LGD
5. data & risk drivers
6. methodology
7. results & conclusions
models with fractional responses using quasi-maximum likelihood estimator applying log-log link function

<table>
<thead>
<tr>
<th>Recovery rate determinants</th>
<th>Subsample A (&gt;1 year)</th>
<th>Subsample B (&lt;1 year)</th>
<th>Whole sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Value</td>
<td>Std. error</td>
<td>P-value</td>
</tr>
<tr>
<td>Exposure at default –EAD</td>
<td>-15.950</td>
<td>3.261</td>
<td>0.000</td>
</tr>
<tr>
<td>Collateral class A as % of EAD</td>
<td>1.802</td>
<td>0.562</td>
<td>0.001</td>
</tr>
<tr>
<td>Collateral class C as % of EAD</td>
<td>1.599</td>
<td>0.359</td>
<td>0.000</td>
</tr>
<tr>
<td>Number of different collateral classes</td>
<td>1.589</td>
<td>0.282</td>
<td>0.000</td>
</tr>
<tr>
<td>Year of loan origination before 1995</td>
<td>-1.032</td>
<td>0.107</td>
<td>0.000</td>
</tr>
<tr>
<td>Overdraft type of loan</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Worm Plot

Normal Q-Q Plot

15th Computing in Economics and Finance, Sydney, July 15-17, 2009
### Introduction

Key issues of LGD discount rate modelling include data and risk drivers, methodology, results, and conclusions.

### Data & Risk Drivers

<table>
<thead>
<tr>
<th>Model</th>
<th>Sample</th>
<th>Exposure at default - EAD</th>
<th>Collateral class A as % of EAD</th>
<th>Collateral class B as % of EAD</th>
<th>Collateral class C as % of EAD</th>
<th>Collateral class D as % of EAD</th>
<th>Age of a counterparty</th>
<th>Length of business relationship</th>
<th>Number of different collateral classes</th>
<th>Year of default before 1995</th>
<th>Year of loan origination before 1995</th>
<th>New industry</th>
<th>Industry not specified</th>
<th>Number of loans</th>
<th>Investment type of loan</th>
<th>Overdraft type of loan</th>
<th>Revolving type of loan</th>
<th>Purpose type of loan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear model</td>
<td>A</td>
<td>-0.230</td>
<td>0.411</td>
<td>0.395</td>
<td>-0.197</td>
<td>0.079</td>
<td>-0.283</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fractional response Logit link</td>
<td>A</td>
<td>1.607</td>
<td>1.679</td>
<td>-0.939</td>
<td>-1.254</td>
<td>-0.725</td>
<td>-1.032</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fractional response Beta - Log-log link</td>
<td>A</td>
<td>1.607</td>
<td>1.679</td>
<td>-0.939</td>
<td>-1.254</td>
<td>-0.725</td>
<td>-1.032</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fractional response Beta - Complementary Log-log link</td>
<td>A</td>
<td>1.607</td>
<td>1.679</td>
<td>-0.939</td>
<td>-1.254</td>
<td>-0.725</td>
<td>-1.032</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fractional response Beta - Complementary Log-log link</td>
<td>A+B</td>
<td>1.607</td>
<td>1.679</td>
<td>-0.939</td>
<td>-1.254</td>
<td>-0.725</td>
<td>-1.032</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ordinal response Logit link</td>
<td>A</td>
<td>-0.250</td>
<td>2.799</td>
<td>2.338</td>
<td>-1.208</td>
<td>0.581</td>
<td>-1.769</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ordinal response Complementary Log-log link</td>
<td>A</td>
<td>-0.250</td>
<td>2.799</td>
<td>2.338</td>
<td>-1.208</td>
<td>0.581</td>
<td>-1.769</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear model</td>
<td>B</td>
<td>2.250</td>
<td>1.329</td>
<td>1.382</td>
<td>0.724</td>
<td>-0.943</td>
<td>0.367</td>
<td>-0.980</td>
<td>0.507</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fractional response Logit link</td>
<td>B</td>
<td>2.250</td>
<td>1.329</td>
<td>1.382</td>
<td>0.724</td>
<td>-0.943</td>
<td>0.367</td>
<td>-0.980</td>
<td>0.507</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fractional response Log-log link</td>
<td>B</td>
<td>2.250</td>
<td>1.329</td>
<td>1.382</td>
<td>0.724</td>
<td>-0.943</td>
<td>0.367</td>
<td>-0.980</td>
<td>0.507</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fractional response Complementary Log-log link</td>
<td>B</td>
<td>2.250</td>
<td>1.329</td>
<td>1.382</td>
<td>0.724</td>
<td>-0.943</td>
<td>0.367</td>
<td>-0.980</td>
<td>0.507</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fractional response Beta - Log-log link</td>
<td>B</td>
<td>2.250</td>
<td>1.329</td>
<td>1.382</td>
<td>0.724</td>
<td>-0.943</td>
<td>0.367</td>
<td>-0.980</td>
<td>0.507</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fractional response Beta - Complementary Log-log link</td>
<td>B</td>
<td>2.250</td>
<td>1.329</td>
<td>1.382</td>
<td>0.724</td>
<td>-0.943</td>
<td>0.367</td>
<td>-0.980</td>
<td>0.507</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fractional response Beta - Complementary Log-log link</td>
<td>B+A</td>
<td>2.250</td>
<td>1.329</td>
<td>1.382</td>
<td>0.724</td>
<td>-0.943</td>
<td>0.367</td>
<td>-0.980</td>
<td>0.507</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ordinal response Logit link</td>
<td>B</td>
<td>-0.250</td>
<td>2.799</td>
<td>2.338</td>
<td>-1.208</td>
<td>0.581</td>
<td>-1.769</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ordinal response Complementary Log-log link</td>
<td>B+A</td>
<td>-0.250</td>
<td>2.799</td>
<td>2.338</td>
<td>-1.208</td>
<td>0.581</td>
<td>-1.769</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Results & Conclusions**

The results include linear models and fractional responses, with coefficients and p-values for each model. The models include Logit, Log-log, Complementary Log-log, Beta-Logit, Beta-Log-log, and Beta-Complementary Log-log links. The conclusions are not explicitly stated in the table.
goodness-of-fit

- Parametric performance measures

<table>
<thead>
<tr>
<th>Model</th>
<th>Subsample A</th>
<th>Subsample B</th>
<th>Whole sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear model</td>
<td>0.603</td>
<td>0.841</td>
<td>0.602</td>
</tr>
<tr>
<td>Fractional response Logit link</td>
<td>0.580</td>
<td>0.846</td>
<td>0.536</td>
</tr>
<tr>
<td>Fractional response Log-log link</td>
<td>0.557</td>
<td>0.829</td>
<td>0.574</td>
</tr>
<tr>
<td>Fractional response Complementary Log-log link</td>
<td>0.573</td>
<td>0.820</td>
<td>0.534</td>
</tr>
<tr>
<td>Fractional response Beta - Logit Link</td>
<td>0.540</td>
<td>0.755</td>
<td>0.550</td>
</tr>
<tr>
<td>Fractional response Beta - Log-log link</td>
<td>0.541</td>
<td>0.784</td>
<td>0.543</td>
</tr>
<tr>
<td>Fractional response Beta - Complementary Log-log link</td>
<td>0.511</td>
<td>0.647</td>
<td>0.550</td>
</tr>
<tr>
<td>Ordinal response Logit link</td>
<td>0.548</td>
<td>n/a</td>
<td>0.610</td>
</tr>
<tr>
<td>Ordinal response Complementary Log-log link</td>
<td>0.563</td>
<td>n/a</td>
<td>0.605</td>
</tr>
</tbody>
</table>

- Non-Parametric performance measures

<table>
<thead>
<tr>
<th>Model</th>
<th>Subsample A (&gt; 1 year)</th>
<th>Subsample B (&lt; 1 year)</th>
<th>Whole Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power Statistic</td>
<td>SE</td>
<td>Power Statistic</td>
<td>SE</td>
</tr>
<tr>
<td>Linear model</td>
<td>64.4%</td>
<td>4.0%</td>
<td>87.2%</td>
</tr>
<tr>
<td>Fractional response Logit link</td>
<td>60.4%</td>
<td>4.3%</td>
<td>88.5%</td>
</tr>
<tr>
<td>Fractional response Log-log link</td>
<td>57.7%</td>
<td>5.0%</td>
<td>89.5%</td>
</tr>
<tr>
<td>Fractional response Complementary Log-log link</td>
<td>59.3%</td>
<td>4.0%</td>
<td>88.7%</td>
</tr>
<tr>
<td>Fractional response Beta - Logit Link</td>
<td>58.6%</td>
<td>4.9%</td>
<td>70.9%</td>
</tr>
<tr>
<td>Fractional response Beta - Log-log link</td>
<td>58.5%</td>
<td>5.4%</td>
<td>69.0%</td>
</tr>
<tr>
<td>Fractional response Beta - Complementary Log-log link</td>
<td>55.7%</td>
<td>4.7%</td>
<td>55.7%</td>
</tr>
<tr>
<td>Ordinal response Logit link</td>
<td>58.3%</td>
<td>4.5%</td>
<td>n/a</td>
</tr>
<tr>
<td>Ordinal response Complementary Log-log link</td>
<td>61.1%</td>
<td>3.8%</td>
<td>n/a</td>
</tr>
</tbody>
</table>
goodness-of-fit

- scatter plots

Linear model

Fractional responses, (QML estimator, logit link)

Fractional responses (QML estimator, log-log link)

Beta distribution log-log link
back-testing results

<table>
<thead>
<tr>
<th></th>
<th>Power in sample</th>
<th>Power out sample</th>
<th>SE in sample</th>
<th>SE out sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fractional response log-log link</td>
<td>66%</td>
<td></td>
<td>4%</td>
<td></td>
</tr>
<tr>
<td>Fractional response log-log link– backtesting (S1, S2, S3)</td>
<td>65%</td>
<td>51%</td>
<td>6%</td>
<td>7%</td>
</tr>
<tr>
<td>Fractional response log-log link– backtesting (LGD ord)</td>
<td>67%</td>
<td>62%</td>
<td>5%</td>
<td>5%</td>
</tr>
<tr>
<td>Fractional response log-log link– backtesting (random)</td>
<td>65%</td>
<td>61%</td>
<td>6%</td>
<td>6%</td>
</tr>
</tbody>
</table>
## Conclusion

- Analyzed several aspects of economic loss
- Appropriate discount factor, timing of the recovery rates, efficient recovery period of workout department
- Statistical models to test empirically the determinant of recovery rates
- Main drivers: certain collateral type, loan size, business connection, year of the loan origination
- Different models provided similar results
- Log-log link model performed better implying asymmetric response of the dependent variable
Thank you for attention!