Modelling Bank Loan LGD of Corporate and SME Segment

Radovan Chalupka, Juraj Kopecsni
Charles University, Prague

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

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
    - o the loss of principal
    - o the carrying costs of non-performing loans (interest income)
    - o 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

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
  - o nominal risk-free rate (r_f)
  - o default premium (δ_{dp})
  - o risk-premium (δ_{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 π 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 δ_{rp} changes, inflation changes, and δ_{dp} should not be used to discount already reduced cash flows
- **lender's cost of equity**
  - sum of r_f and δ_{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

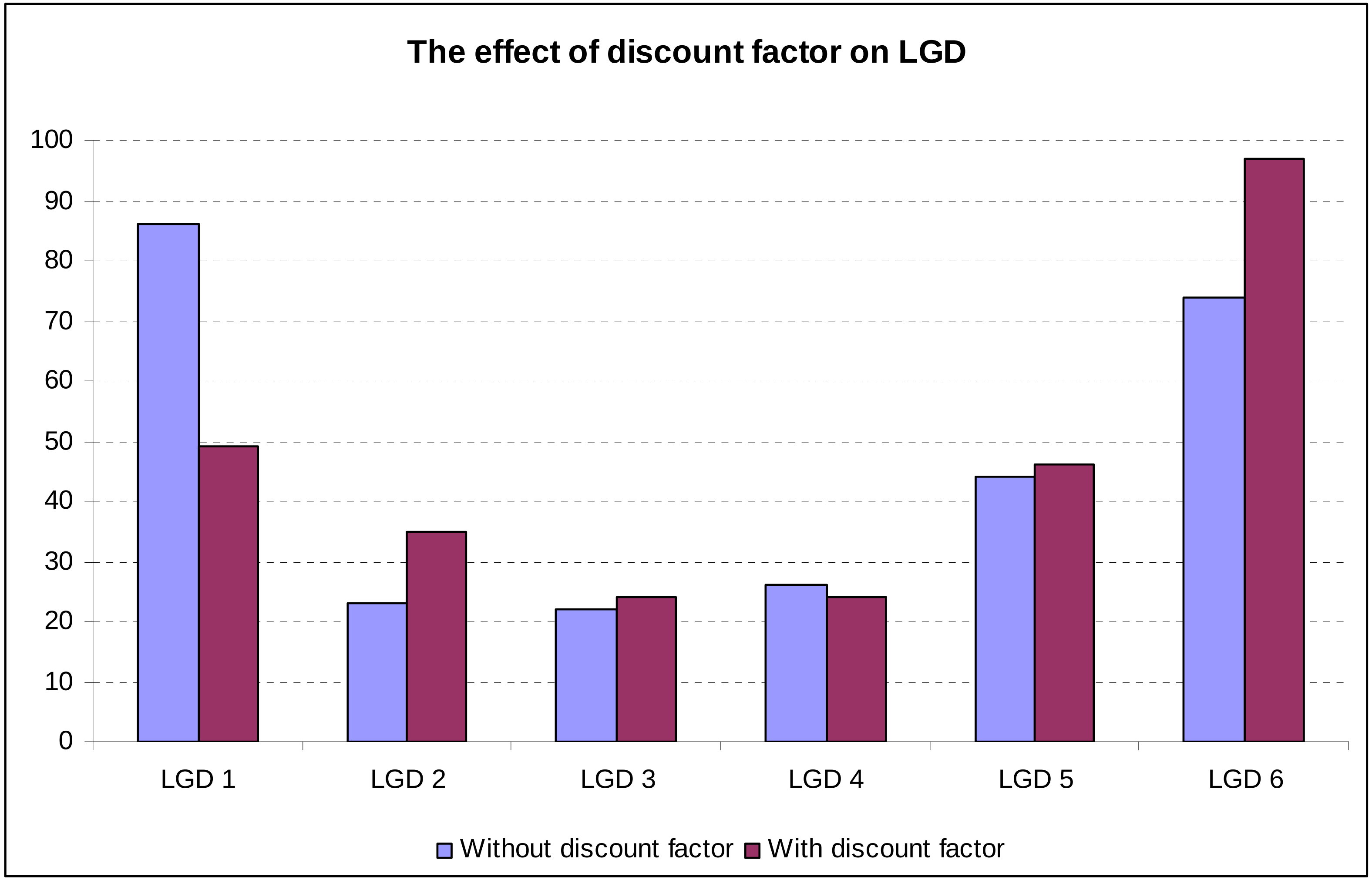
risk premiums above r_f

- 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



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
- Asarnov and Edwards (1995), Carey (1998) and Gupton et al (2000) find that seniority and collateral matter

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

- Asarnov and Edwards (1995), Carty and Lieberman (1996) and Thornburn (2000) find no relationship between LGD and size of loan on US market
- Hurt and Felsovalyi (1998) show that large loan default exhibiting lower recovery rates.

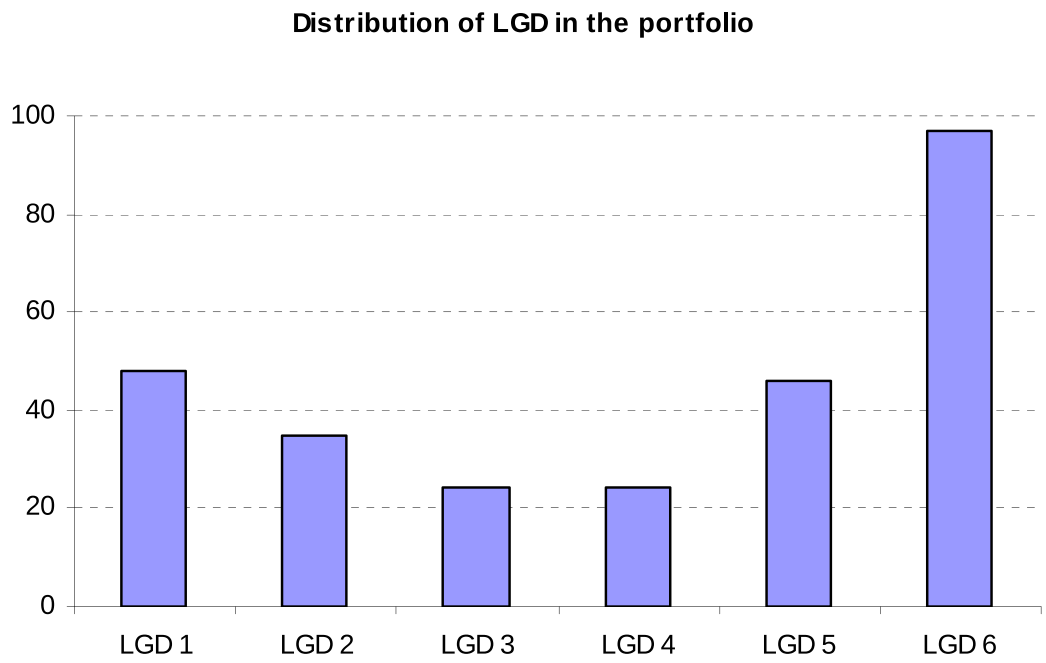
1. introduction
2. key issues of LGD
3. discount rate
4. modelling LGD
5. **data & risk drivers**
6. methodology
7. results & conclusions

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% <=LGD<10% LGD2 10%<=LGD<30%
- LGD3 30%<=LGD<50% LGD4 50%<=LGD<70%
- LGD5 70%<=LGD<90% LGD6 90%<=LGD<100%



explanatory variables

- **counterparty related factors**
  - industry classification, age of the company, year of default, year of loan origination, length of business connection
- **contract related factors**
  - type of the contract, exposure at default, interest rate on the loan, number of different type of contracts
- **collateral related factors**
  - 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	Type	Correlation
Counterparty related factors		
Age of a counterparty	Continous	Positive
Length of business relationship	Continous	?
Year of default before 1995	Dummy	Negative
Year of loan origination before 1995	Dummy	Negative
New industries	Dummy	?
Industry not specified	Dummy	?
Contract related factors		
Exposure at default	Continous	Negative
Number of loans	Categorical	?
Investment type of loan	Dummy	?
Overdraft type of loan	Dummy	?
Revolving type of loan	Dummy	?
Purpose type of loan	Dummy	?
Collateral related factors		
Collateral value of A relative to EAD	Continous	Positive
Collateral value of B relative to EAD	Continous	Positive
Collateral value of C relative to EAD	Continous	Positive
Collateral value of D relative to EAD	Continous	Positive
Number of different collaterals	Categorical	Positive

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

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

- 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

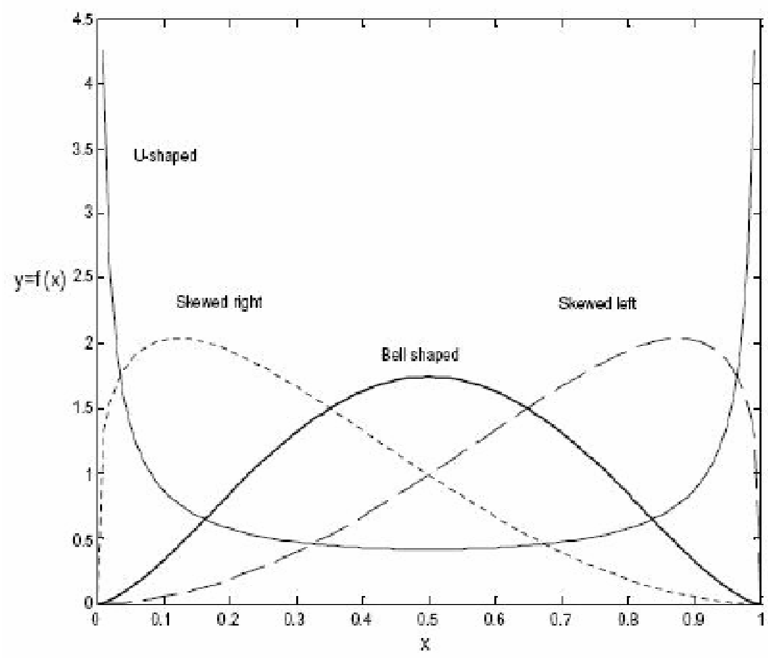
functions

- Symmetric Logit link $$G(\alpha + \boldsymbol{\beta}' x) = \frac{\exp(\alpha + \boldsymbol{\beta}' x)}{1 + \exp(\alpha + \boldsymbol{\beta}' x)}$$
- Asymmetric Log-log link $$G(\alpha + \boldsymbol{\beta}' x) = e^{-e^{-\alpha + \boldsymbol{\beta}' x}}$$

beta inflated distribution

$$f_Y(y \mid \mu,\sigma,\nu,\tau) = \begin{cases} p_0 & if \; y = 0 \\ (1 - p_0 - p_1)\dfrac{1}{B(\alpha,\beta)}\, y^{\alpha-1}(1-y)^{\beta-1} & f \; 0 < y < 1 \\ p_1 & f \; y = 1 \end{cases}$$

for $0 \le y \le 1$, 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

$$logit[P(Y \leq j \mid \mathbf{x})] = \log \frac{P(Y \leq j \mid \mathbf{x})}{1 - P(Y \leq j \mid \mathbf{x})}$$

$$= \log \frac{\pi_1(\mathbf{x}) + \ldots + \pi_j(\mathbf{x})}{\pi_{j+1}(\mathbf{x}) + \ldots + \pi_J(\mathbf{x})}, \quad j = 1, \ldots, J-1$$

- each cumulative logit uses all *J* response categories, a model for 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

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

- each cumulative logit has its own intercept, the $\{\alpha_j\}$ are increasing in *j*, since P($Y \leq j|\mathbf{x}$) increases in j for fixed x, and the logit is an increasing function of this probability, this model has the same effects **β** 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

$$\log\{-\log[1-P(Y \le j \mid \mathbf{x})]\} = \alpha_j + \boldsymbol{\beta}'\mathbf{x}, \quad j = 1, \ldots, J-1$$

- this complementary log-log link has the property

$$P(Y \le j \mid \mathbf{x}_1) = [P(Y \le j \mid \mathbf{x}_2)]^{\exp[\boldsymbol{\beta}'(\mathbf{x}_1 - \mathbf{x}_2)]}$$

- with this link, P(Y≤j) approaches 1 at a faster rate than it approaches 0, the related log/log link log{–log[P(Y ≤ 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
  - continuous variables were plotted for each LGD grade against the value to get “a feel” of the underlying relationship
  - categorical variables were tabulated to form an expectation of a potential relationship
  - the frequency table gives information whether there are enough counts for each cell to estimate reliably the effect
  - 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 the lowest t-statistic

selecting the appropriate model

- in order to select the most appropriate model, some commonly used procedures were followed
  - 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 (backward elimination) on Akaike (AIC) and Schwarz information criteria (SIC)
  - o Worm plot for residuals and QQ-plots were utilised to have a visual indication of normality of residuals
  - o Normality of residuals was tested by Shapiro-Wilk normality test

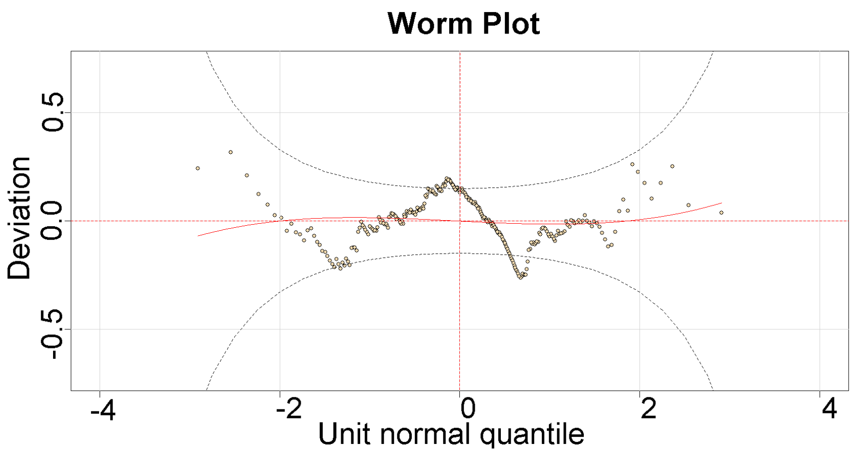
model evaluation: back-testing

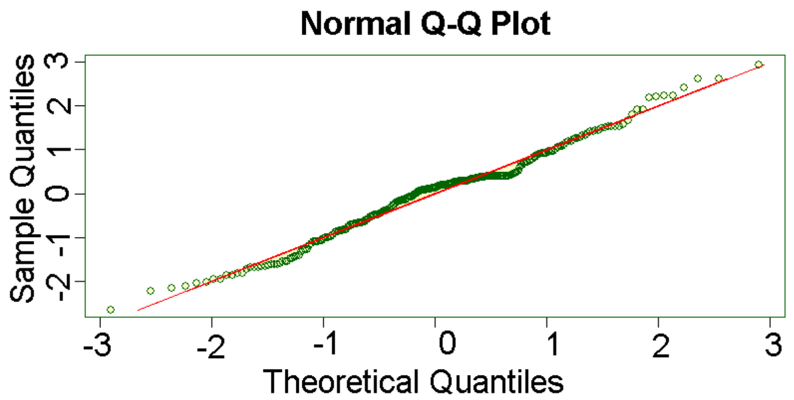
- process of assessing the model predictive power by using historical data:
  - In sample back-testing
  - Out of sample back-testing
  - Out of time back-testing
- significant differences between the observed and predicted values indicate that the model is not robust (over fitting) with regard to changes 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

Recovery rate determinants	Subsample A (>1 year)			Subsample B (<1 year)			Whole sample		
	Value	Std. error	P-value	Value	Std. error	P-value	Value	Std. error	P-value
Exposure at default –EAD				-15.950	3.261	0.000	-1.128	0.271	0.000
Collateral class A as % of EAD	1.802	0.562	0.001				1.491	0.546	0.006
Collateral class C as % of EAD	1.599	0.359	0.000				1.612	0.358	0.000
Number of different collateral classes				1.589	0.282	0.000			
Year of loan origination before 1995	-1.032	0.107	0.000				-1.128	0.112	0.000
Overdraft type of loan							0.825	0.194	0.000





Model	Sample	Exposure at default - EAD	Collateral class A as % of EAD	Collateral class B as % of EAD	Collateral class C as % of EAD	Collateral class D as % of EAD	Age of a counterparty	Length of business relationship	Number of different collateral classes	Year of default before 1995	Year of loan origination before 1995	New industries	Industry not specified	Number of loans	Investment type of loan	Overdraft type of loan	Revolving type of loan	Purpose type of loan
Linear model	A	-0.230	0.411		0.395			-0.197	0.079		-0.283							
Fractional response Logit link	A		2.552		2.565			-1.225			-1.622							
Fractional response Log-log link	A		1.802		1.599						-1.032							
Fractional response Complementary Log-log link	A		1.607		1.679			-0.939			-1.254							
Fractional response Beta - Logit Link	A	-1.426			0.963						-1.364		-0.725					
Fractional response Beta - Log-log link	A	-0.730			0.716						-0.797		-0.424					
Fractional response Beta - Complementary Log-log link	A	-1.230									-1.121		-0.611					
Ordinal response Logit link	A	-2.500	2.799		2.338			-1.208	0.581		-1.769							
Ordinal response Complementary Log-log link	A	-1.648	1.329		1.382		0.724	-0.943	0.367		-0.980		-0.507					
Linear model	B	-2.250							0.211		-0.369			-0.097	0.154	0.125		0.143
Fractional response Logit link	B	-27.240							2.149									
Fractional response Log-log link	B	-15.950							1.589									
Fractional response Complementary Log-log link	B	-13.096							0.936									
Fractional response Beta - Logit Link	B	-24.382		1.854	-2.227		1.329	1.900							2.984	1.718	0.909	1.443
Fractional response Beta - Log-log link	B	-20.540		1.766	-2.318		1.074	2.014				0.814			2.418	1.702	1.099	1.527
Fractional response Beta - Complementary Log-log link	B	-9.435		0.456											0.581			
Linear model	A+B	-0.330	0.359		0.329			-0.179	0.103		-0.298					0.186		
Fractional response Logit link	A+B	-2.873							0.666		-1.567					1.008		
Fractional response Log-log link	A+B	-1.128	1.491		1.612						-1.128					0.825		
Fractional response Complementary Log-log link	A+B	-2.254							0.471		-1.247					0.591		
Fractional response Beta - Logit Link	A+B	-1.946							0.311		-1.390		-0.695			0.845		
Fractional response Beta - Log-log link	A+B	-0.989							0.191		-0.830		-0.445			0.636		
Fractional response Beta - Complementary Log-log link	A+B	-1.504							0.237		-1.083		-0.454			0.500		
Ordinal response Logit link	A+B	-3.471	2.242		1.802		1.202	-1.348	0.652		-1.796		-0.811			1.133		
Ordinal response Complementary Log-log link	A+B	-2.218	1.144		1.050		0.725	-0.833	0.437		-1.002		-0.469			0.632		

goodness-of-fit

- Parametric performance measures

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

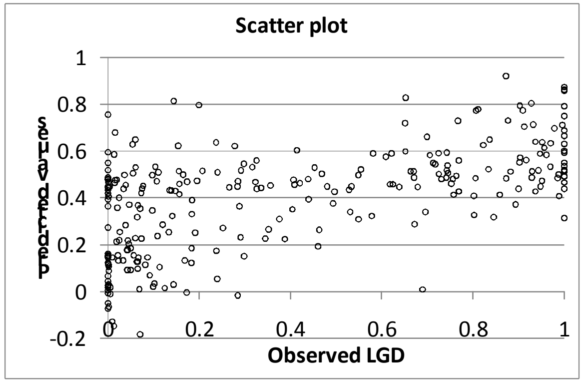
- Non-Parametric performance measures

Model	**Subsample A (> 1 year)**		**Subsample B (< 1 year)**		**Whole Sample**	
	Power Statistic	**SE**	**Power Statistic**	**SE**	**Power Statistic**	**SE**
Linear model	64.4%	4.0%	87.2%	8.8%	70.8%	3.6%
Fractional response Logit link	60.4%	4.3%	88.5%	5.7%	66.5%	3.9%
Fractional response Log-log link	57.7%	5.0%	89.5%	6.4%	65.5%	3.5%
Fractional response Complementary Log-log link	59.3%	4.0%	88.7%	7.8%	66.6%	4.2%
Fractional response Beta - Logit Link	58.6%	4.9%	70.9%	16.8%	67.1%	3.6%
Fractional response Beta - Log-log link	58.5%	5.4%	69.0%	19.8%	66.8%	3.8%
Fractional response Beta - Complementary Log-log link	55.7%	4.7%	55.7%	21.5%	67.4%	3.6%
Ordinal response Logit link	58.3%	4.5%	n/a	n/a	72.0%	2.8%
Ordinal response Complementary Log-log link	61.1%	3.8%	n/a	n/a	71.8%	3.8%

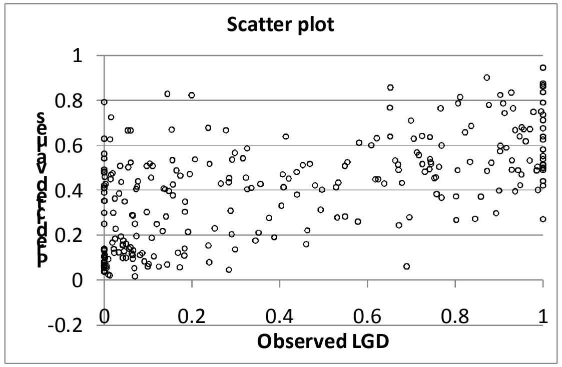
goodness-of-fit

- scatter plots

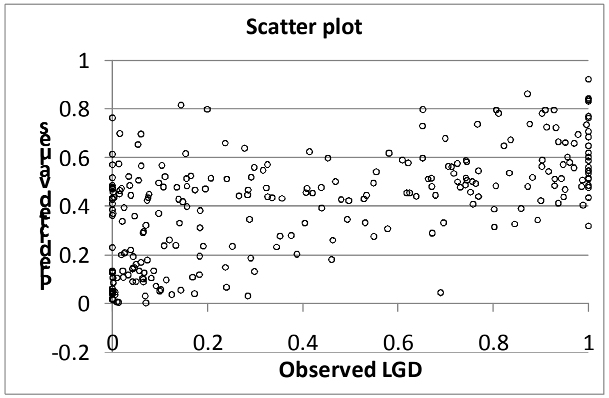
Linear model



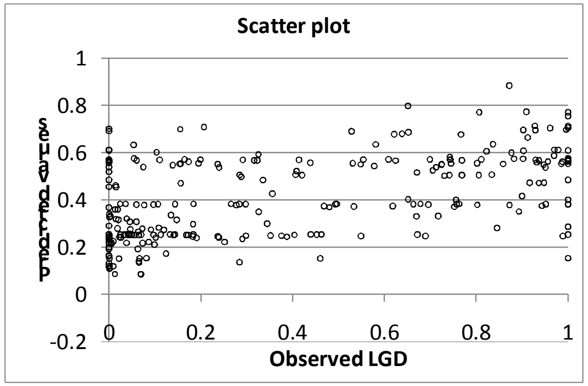
Fractional responses, (QML estimator, logit link)



Fractional responses (QML estimator, log-log link)



Beta distribution log-log link



back-testing results

	Power in sample	Power out sample	SE in sample	SE out sample
Fractional response log-log link	66%		4%	
Fractional response log-log link– backtesting (S1, S2, S3)	65%	51%	6%	7%
Fractional response log-log link– backtesting (LGD ord)	67%	62%	5%	5%
Fractional response log-log link– backtesting (random)	65%	61%	6%	6%

Conclusion

- Analyzed several aspect 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!