

# **Application of Scoring Approach in the LGD Estimation**

**Ivana Zohova**

**Credit Scoring and Credit Control XIV, Edinburgh, 26-28 August 2015**

# Agenda

## I Introduction

- **LGD in Credit Risk Measurement and Management**
- **Assumptions on the LGD Segmentation**
- **Areas of Application**

## II Statistical Techniques

- **Modelling Techniques for Fractional Response Variables**
- **Comparison of Modelling Approaches**
- **Fractional Regression: Assumptions and Estimation**

## III Application LGD Scorecard

- **Model Variables and Performance**
- **LGD Grades and Calibration**
- **Comparison: Performing vs. Defaulted Population**
- **Comparison: Product vs. Scoring Approach**

# I Introduction

# Loss Given Default

## Credit Risk Measurement and Management

- Describes the portion of an exposure lost by the bank in case a borrower defaults. It is expressed as a percentage of the outstanding amount at time of default.
- An important **input** for the calculation of:
  - Regulatory and economic capital requirement,
  - Loan loss provisions,
  - Credit risk pricing.

An indicator for evaluation of efficiency of the workout process.

- Individual LGD usually calculated using the **workout approach**:

$$LGD_i(t) = 1 - \frac{\sum_{j=1}^t PV(CF_{ij})}{EAD_i},$$

where

- $LGD_i(t)$  = LGD of a defaulted exposure  $i$  at time  $t$  after default,
- $CF_{ij}$  = cash flow resulting from the workout process of the exposure  $i$  registered at time  $j$ ,
- $EAD_i$  = outstanding of the exposure  $i$  at time of default,
- $PV(.)$  = present value function.

# Motivation

## Segmentation Assumptions

### Homogeneity

- The LGD segmentation should be **homogeneous** with regard to all main risk drivers. In practice, however, the granularity of the LGD segmentation is often limited by data availability.

### Representativeness

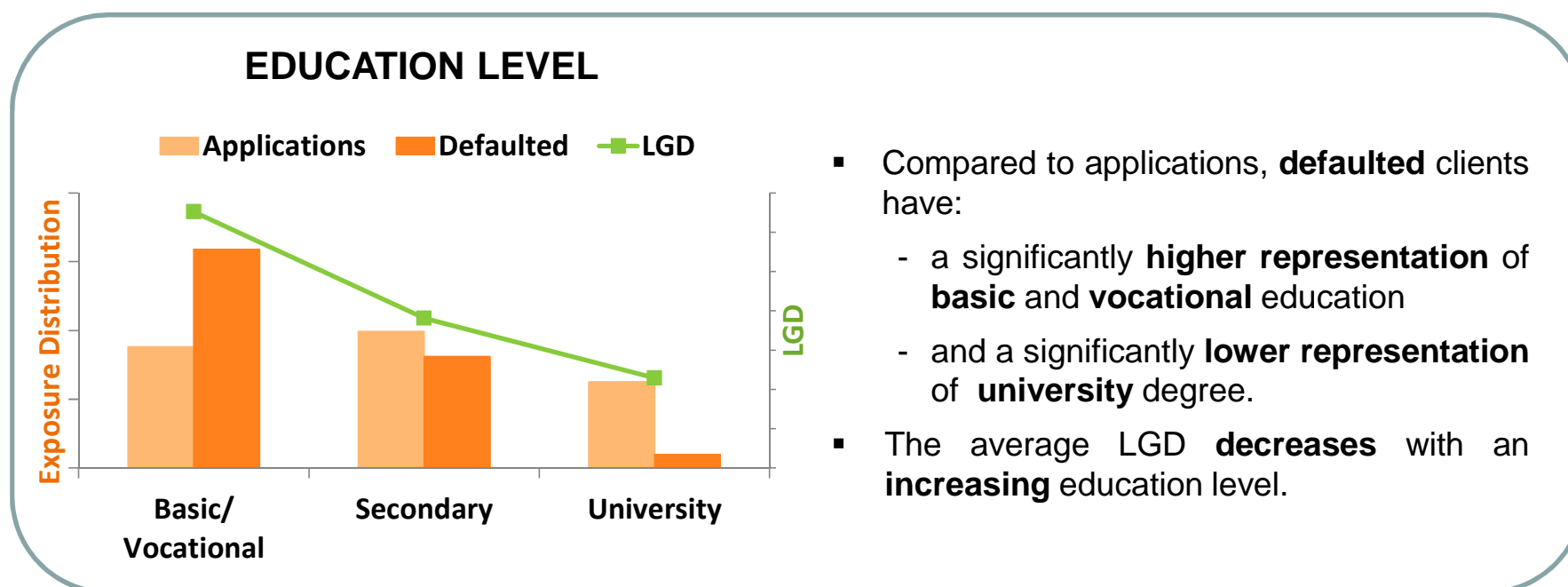
- In the most simplistic way, LGD may be calculated as an **average loss** observed on **defaulted** exposures of an expertly set LGD segment (e.g. a product pool, a loan amount bucket, etc.) and applied to all **non-defaulted** exposures of the same segment.
- To avoid an estimation error, defaulted exposures of each LGD segment need to be **representative** of the non-defaulted population of the given segment.

- However, **non-defaulted** and **defaulted** populations of the same expert based segment (product, amount bucket, etc.) often **differ** in terms of client, loan or collateral characteristics and consequently, LGD estimates may be severely **biased**.

- A comparison of real bank non-defaulted and defaulted portfolios for selected client variables presented on the following slides.

# Motivation

## Applications vs. Defaulted Population [1/4]



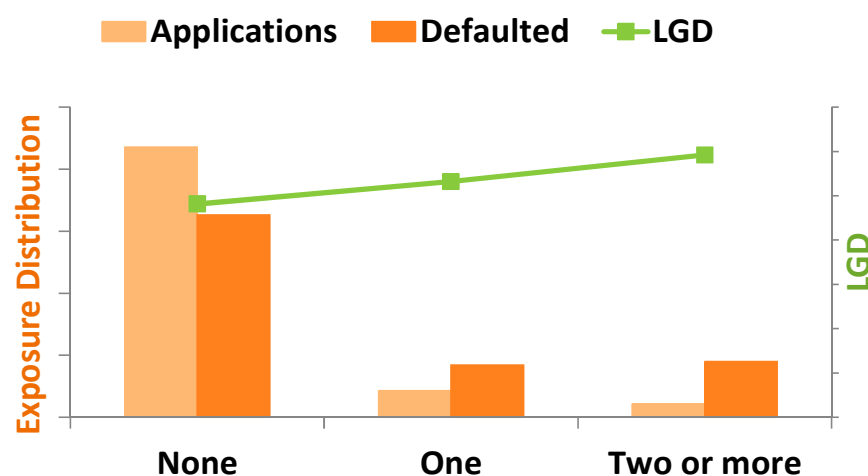
### Note:

- Analysis conducted on a selected retail unsecured product in one of Erste Group's subsidiaries.
- Presented grouping may not be identical to that potentially employed in the LGD scorecard.
- The horizontal axis may not intersect the "Exposure Distribution" axis at point 0.
- The value of LGD per categories not disclosed.

# Motivation

## Applications vs. Defaulted Population [2/4]

### NUMBER OF REJECTED APPLICATIONS



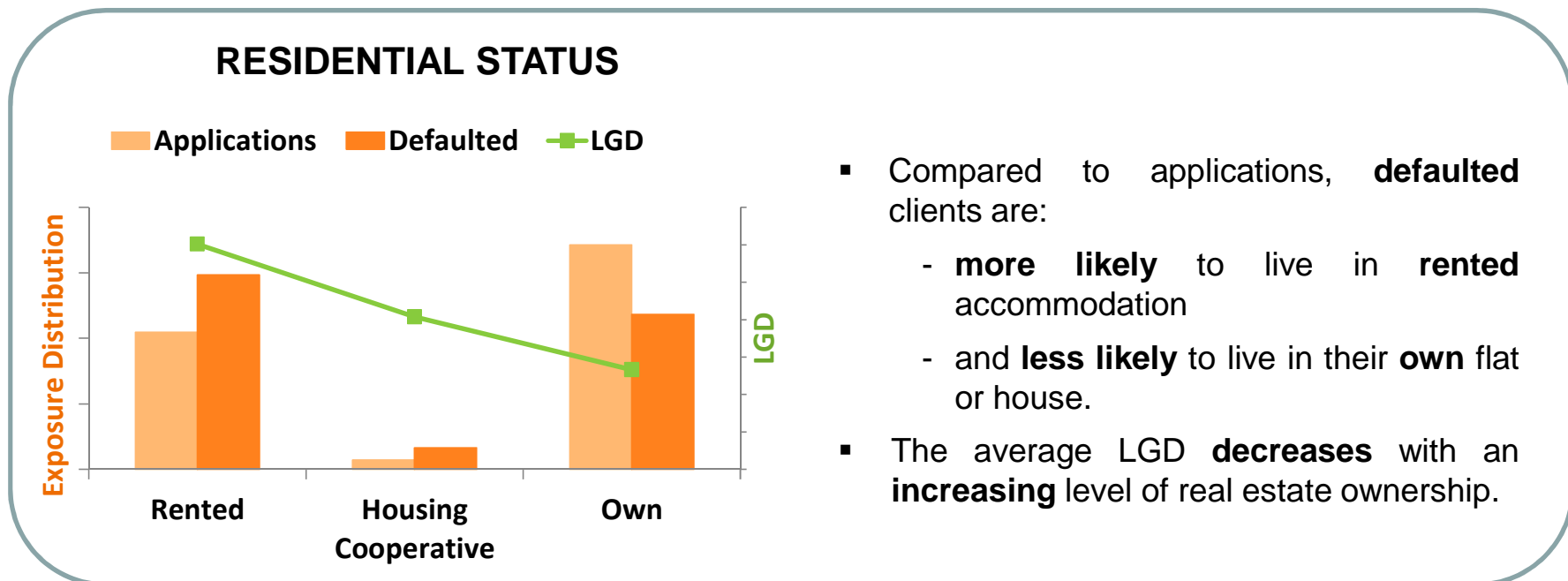
- Number of applications rejected by banks as reported in Bureau
- Compared to applications, **defaulted** clients tend to have a **higher number** of applications rejected by banks.
- The average LGD **increases** with an **increasing** number of rejected applications.

#### Note:

- Analysis conducted on a selected retail unsecured product in one of Erste Group's subsidiaries.
- Presented grouping may not be identical to that potentially employed in the LGD scorecard.
- The horizontal axis may not intersect the "Exposure Distribution" axis at point 0.
- The value of LGD per categories not disclosed.

# Motivation

## Applications vs. Defaulted Population [3/4]

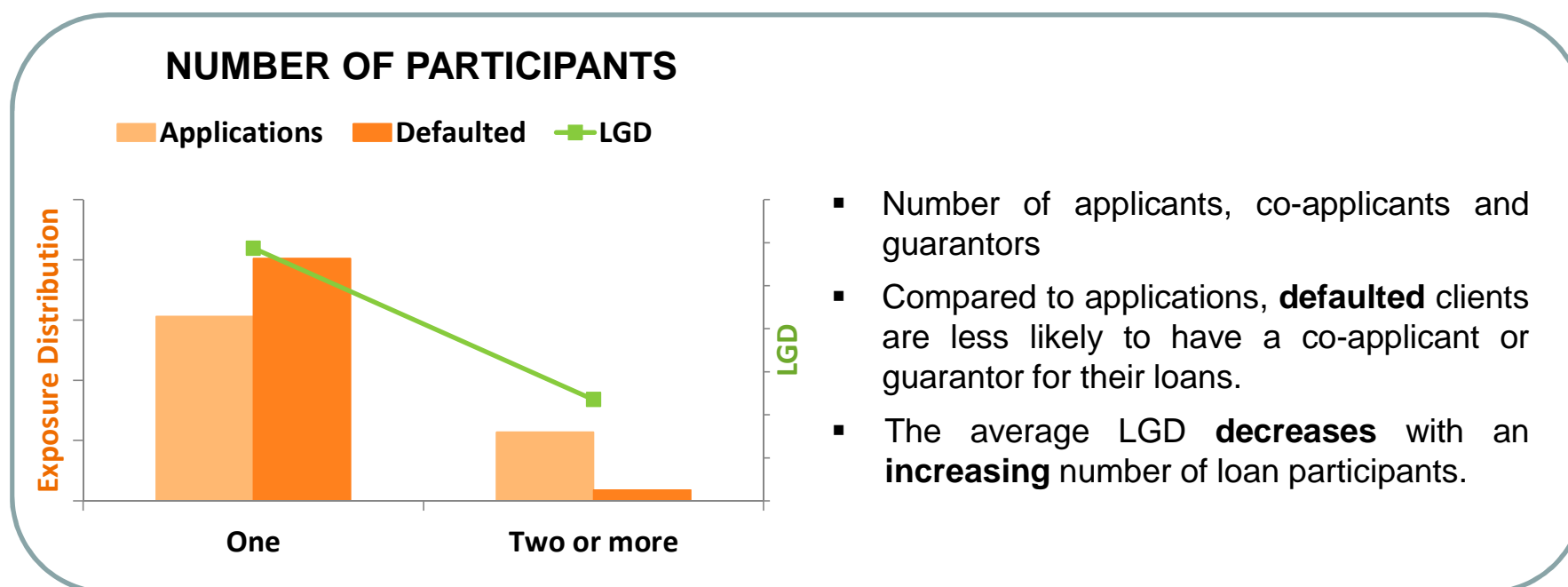


### Note:

- Analysis conducted on a selected retail unsecured product in one of Erste Group's subsidiaries.
- Presented grouping may not be identical to that potentially employed in the LGD scorecard.
- The horizontal axis may not intersect the "Exposure Distribution" axis at point 0.
- The value of LGD per categories not disclosed.

# Motivation

## Applications vs. Defaulted Population [4/4]



### Note:

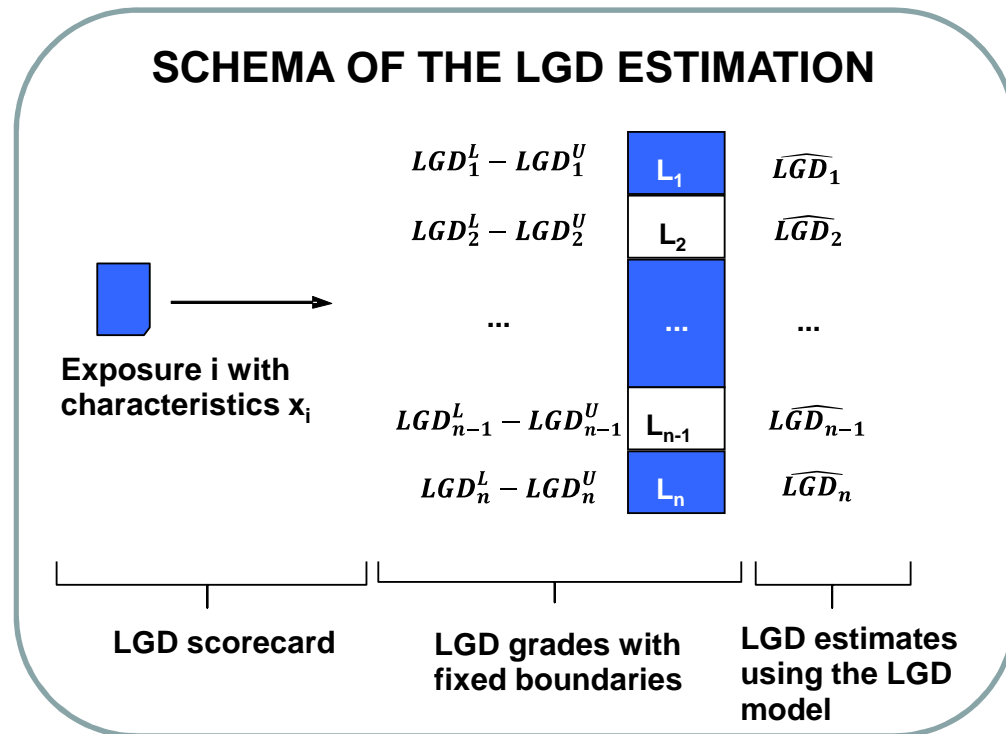
- Analysis conducted on a selected retail unsecured product in one of Erste Group's subsidiaries.
- Presented grouping may not be identical to that potentially employed in the LGD scorecard.
- The horizontal axis may not intersect the "Exposure Distribution" axis at point 0.
- The value of LGD per categories not disclosed.

# Motivation

## Scoring-Driven Segmentation

- Simple expert based segmentations (product pools, loan amount buckets, etc.) cannot take into account a complex relationship between LGD and **various variables**:
  - socio-demographic information,
  - behavioural characteristics,
  - exposure values,
  - collateral attributes,
  - etc.

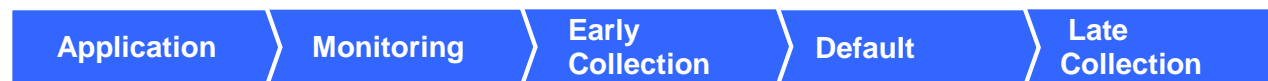
- Such segmentations should be replaced with **LGD grades**, i.e. LGD homogeneous segments with fixed boundaries determined by **LGD scorecards**.



# Areas of Application

## Life Cycle Phases

- **Data sources** available for the LGD scoring **change** over the credit life cycle of an exposure.  
→ **Different models** needed at different phases of an exposure's life.



### Application scoring

- Predicts loss given default occurs within one year (or any other period) **after granting** a new loan.
- Can be used in the **approval** process (to help decide on granting of a new loan, determine pricing, set up the limit system, etc.).

### Behavioral scoring

- Predicts loss given default occurs within one year (or any other period) **after reporting date**.
- Can be used to quantify and manage credit risk in the **existing** portfolio (e.g. calculation of loan loss provisions, regulatory and economic capital requirement, pre-approving loans, etc.).

### Collection scoring

- Predicts ultimate loss for exposures being **in the collection process**.
- Can be used to optimize the workout strategy and help decide on a sequence of workout actions based on client's recovery potential.

# II Statistical Techniques

# Statistical Techniques

## Models for Fractional Response Variables

### Individual LGD

- A **continuous variable** typically ranging the interval **[0,1]**. Values below 0 and above 1 possible.
  - **Below 0**: An exposure fully repaid with paid interest accruals and fees covering more than collection costs and the discounting effect.
  - **Above 1**: Discounted recovered amount not even sufficient to cover collection costs.
- For the scorecard estimation the LGD **restricted** to the interval [0,1].

### Models for a fractional outcome

- Different **models** tested in the pilot phase, including:
  - **Logistic regression with a fixed threshold** (LGD above a specific threshold categorized as *high*, otherwise as *low*. The probability of being *high* estimated via the logistic regression.)
  - **Fractional regression of LGD with the logit link**
  - **Normal linear regression of transformed LGD** (LGD transformed from its beta distribution to a normal distribution.)
  - **Beta regression of LGD with the logit link**

- The tested models led to similar results in terms of Somers' D, with the fractional regression slightly outperforming the other models. The **fractional regression** chosen as the final model form.

# Statistical Techniques

## Models Comparison

	Model	$wR^2$	$wMSE$	$wMAE$	Somers' D
[LR]	Logistic Regression with Threshold $h = 0.2$	15.2%	0.1236	0.3158	25.2%
[FR]	Fractional Regression with Logit Link	17.0%	0.1208	0.3062	25.6%
[TV]	Linear Regression of Transformed Variable with Offset $\sim U(0,10^{-5})$	-1.0%	0.1471	0.3083	25.2%
	Linear Regression of Transformed Variable with Offset $\sim U(0,10^{-4})$	1.1%	0.1441	0.3090	25.2%
	Linear Regression of Transformed Variable with Offset $\sim U(0,10^{-3})$	3.7%	0.1402	0.3094	25.2%
[BR]	Beta Regression with Logit Link and Offset $\sim U(0,10^{-5})$	9.4%	0.1320	0.3324	25.1%
	Beta Regression with Logit Link and Offset $\sim U(0,10^{-4})$	9.8%	0.1313	0.3314	25.1%
	Beta Regression with Logit Link and Offset $\sim U(0,10^{-3})$	10.3%	0.1306	0.3302	25.1%

EAD used as weights in the calculation of  $wR^2$ ,  $wMSE$  and  $wMAE$ .

- All the models performed similarly in terms of Somers' D, [FR] slightly outperforming the other models.
- In terms of  $wR^2$ , [LR] performed slightly worse than [FR]. This is because [LR] reduces the LGD estimates to a combination of just two scenarios: *high* and *low* LGDs.
- In case of [TV] and [BR], the results in terms of  $wR^2$  show a high sensitiveness to the treatment of the border LGD values, especially in case of [TV]. This is because one parametric beta distribution does not describe the empirical distribution of our LGDs sufficiently well and the transformed LGDs do not follow the normal distribution with the zero mean. [TV] is estimated in terms of transformed LGDs and the backward conversion to the LGD space leads in our case to the non-zero average estimation error. [BR] shows better results, as it allows observations to follow different beta distributions (depending on their attributes) and mainly, it enables a direct estimation of the mean of LGDs without need of any transformation.

# Statistical Techniques

## Fractional Regression [1/2]

Introduced in Papke, L. E. and Wooldridge, J. M. (1996): *Econometric Methods for Fractional Response Variables with an Application to 401 (K) Plan Participation Rates*, Journal of Applied Econometrics, vol. 11, p. 619-632.

### Assumptions

- Observations  $\{(y_i, \mathbf{x}_i), i = 1, \dots, N\}$  are mutually **independent**, their distributions not specified and not required to be identical.  
The response variable is a fraction, i.e.  $y \in [0, 1]$ .

- For  $\forall i$  it holds:

$$E(y_i | \mathbf{x}_i) = G(\boldsymbol{\beta}^T \mathbf{x}_i),$$

where

- $\boldsymbol{\beta}^T \mathbf{x}_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}$ ,  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)$  being a vector of regression parameters
- and  $G: \mathbb{R} \rightarrow [0, 1]$  is a **link function**, e.g. the **logit link**:

$$G(\boldsymbol{\beta}^T \mathbf{x}_i) = \frac{\exp(\boldsymbol{\beta}^T \mathbf{x}_i)}{1 + \exp(\boldsymbol{\beta}^T \mathbf{x}_i)}$$

# Statistical Techniques

## Fractional Regression [2/2]

### Estimation procedure

- The parameter vector  $\beta$  estimated using **quasi-maximum likelihood (QML) method**:

$$\max_{\mathbf{b}} \sum_{i=1}^N l_i(\mathbf{b}),$$

where  $l_i(\mathbf{b})$  is the **Bernoulli log-likelihood function**:

$$l_i(\mathbf{b}) = y_i \cdot \log(G(\mathbf{b}^T \mathbf{x}_i)) + (1 - y_i) \cdot \log(1 - G(\mathbf{b}^T \mathbf{x}_i)).$$

- QML estimator  $\hat{\beta}$ , which is **consistent** and  **$\sqrt{N}$ -asymptotically normal** (regardless of the conditional distribution of  $y_i$ ).
- A restrictive assumption on the **conditional variance**:

$$\text{VAR}[y_i | \mathbf{x}_i] = \sigma^2 \cdot G(\beta^T \mathbf{x}_i) \cdot (1 - G(\beta^T \mathbf{x}_i)), \sigma^2 > 0$$

- **Asymptotically robust inference** for the conditional mean parameters

# III Application Scorecard

# Application Scorecard

## Introduction

- Targets the average **LGD on unsecured products** given a client defaults within **one year** after granting a new loan (either unsecured or secured)
- Developed on a mixture of retail unsecured products in one subsidiary of Erste Group
- For explanatory variables **Application** and **Bureau** data used
- Produces an unsecured application **LGD grade** for each client applying for a new loan

### Development phase

- **Allocates** clients to LGD grades based on the expected value of LGD on their unsecured products
- Incomplete cases excluded: Only clients with defaults either already resolved or old enough are considered.

### Calibration phase

- **Assigns** to each LGD grade (and product group) an **LGD value** using the internal LGD model.
- Incomplete cases included: Defaults with open collection process treated by the internal LGD model.

# Application Scorecard

## Model Variables and Performance

- The scorecard used for **order** exposures by their expected LGD, not to directly determine the LGD value itself.

The list of **explanatory variables**:

- **Duration of client relationship**
- **Client exposure on unsecured products**
- **Education**
- **Number of products refused by banks** (in Bureau)
- **Geographical information** (The credit risk of each region is evaluated on the bank's existing clients living in the particular area.)
- **Total number of loan participants**
- **Residential status**
- **Citizenship**

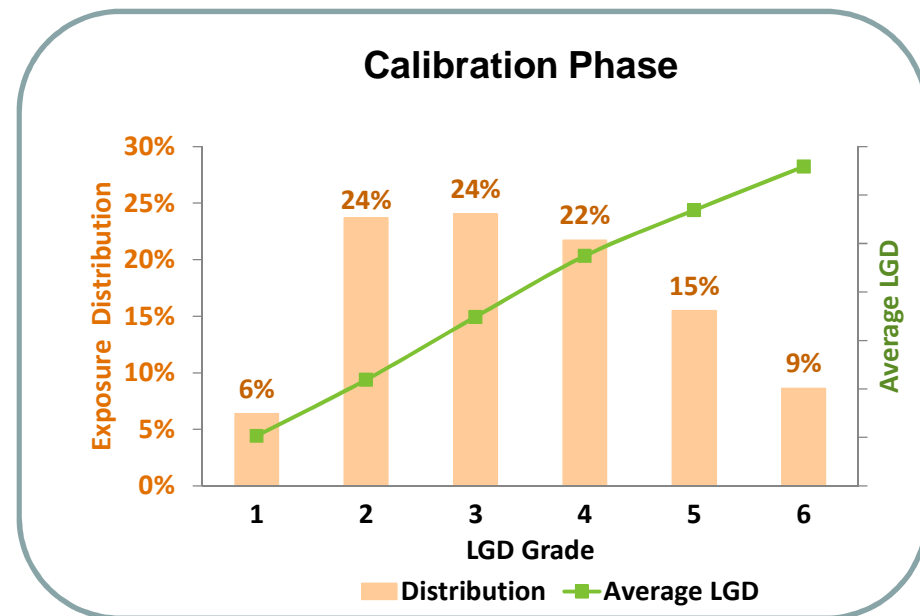
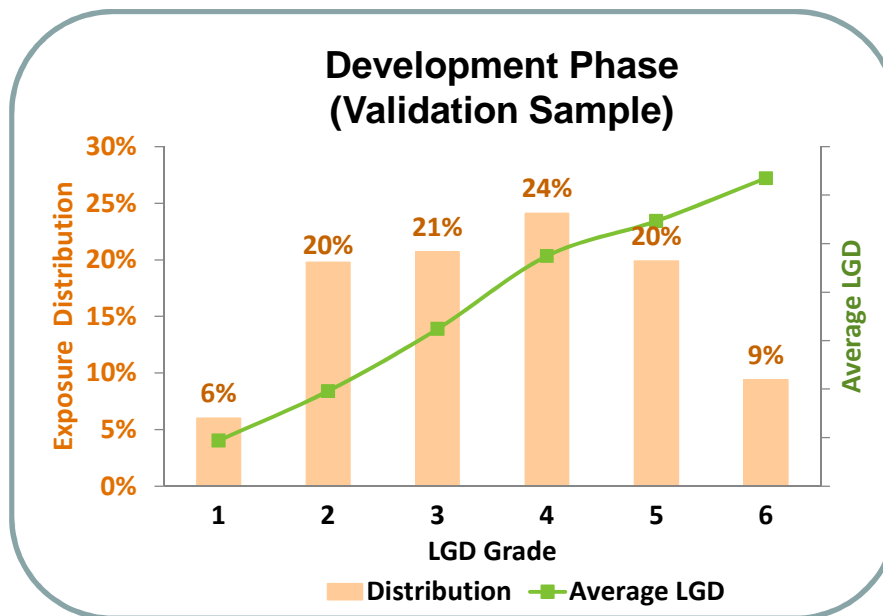
Performance statistics of the LGD scorecard

Sample	R-Squared	MSE	MAE	Somers' D
Training	17.5%	0.1204	0.3063	25.7%
Validation	17.0%	0.1208	0.3062	25.6%

# Application Scorecard

## LGD Grades

- Each **LGD grade** is associated with a distinct LGD interval. However, during the calibration process the LGD value may move up or down due to the inclusion of incomplete defaults into the sample.
- Distribution of defaulted applications by LGD grade and their observed (development phase) or expected (calibration phase) LGD per each grade:**



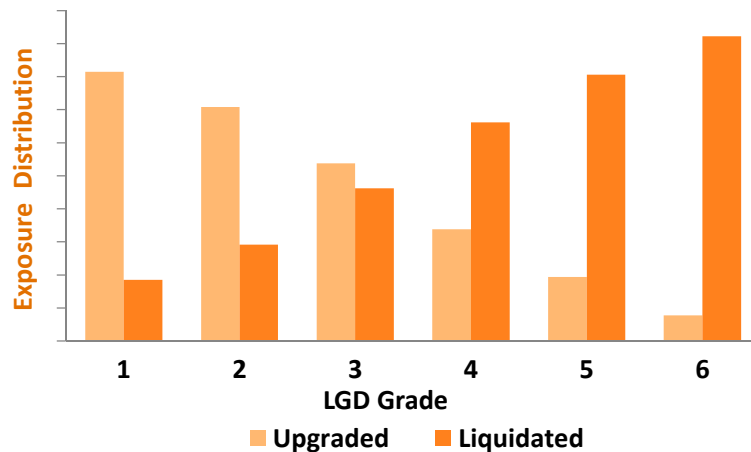
**Note:** Average LGD values per LGD grades not disclosed.

# Application Scorecard

## Calibration

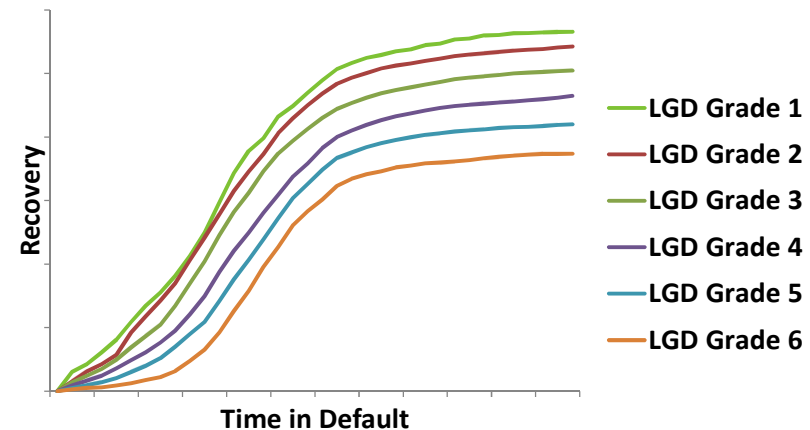
- A portion of upgraded and liquidated defaults, as well as their ultimate recovery and the speed of recovery saturation differ per LGD grades.

**Upgraded vs. Liquidated Defaults by LGD Grade**



- A portion of liquidated defaults increases with an increasing (i.e. worse) LGD grade.
- The horizontal axis may not intersect the “Exposure Distribution” axis at point 0.

**Average Cumulative Recovery Curve on Liquidated Defaults by LGD Grade**



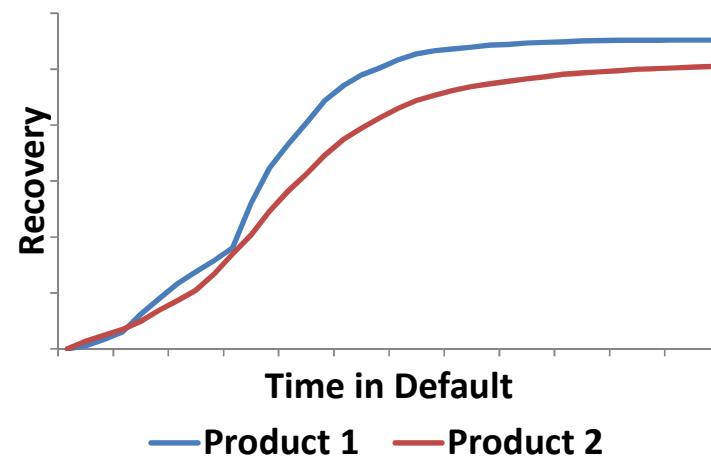
- Average recovery on liquidated defaults decreases with an increasing (i.e. worse) LGD grade.

# Application Scorecard

## Product Differences

- Products may **differ** within **the same LGD grade** due to differences in product characteristics (e.g. limit, contractual interest rate, etc.), as well as differences in the bank's collection strategies.  
→ In that case the **segmentation** should reflect, besides **LGD grades**, also relevant **product groups**.
- Cumulative **recovery curves** for a given LGD grade differentiate by unsecured products both in their **shape** and **end-point**:
  - In case of **upgraded** defaults the end-point is affected mainly by the average value of the **effective interest rate** and the shape by conditions of **client restructuring**.
  - In case of **liquidated** defaults the end-point is affected mainly by the average **exposure amount**. Smaller exposures are generally easier to recover.

**Average Cumulative Recovery Curve on Liquidated Defaults for a Selected LGD Grade and Two Different Unsecured Products**

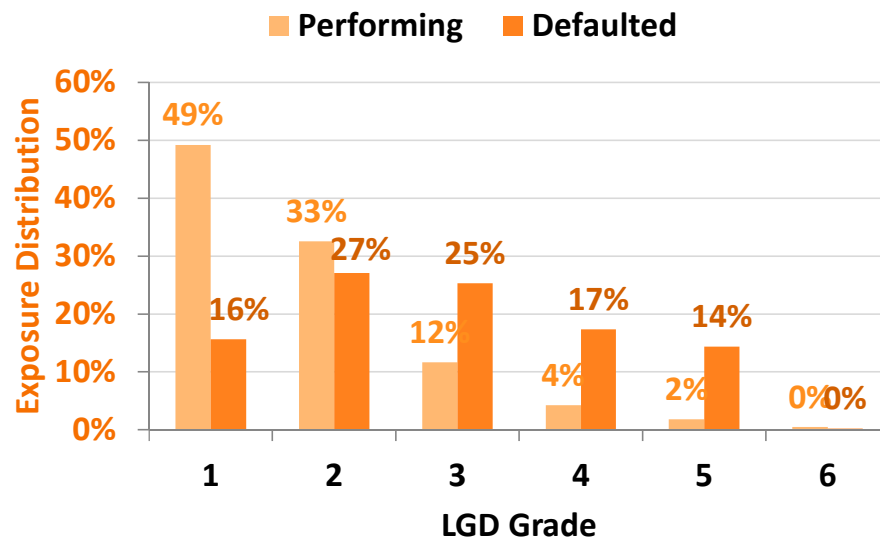


# Application Scorecard

## Performing vs. Defaulted Applications

### Distribution of applications for a selected product by LGD grade

- Clients **not defaulted** during the first year after being granted a new loan
- Clients **defaulted** during the first year after being granted a new loan



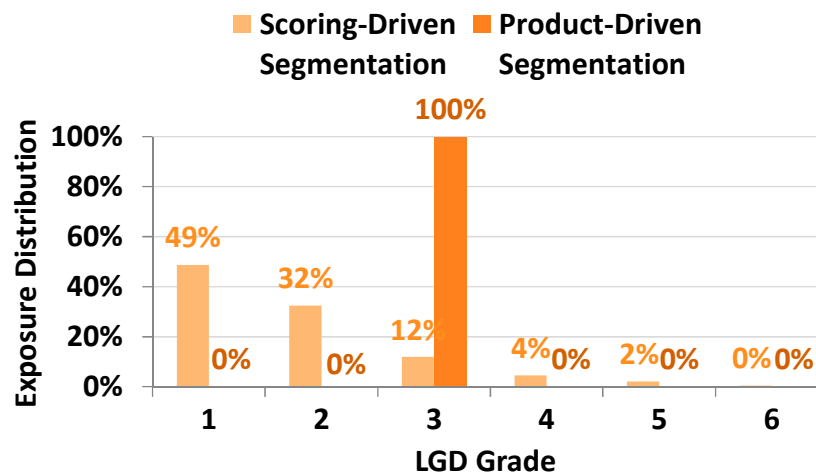
- Performing exposures have on average a considerably **better application LGD grade** than defaulted exposures.
- The application scorecard targets the LGD provided default occurs within the first year after application. Generally it is assumed that losses from exposures defaulting after a longer period than one year from application shall not be higher than losses from the “first-year” defaults.

# Application Scorecard

## Product vs. Scoring-Driven Segmentation

### Distribution of applications for a selected product by LGD range

- Under the **product approach**, an application is assigned an average  $LGD_p$  measured on defaulted exposures belonging to the corresponding product  $P$ .
- Under the **scoring approach**, an application is assigned an average  $LGD_{pi}$  measured on defaulted exposures belonging to the corresponding product  $P$  and the corresponding LGD grade  $i$ .



The product-level  $LGD_p$  falls for the given product within the interval assigned to the LGD grade 3.

- Under the scoring approach, **93%** of applications are assigned a **lower LGD** value and **7%** of applications a **higher LGD** value than under the product segmentation.
- The **average** of LGD values under the scoring approach is **lower by 29%** compared to under the product segmentation.

Ivana Žohová

Erste Group – Credit Risk Models

---

Česká spořitelna, a.s.

Olbrachtova 1929/62 · 140 00 Prague 4 · Czech Republic

E-mail: [IZohova@csas.cz](mailto:IZohova@csas.cz)

[www.csas.cz](http://www.csas.cz), [www.erstegroup.com](http://www.erstegroup.com)