State of the Art in Credit Risk Modeling

Prof. dr. Bart Baesens

Department of Decision Sciences and Information Management
KU Leuven (Belgium)

School of Management
University of Southampton (United Kingdom)
Bart.Baesens@kuleuven.be

YouTube/Facebook/Twitter: DataMiningApps
Presenter: Bart Baesens

- Studied at KU Leuven (Belgium)
  - Business Engineer in Management Informatics, 1998
  - PhD. in Applied Economic Sciences, 2003
- PhD.: Developing Intelligent Systems for Credit Scoring Using Machine Learning Techniques
- Professor at KU Leuven, Belgium
- Lecturer at the University of Southampton, UK
- Research: Big Data & Analytics, Credit Risk, Fraud, Marketing, ...
- YouTube/Facebook/Twitter: DataMiningApps
- www.dataminingapps.com
- Bart.Baesens@kuleuven.be
Books

- Analytics in a Big Data World
- Credit Risk Analytics: Measurement Techniques, Applications, and Examples in SAS
- Fraud Analytics: Using Descriptive, Predictive, and Social Network Techniques
Overview

• Introduction
• Credit Risk Components
• Basel versus IFRS 9
• Modeling impact
  – Survival analysis/mixture cure models
  – Discrimination versus Calibration
• Model monitoring
Strategic impact of credit risk analytics

• More than ever before, analytical models steer strategic decisions of financial institutions!
• Minimum equity (buffer capital) and provisions a financial institution holds are directly determined, a.o., by
  – credit risk models
  – market risk models
  – operational risk models
  – insurance risk models
  – …
• Analytics typically used to build all these models!
• Often subject to regulation (e.g. Basel II/Basel III, IFRS 9, …)!
• Model errors directly affect profitability, solvency, shareholder value, macro-economy, …, society as a whole!
Credit Risk Components

- **Probability of default (PD) (decimal):** probability of default of a counterparty

- **Exposure at default (EAD) (currency):** amount outstanding

- **Loss given default (LGD) (decimal):** ratio of the loss on an exposure due to default of a counterparty to the amount outstanding

- **Expected loss (Basel, IFRS9) =** PD x LGD x EAD

- **Unexpected loss (Basel) =** $f(PD, LGD, EAD)$
## Basel versus IFRS9

<table>
<thead>
<tr>
<th>Basel</th>
<th>IFRS 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default definition: 90 days payment arrears</td>
<td>No default definition</td>
</tr>
<tr>
<td>One year PD</td>
<td>Lifetime PD for stage 2 assets</td>
</tr>
<tr>
<td>TTC rating philosophy (focus on long run average PD)</td>
<td>PIT rating philosophy (focus on reporting date)</td>
</tr>
<tr>
<td>Downturn LGD (both direct + indirect costs)</td>
<td>Best estimate LGD (only direct costs)</td>
</tr>
<tr>
<td>Downturn EAD</td>
<td>Best estimate EAD</td>
</tr>
<tr>
<td>EL=PD<em>LGD</em>EAD</td>
<td>EL=PD*PV of cash shortfalls</td>
</tr>
<tr>
<td>Conservative calibration</td>
<td>-</td>
</tr>
<tr>
<td>Regulatory PD/LGD floors</td>
<td>-</td>
</tr>
</tbody>
</table>
Credit Risk Model Architecture

1. Create the model
   - application scorecard
   - behavioral scorecard

2. Define ratings and calibrate the model
   - PD calibration
   - risk ratings definition

0. Prepare the data
   - internal data
   - external data
   - expert judgment

PD Modeling
LGD Modeling
EAD Modeling
## Basel: Performance benchmarks

<table>
<thead>
<tr>
<th>Context</th>
<th>Number of Characteristics</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>PD Application Credit Scoring</td>
<td>10-15</td>
<td>AUC 70%-85%</td>
</tr>
<tr>
<td>PD Behavioural Credit Scoring</td>
<td>10-15</td>
<td>AUC 80%-90%</td>
</tr>
<tr>
<td>LGD</td>
<td>5-10</td>
<td>R-squared 15%-30%</td>
</tr>
</tbody>
</table>


Impact of IFRS 9: Survival analysis

• Predict timing of default
• Deal with censored data + time varying covariates (e.g. macro-economic fluctuations)
• Common techniques:
  – Kaplan-Meier analysis
  – Parametric survival analysis
  – Proportional hazards regression
  – Spline based models (complex!)
Mixture cure models
Mixture cure models

- Let $Y=1$ when an account is susceptible to default and 0 otherwise; let $x$ and $z$ be customer characteristics.

- Mixture cure model: $S(t|\mathbf{x},\mathbf{z}) = \pi(z) S(t|Y=1,\mathbf{x}) + 1 - \pi(z)$
  - $\pi(z) = P(Y=1|z)$: the incidence model component, modeled using e.g. logistic regression.
  - $S(t|Y=1,\mathbf{x})$: the latency model component, modeled using e.g. proportional hazards regression.

- Parameters can be estimated by formulating a ML function and optimizing it with the EM algorithm.

- See Dirick L., Claeskens G., Baesens B., 2015
Model discrimination versus Model calibration

• **Model discrimination**
  – Rank order (score) entities with respect to likelihood of event occurring
  – Despite traditional focus in credit risk, this is no longer sufficient!
  – We need to know the **EXACT** probability of the event occurring!

• **Model calibration**
  – Provide well-calibrated probabilities based on
    • Historical data
    • Expectations with respect to the future (e.g. GDP contraction versus expansion)
  – Example
    • $P(\text{Bart defaults})=0.90; P(\text{Victor defaults})=0.75$

**BRING THE MACRO-ECONOMY INTO THE MODEL!**
Model discrimination versus Model calibration

### Model Discrimination

<table>
<thead>
<tr>
<th>Characteristic Name</th>
<th>Attribute</th>
<th>Scorecard Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE 1</td>
<td>Up to 26</td>
<td>100</td>
</tr>
<tr>
<td>AGE 2</td>
<td>26 - 35</td>
<td>120</td>
</tr>
<tr>
<td>AGE 3</td>
<td>35 - 37</td>
<td>185</td>
</tr>
<tr>
<td>AGE 4</td>
<td>37+</td>
<td>225</td>
</tr>
<tr>
<td>GENDER 1</td>
<td>Male</td>
<td>90</td>
</tr>
<tr>
<td>GENDER 2</td>
<td>Female</td>
<td>180</td>
</tr>
<tr>
<td>SALARY 1</td>
<td>Up to 500</td>
<td>120</td>
</tr>
<tr>
<td>SALARY 2</td>
<td>501-1000</td>
<td>140</td>
</tr>
<tr>
<td>SALARY 3</td>
<td>1001-1500</td>
<td>160</td>
</tr>
<tr>
<td>SALARY 4</td>
<td>1501-2000</td>
<td>200</td>
</tr>
<tr>
<td>SALARY 5</td>
<td>2001+</td>
<td>240</td>
</tr>
</tbody>
</table>

### Model Calibration

**historical 12 month default rate credit rating B**

Historical probability of default (PD) calibration for customer segment B!

Example application scorecard
Model Calibration: example approach

• Analytical models typically built using a snapshot in time
• Cluster model outputs (e.g. scores) into pools\ratings
  – Scores are too fine granular anyway!
  – Essentially, a semi-supervised learning exercise
  – Score 200-300: pool A; score 301-500: pool B, score 501-650: pool C, ...
• For each pool, calibrate event probability using
  – Forecasting techniques (ARIMA, VAR, ...)
  – Dynamic models/Markov Chains
• Model transitions between pools
  – Gives an idea about customer volatility/model stability
  – Do I have a point-in-time (PIT) or through the cycle (TTC) model?
Summarising: Model architecture

- Calibration
  - Dynamic macro-economic models

- Discrimination
  - Scorecard

- Data
  - Internal/External Data
  - Expert Input

### Discrimination Scorecard

<table>
<thead>
<tr>
<th>Characteristic Name</th>
<th>Attribute</th>
<th>Scorecard Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE 1</td>
<td>Up to 26</td>
<td>100</td>
</tr>
<tr>
<td>AGE 2</td>
<td>26 - 35</td>
<td>120</td>
</tr>
<tr>
<td>AGE 3</td>
<td>35 - 37</td>
<td>185</td>
</tr>
<tr>
<td>AGE 4</td>
<td>37+</td>
<td>225</td>
</tr>
<tr>
<td>GENDER 1</td>
<td>Male</td>
<td>90</td>
</tr>
<tr>
<td>GENDER 2</td>
<td>Female</td>
<td>180</td>
</tr>
<tr>
<td>SALARY 1</td>
<td>Up to 500</td>
<td>120</td>
</tr>
<tr>
<td>SALARY 2</td>
<td>501-1000</td>
<td>140</td>
</tr>
<tr>
<td>SALARY 3</td>
<td>1001-1500</td>
<td>160</td>
</tr>
<tr>
<td>SALARY 4</td>
<td>1501-2000</td>
<td>200</td>
</tr>
<tr>
<td>SALARY 5</td>
<td>2001+</td>
<td>240</td>
</tr>
</tbody>
</table>
Side benefit: stress testing

• By introducing the macro economy into the model, one can do stress testing
  – “evaluate the potential impact on a firm of specific adverse events and/or movements in a set of financial variables” (BIS, 2005)
• Sensitivity analysis
  – Single variable versus multiple variables
• Scenario analysis
  – Historical or hypothetical
  – E.g. 3 successive years of GDP contraction, house prices drop by 5%, ...
• Common challenges/problems:
  – Lack of historical data
  – Correlations break down during stress
  – Integrate risks
  – What is stress??
Model Risk

• “Essentially, all models are wrong, but some are useful“ (George E. P. Box, 1987)

• Models are not perfect
  – PD: **good** performance (AUC around 80%-85%)!
  – LGD: **awful** performance (R² typically < 0.30)

• Model imperfection is typically dealt with by
  – Improving data quality
  – More powerful modeling techniques (?)
  – Conservative parameter calibration (Basel perspective!)
    – aka **economic downturn** calibration
Model monitoring

• Why PD/LGD/EAD models may degrade in performance?
  – Sample effects (models estimated on limited samples)
  – Macro-economy (downturn versus upturn)
  – Internal effects (e.g. strategy change, population drift, M&A)
  – In reality: a very nice (?) mixture of these!

• Need to constantly monitor outcomes of models

• Crucial since models more and more steer strategic decisions of the firm (Basel, IFRS 9)

• Quantitative versus Qualitative validation
Model validation

• Quantitative validation
  – Backtesting
  – Benchmarking

• Qualitative validation
  – Data quality
  – Model design
  – Documentation
  – Corporate governance and management oversight
Backtesting

• Contrasting ex-post realised numbers with ex-ante predictions
• Using statistical tests and performance measures
• Examples
  – Use binomial/Vasicek test for comparing default rates
  – Monitor decrease in AUC (Gini) over time
• Basel versus IFRS:
  – TTC (Basel): Backtesting should find that realized default rates vary around forecast PD (rising in downturns and falling in upturns)
  – PIT (IFRS): Backtesting should find that realized default rates are close to forecast PD
Backtesting Survival Analysis Models

• Statistical significance of both the model as well as the individual covariates
• Take a snapshot of the survival probabilities at a specific time $t$ (e.g., 12 months), compare with event time indicator and calculate ROC
  – Indicates how well the model ranks the observations for each
• Evaluate interpretability of model by using univariate sign checks on the covariates
# Backtesting: examples

<table>
<thead>
<tr>
<th>Score Range</th>
<th>Expected (training) %</th>
<th>Observed (actual) % at t</th>
<th>Observed (actual) % at t + 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-169</td>
<td>6%</td>
<td>7%</td>
<td>6%</td>
</tr>
<tr>
<td>170-179</td>
<td>10%</td>
<td>8%</td>
<td>7%</td>
</tr>
<tr>
<td>180-189</td>
<td>9%</td>
<td>7%</td>
<td>10%</td>
</tr>
<tr>
<td>190-199</td>
<td>12%</td>
<td>9%</td>
<td>11%</td>
</tr>
<tr>
<td>200-209</td>
<td>12%</td>
<td>11%</td>
<td>10%</td>
</tr>
<tr>
<td>210-219</td>
<td>8%</td>
<td>11%</td>
<td>9%</td>
</tr>
<tr>
<td>220-229</td>
<td>7%</td>
<td>10%</td>
<td>11%</td>
</tr>
<tr>
<td>230-239</td>
<td>8%</td>
<td>12%</td>
<td>11%</td>
</tr>
<tr>
<td>240-249</td>
<td>12%</td>
<td>11%</td>
<td>10%</td>
</tr>
<tr>
<td>250+</td>
<td>16%</td>
<td>14%</td>
<td>15%</td>
</tr>
<tr>
<td>SSI versus Expected</td>
<td>0.0605</td>
<td>0.494</td>
<td></td>
</tr>
<tr>
<td>SSI versus t - 1</td>
<td></td>
<td>0.0260</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of observations</th>
<th>Number of defaulters</th>
<th>AR</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR model</td>
<td>5866</td>
<td>105</td>
</tr>
<tr>
<td>AR 2006</td>
<td>5677</td>
<td>97</td>
</tr>
<tr>
<td>AR 2005</td>
<td>5462</td>
<td>108</td>
</tr>
<tr>
<td>AR 2004</td>
<td>5234</td>
<td>111</td>
</tr>
<tr>
<td>AR 2003</td>
<td>5260</td>
<td>123</td>
</tr>
<tr>
<td>AR 2002</td>
<td>5365</td>
<td>113</td>
</tr>
<tr>
<td>AR 2001</td>
<td>5354</td>
<td>120</td>
</tr>
<tr>
<td>AR 2000</td>
<td>5306</td>
<td>119</td>
</tr>
<tr>
<td>AR 1999</td>
<td>4970</td>
<td>98</td>
</tr>
<tr>
<td>AR 1998</td>
<td>4501</td>
<td>62</td>
</tr>
<tr>
<td>AR 1997</td>
<td>3983</td>
<td>60</td>
</tr>
<tr>
<td>Average AR</td>
<td>5111.2</td>
<td>101.1</td>
</tr>
</tbody>
</table>

SSI versus Expected: SSI = 0.0605, Expected = 0.494
SSI versus t - 1: SSI = 0.0260
Action plans

Model calibration

Model discrimination

Data stability

Continued using model

OK

NOT OK

Re-estimate model

Tweak model

Re-calibrate model

Data stability

OK

NOT OK

Re-estimate model

Tweak model

NOT OK

Re-calibrate model

Tweak model
Conclusions

• Introduction

• Credit Risk Components

• Basel versus IFRS 9

• Modeling impact
  – Survival analysis/mixture cure models
  – Discrimination versus Calibration

• Model monitoring
References


• See [www.dataminingapps.com](http://www.dataminingapps.com)
Books
Self-Paced E-learning course

Self-Paced E-learning course: Credit Risk Modeling

See: https://support.sas.com/edu/schedules.html?ctry=us&id=2455

The E-learning course covers both the basic as well some more advanced ways of modeling, validating and stress testing Probability of Default (PD), Loss Given Default (LGD) and Exposure At Default (EAD) models. Throughout the course, we extensively refer to our industry and research experience. Various business examples and small case studies in both retail and corporate credit are also included for further clarification. The E-learning course consists of more than 20 hours of movies, each 5 minutes on average. Quizzes are included to facilitate the understanding of the material. Upon registration, you will get an access code which gives you unlimited access to all course material (movies, quizzes, scripts, ...) during 1 year. The course focusses on the concepts and modeling methodologies and not on the SAS software. To access the course material, you only need a laptop, iPad, iPhone with a web browser. No SAS software is needed. See https://support.sas.com/edu/schedules.html?ctry=us&id=2455 for more details.