Knowledge Discovery through Risk Classification in Credit Portfolios

Eduardo Rodriguez T.
BSc Math, MBA, MSc Math. PhD
Principal IQAnalytics
Canada
• Introduction
• Objectives
• The problem
  – The simple model
  – LGD
• The real experience:
  – Some models
  – Not only models, where are the tools?
  – Digging deeper in what we know and what we can know
• Conclusions
• Utilitarianism Paradigm (Example health care and elderly)
• Probability concept
• Operations Research: Data Mining / Scoring / Classification
• The risk definition: Not only Loss but also we need to work on the Gain too. Not only work on the object but also on the perception
  • 50 years of data mining and OR: upcoming trends and challenges 2009 Journal of the Operational Research Society, B Baesens et al.
  • Comprehensible credit scoring models using rule extraction from support vector machines (2007) European Journal of Operational Research, David Martens et al.
From some of the most influent Professors in Risk Management

The Known, the Unknown, and the Unknowable in Financial Risk Management

Francis X. Diebold, Neil A. Doherty, and Richard J. Herring 2010
• Actuary is a risk problem solver in many industries
• The opportunities in quantitative world are growing
• Marketing and operations require more quantitative analysis
• Convergence with OR and IS-IT is higher than years before and highly needed
• Using techniques once only for Actuarial now in Operational Risk for example
• Changes in the bank business, opening to the search of strategies to reduce risk and/or increase returns

• The issue of changing business models and not showing in the balance sheets the assets creating what is called a SPV Special Purpose Vehicle or entity or SIV Structured Investment Vehicle
What I want to share based on the LGD problem

– Some words about the new settings in risk analysis-business
– Reflection about the capacity of models and the needs in other areas

What I expect you to think and learn after this talk

– Need and responsibility of actuary in risk management
– Development of new answers to current and future problems
• “Models are not decision makers; people are.” (Champion, 2009)
• “Models increased risk exposure instead of limiting it.” (The Economist, 2010)
• “So chief executives would be foolish to rely solely, or even primarily, on VAR to manage risk.” (The Economist, 2010)
• “Remember that the biggest risk lies with us: we overestimate our abilities and underestimate what can go wrong.” (Taleb et al., 2009)
• Risk as potential organisational losses and not on the organisational positive results or variance view of the outcomes; equally, managers concentrate more on the value of the loss than on the probability of the events with the observation that the attitude is that managers are not strongly oriented to measure the risk but to perceive it (March and Shapira, 1987)
From the EU agreement Solvency II:
“Solvency II cannot just be about capital requirements; no amount of capital can substitute for the capacity to understand, measure and manage risk and no formula or model can capture every aspect of the risks an insurer faces.”

and from Basel III
“reforms to strengthen global capital and liquidity rules with the goal of promoting a more resilient banking sector. The objective of the reforms is to improve the banking sector’s ability to absorb shocks arising from financial and economic stress, whatever the source, thus reducing the risk of spillover from the financial sector to the real economy”

- Review the legal. Regulation framework
  - New work on the capital side
- OR and Data Mining
  - Looking for new tools
- OR and Credit Scoring for profit planning
  - Looking for new applications
- Other disciplines
From Alan Forrest Group Risk Analytics Independent Model Validation RBS Group (2011)

“Model Risk assessment is an essential part of Credit Risk model development and validation.”

And additionally expressed:

“The risk that a model is built or implemented in a way that turns out to be unfit for purpose.
– How different could the model have been? How well will it hold up?

Can be caused by:
– limitations or bias in the data used to develop the model;
– changes in the meaning or population of the data over time;
– assumptions that define the structure of the model;
– assumptions about the market underlying the portfolio, past and future;
– uncertainty about model’s response to stress or downturn”
• What? How to estimate the parameters
  – Loss distribution
  – EL = PD \times LGD \times EAD

• How?
• How to use?
• How to communicate and create actions?
The PD problem has been more studied, is based on the obligor knowledge, a rating system, has data, is dichotomous and a figure.

The LGD is based on the transaction, can be a number or a distribution, related to human actions (recoveries), data are scarce and multiple complex factors involved, the standard deviation are very high, problem with the time to define the loss, multiple products…

The problem: What about the estimation of the parameters
• For the PD read “The value of non-financial information in SME risk management” Altman et al. 2009

• The statistical analysis will drive the best selection of mitigants:
  – Financial Collateral
  – Non-Financial Collateral
  – Guarantees
  – Credit Derivatives
A percentage of Recovery
A loss distribution
Segments with different behaviour
Adverse selection
Crisis vs. No crisis
Risk Management Knowledge Systems
Relationship between variables & operations
Cut offs
Portfolio optimization with more restrictions
The problem:
How to measure long run LGD

• Historic data
  – Use the averages from the historical results
  – Take into considerations the value of the assets-
    Exposure through the history
  – Review the portfolio concentration through time

• The future LGD
  – Use the indicators and use time series for
    supporting the future estimation
In the financial crisis for example most of the assumptions related LGD proved faulty. It was thought that the LGD for mortgages would be low. The reason, it was expected no the lost of houses value.
The problem: how to start, how to proceed, what to use...

• Definitions
• Data
• Models
• Interpretation
• Reviews, tests
• Understanding
• Again evaluation
• Segmentation
The problem: The real problem...

If people do not believe that mathematics is simple, it is only because they do not realize how complicated life is. ~John Louis von Neumann
Loans and options
Pay the loan when the assets value is higher than the credit amount and to default in case of assets value is smaller than loan
• The literature refers to two kinds of models for PD
  • Structural default prediction models (Moody’s KMV Loss Calc)
  • Reduced form Models (Kamakura’s risk manager)
  • Kamakura’s model has some advantages: one because of easier calculation and second for including other variables
• The KMV model is based on the stock price volatility and the stock price. No applicable for private companies, possibly no probability of default.

• There are two concepts DD Distance to Default and Expected Default Frequency
Kamakura’s Risk Manager: Use debt and other security prices to describe default probabilities

• This is an answer to avoid the assumption of the LGD as constant or proportional to the bond value

• Based on the article of Chava and Jarrow (2004) plus some pieces of Merton model

• This includes the firm specific information, industry information, economic environment, and macroeconomic factors
The real experience: Types of models

- Credit scoring systems: example Altman
- Mortality rate systems: tables with bonds mortality rates and with a combination to calculate LGD and EL
- Neural Networks
• A data base where 1% of the obligors are possible to analyze with the previous models (Working with companies around the world)
• Most of the companies are not public they are private.
• A need for testing other variables. Which are significant (Example years in business, experience of the CEO…) To remark in some cases with more contribution than the financial variables
The real experience How to know more about LGD?

• Obligor size, market information
• Economic sector and country
• Type of firm (industry)
• External-internal information of the organization
• Size of loan
• Firm specific characteristics
  – E.g. revenue, total assets, net worth, total debt / capital at default, current liability/total liability at one year before default, ...
• Economic conditions
  – E.g. GDP growth, average default rates, inflation rates, ...
• Country-related features
  – geographical region (e.g. zipcode)
  – e.g. how creditor-friendly is the bankruptcy regime?
• Seniority of debt relative to other creditors
The real experience: Classification using other variables

- Segmentation
- Seniority of debt
- Legal systems
- Credit rights
- Behavior
The purpose is to include variables that can affect the LGD, mainly those related to the country.

Understanding and learning from the experience. Is it possible to observe behaviors that support the loss control?

This is a first step in a learning curve and continuous improvement process.

Better understanding produces better segmentation and pricing allocation.

The variables applied to the database are:

- Data from "Private Credit in 129 Countries" Djankov, Simeon, Caralee McLiesh and Andrei Shleifer Journal of Financial Economics, 2007
- Debt enforcement database
• Identification of variables that affect LGD
  – Recovery %
    – Common v Civil Law
    – Macro Variables (GDP & Inflation)
    – Industry Sectors
  – Recovery from Insolvency
    – Creditors Rights in Bankruptcy
  – Legal Uncertainty
    – Days to Resolution
    – Recovery Forecast Errors
The real experience
Most of the Kernel distributions were like these, an issue for % description

Recovery: Rest of World (Not CAN & USA) - By Area

Graphs by world0area

The real experience
Most of the Kernel distributions were like these, an issue for % description
1. Common Law - Significant
2. Inflation – Not Significant
3. GDP Growth – Not Significant
4. Paid value $ - Not Significant
5. Original Estimate of Recovery – Significant
6. Years in Business (Positive Significant)
7. Cash Assets (Not Significant)
8. Asset Size( Significant Negative – (Smaller Firm more recovery)


The real experience:
looking just for significance of the variables
Desmitifying R-square or others meausres. Analyzing residuals
9. Covered Years (Negative and Significant) – fewer years more recovery
10. Market Sector – Not Significant
11. Overdue Code – Not Significant
12. Number of Invoices - (Negative and Significant) More invoices – recover less
13. Terms – Not Significant
Analyzing time series for risk indicators

<table>
<thead>
<tr>
<th>years</th>
<th>Transportation</th>
<th>Resources</th>
<th>Light Manufacturing</th>
<th>Infrastructure and Environment</th>
<th>Info. &amp; Com Technology</th>
<th>Extractive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998</td>
<td>1.13%</td>
<td>1.04%</td>
<td>1.77%</td>
<td>0.83%</td>
<td>1.32%</td>
<td>1.07%</td>
</tr>
<tr>
<td>1999</td>
<td>0.81%</td>
<td>1.03%</td>
<td>2.08%</td>
<td>1.36%</td>
<td>1.35%</td>
<td>0.63%</td>
</tr>
<tr>
<td>2000</td>
<td>1.68%</td>
<td>1.46%</td>
<td>2.11%</td>
<td>1.17%</td>
<td>1.98%</td>
<td>0.49%</td>
</tr>
<tr>
<td>2001</td>
<td>1.54%</td>
<td>1.29%</td>
<td>2.34%</td>
<td>1.51%</td>
<td>1.90%</td>
<td>1.37%</td>
</tr>
<tr>
<td>2002</td>
<td>1.17%</td>
<td>1.32%</td>
<td>2.35%</td>
<td>1.38%</td>
<td>1.74%</td>
<td>0.95%</td>
</tr>
<tr>
<td>2003</td>
<td>0.54%</td>
<td>1.12%</td>
<td>1.68%</td>
<td>1.11%</td>
<td>1.23%</td>
<td>0.73%</td>
</tr>
<tr>
<td>2004</td>
<td>1.19%</td>
<td>1.03%</td>
<td>1.14%</td>
<td>0.74%</td>
<td>1.06%</td>
<td>0.43%</td>
</tr>
<tr>
<td>2005</td>
<td>0.85%</td>
<td>1.03%</td>
<td>1.17%</td>
<td>0.67%</td>
<td>0.85%</td>
<td>0.21%</td>
</tr>
<tr>
<td>2006</td>
<td>0.69%</td>
<td>0.81%</td>
<td>1.27%</td>
<td>0.75%</td>
<td>0.66%</td>
<td>0.58%</td>
</tr>
<tr>
<td>2007</td>
<td>0.79%</td>
<td>0.86%</td>
<td>1.27%</td>
<td>0.68%</td>
<td>0.85%</td>
<td>0.50%</td>
</tr>
<tr>
<td>2008</td>
<td>1.53%</td>
<td>1.53%</td>
<td>2.28%</td>
<td>0.98%</td>
<td>1.21%</td>
<td>1.03%</td>
</tr>
<tr>
<td>2009</td>
<td>1.08%</td>
<td>1.20%</td>
<td>1.62%</td>
<td>0.84%</td>
<td>1.02%</td>
<td>0.76%</td>
</tr>
</tbody>
</table>
Developing means to identify segments for analysis; for instance, using classification trees.
The real experience:

"Only professional mathematicians learn anything from proofs. Other people learn from explanations." — Ralph Boas
• Application Main Menu

• Options According to the Audience

The real experience: Creating tools to know?
• Search of thresholds, early warning systems...
• Premium
• Expected losses
• Claims
• Exposure
• By regions, products, sectors, risk levels
### DISTRIBUTION ANALYSIS

#### Exposure and Change by Period

<table>
<thead>
<tr>
<th>Excess</th>
<th>June 2010</th>
<th>Quarter</th>
<th>6 Months</th>
<th>12 Months</th>
<th>June 2011</th>
<th>Quarter</th>
<th>6 Months</th>
<th>12 Months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>28,880,519</td>
<td>62.94%</td>
<td>15.30%</td>
<td>0.48%</td>
<td>28,880,519</td>
<td>62.92%</td>
<td>15.26%</td>
<td>0.45%</td>
</tr>
<tr>
<td>Top 100</td>
<td>18,951,286</td>
<td>65.21%</td>
<td>16.95%</td>
<td>0.52%</td>
<td>18,951,286</td>
<td>65.20%</td>
<td>16.94%</td>
<td>0.51%</td>
</tr>
<tr>
<td>1-25</td>
<td>6,863,624</td>
<td>23.74%</td>
<td>9.25%</td>
<td>0.34%</td>
<td>6,863,624</td>
<td>23.71%</td>
<td>9.24%</td>
<td>0.33%</td>
</tr>
<tr>
<td>26-50</td>
<td>4,899,236</td>
<td>17.02%</td>
<td>5.37%</td>
<td>0.25%</td>
<td>4,899,236</td>
<td>17.01%</td>
<td>5.36%</td>
<td>0.24%</td>
</tr>
<tr>
<td>51-100</td>
<td>349,527</td>
<td>1.21%</td>
<td>0.73%</td>
<td>0.03%</td>
<td>349,527</td>
<td>1.21%</td>
<td>0.73%</td>
<td>0.03%</td>
</tr>
<tr>
<td>101-250</td>
<td>128,742</td>
<td>0.45%</td>
<td>0.45%</td>
<td>0.02%</td>
<td>128,742</td>
<td>0.45%</td>
<td>0.45%</td>
<td>0.02%</td>
</tr>
<tr>
<td>Total and Change by Period</td>
<td>28,880,519</td>
<td>62.94%</td>
<td>15.30%</td>
<td>0.48%</td>
<td>28,880,519</td>
<td>62.92%</td>
<td>15.26%</td>
<td>0.45%</td>
</tr>
</tbody>
</table>

#### Distribution by Risk

- **High:** 33.30%
- **Medium High:** 26.46%
- **Medium:** 20.29%
- **Medium Low:** 10.09%
- **Low:** 9.95%

#### Summary

- **June 2010:** Total Exposure $28,880,519, with 62.94% in the Top 100, 23.74% in 1-25, 17.02% in 26-50, and 1.21% in 101-250.
- **June 2011:** Total Exposure $28,880,519, with 62.92% in the Top 100, 23.71% in 1-25, 17.01% in 26-50, and 1.21% in 101-250.

#### Prospective Premiums

- **Prospective Premium:** $8,918,300 (6.94% of Total Exposure 12 Months).

---

**Note:** The above data represents a snapshot of the distribution analysis for June 2010 and June 2011, highlighting the exposure and change by period, as well as the distribution by risk. This analysis helps in comparing results and creating control systems for the main risk indicators.
• Knowing how the portfolio evolves

• Example of evolution through transition matrices

The real experience:
...tools...Understanding a portfolio evolution

Risk Levels Transition Matrices

Current Transition Matrix

Historical Transition Matrices

Forecast Matrix (n+1)-period

<table>
<thead>
<tr>
<th>Probabilities Transition Matrix June 09 to December 09</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOW TO</td>
</tr>
<tr>
<td>LOW</td>
</tr>
<tr>
<td>MOD</td>
</tr>
<tr>
<td>HI</td>
</tr>
<tr>
<td>CRI</td>
</tr>
<tr>
<td>PRI</td>
</tr>
</tbody>
</table>
The real experience: Additionally doing a work on Exposure Management

- LGD: Current and comparison capability
- Supporting Qualitative Analysis Knowing more about specific cases
• Integrating numbers and interpretation

• Creating the road orientation according to the key questions

The real experience: Creating means to communicate effectively:
See work from Rodriguez and Edwards and SerenKo

Qualitative Analysis

- Concentration-Clusters
- Changes to remark
- Expected Results

Standards and Scoring Team
• Definition of LGD
• Dependent variable continuous between 0% and 100, or 0% or 100%
• Review the judgemental considerations
• Use segmentation and regression models
• Analyzing the bimodal distribution
• Need loss estimates for bad accounts
• Decide on time horizon
  – Survival analysis
• Dynamic models: incorporate changes in economic factors, cycles
• Organize scores for profit analysis and look for classification based on Expected losses
Models for data that is not with continuous pricing is requires.
The need of fitting loss distributions is clear.
The validation and modification of assumptions is required:
  - Recovery rate is not normal distribution.
  - Beta distribution is better fit (Moody’s LossCalc assumes this).
  - Regression assumes errors (and hence dependent variable) is normal.

The real experience: Digging deeper in what we know and what we can know.
• Fit Loss Distribution
• Use Mixed Loss Distributions
• Review the areas of risk and segments of obligors
• Study the quantiles and moments, indicate the levels of concentration
• Analyzing at least three groups: 0 LGD, in the interval (0,100) % and 100% LGD
• Use logistic regression for 0 and 100% LGD and linear regression for the interval (0,100)%
• Use segments for better approach and estimation
A life actuary designed a new coverage “Senility Insurance”. He expected low claims because "If you remember that you have a policy, it is proof that you are not senile."
Virtually no correlation between LGD of secured exposures and business cycle!

The real experience:
Digging deeper in what we know and what we can know

Recovery rate/default rate association
Altman defaulted bonds data set (1982-2000)
Dollar weighted average recovery rates to dollar weighted average default rates

\[ y = -2.617x + 50.9 \]
\[ R^2 = 0.4498 \]

\[ y = 0.5609x^2 - 8.7564x + 60.61 \]
\[ R^2 = 0.6091 \]

\[ y = 52.739 \times ^{-0.2854} \]
\[ R^2 = 0.6004 \]

Altman, Brady, Resti, Sironi 2003
• Correlation of PDs across obligors
• Correlation between PD and LGD
• Correlation between PD and EAD
• Impact on e.g. stress testing
• Time series analysis of the indicators
• Developing indicators with data-mining techniques
• Use perception surveys and develop indicators
• Connect the organization’s performance evaluation and risk indicators
• Use testing
• Data quality
• Model design
• Documentation
• Corporate governance and management oversight
• Cleaning data
• Observing segments, observing the descriptive statistics
• Looking for fitting models for the loss distribution and for the % of recovery
• Looking for clusters, models with machine learning (trees)
• Loan information includes data: demographic, collateral, guarantee rates
• Guarantee exposure, guaranteed LGD
• Assessment about contracts, recoveries effectiveness
• Tables of assumptions: what are the percentage of guarantees in the portfolio and loan
• LGD by portfolio percentile
• Compliance: only what regulations say to do
• Not to use the body of knowledge in lessons learned and from information that is accumulated
• Lack of risk culture
• Poor adaptation to the complexity of financial instruments
The real experience
Expected Loss versus LGD

- In summary we need to know how to classify using the expected loss or better the mix between PD and LGD

<table>
<thead>
<tr>
<th>PD\LGD</th>
<th>High</th>
<th>Average</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>8.5%</td>
<td>6%</td>
<td>2.0%</td>
</tr>
<tr>
<td>Medium</td>
<td>4.75%</td>
<td>2.0%</td>
<td>1.1%</td>
</tr>
<tr>
<td>Low</td>
<td>1.15%</td>
<td>1.15%</td>
<td>0.52%</td>
</tr>
</tbody>
</table>

Expected Loss evaluation

- good
- bad
Conclusions

1. Many things that we know and need to understand better
2. Many things that we have to find a way to know
3. Many things that we need to dig deeper to improve the knowledge
4. Many things that are together not only modeling, we need systems, we need to communicate, we need to develop more competencies
5. We have to find more and better relationships with other disciplines
• Journal of Risk Management in Financial Institutions
• Journal of Credit Risk
• Journal Insurance Mathematics and Economics
• Journal of Banking and Finance
• Risk Magazine
• Journal of the Operational Research Society
• European Journal of Operational Research
• Management Science
• IMA Journal of Mathematics
References

- Rodriguez E. Estrategia en el Mercado de Servicios Financieros -Strategy in Financial Services Book to be published 2011
- Book - ISBN 978-958-8307-08-4
- Rodriguez E. (2005) Gerencia del Conocimiento y Pensamiento Estratégico (Knowledge Management and Strategic Thinking)
• http://www.crc.man.ed.ac.uk/conference/
• www.defaultrisk.com
• http://www.bis.org/
• Websites of regulators:
  – www.fsa.gov.uk (United Kingdom)
  – www.hkma.gov.uk (Hong Kong)
  – www.apra.gov.au (Australia)
  – www.mas.gov.sg (Singapore)
• In the U.S.
Thank you very much!!

Eduardo Rodriguez T. PhD
Principal IQAnalytics
Ottawa, Ontario Canada
www.iqanalytics.com
613-2984978/ 613-8000674
8-2049 Baseline RD K2C 0C8
eduardo.rodriguez@iqanalytics.com