Credit Research Centre
Credit Scoring and Credit Control IX

7-9 September 2005

The University of Edinburgh Management School

"Automatic modelling of credit risk through internal ratings: An application of advanced statistical models and Machine Learning techniques"

Raquel Flórez-López

Department of Economics and Business Administration

University of León (SPAIN)

E-mail: dderfl@unileon.es

(*For additional information or comments, please contact the author*)

Overview

Introduction

- The **Basel II Capital Agreement (2004)** provides a 'risk-sensitive' framework for credit, market and operational risks management in banks.

- About **credit risk**, two broad methodologies are proposed :
  - ***Standardised Approach***: Supported by external credit assessments. Fixed credit weights are established for 13 different categories of credit exposures.
  - ***Internal Rating-based (IRB) Approach***: Supported by internal ratings and own estimates of risk components (PD, LGD, EAD, M). Given risk componentes, IRB risk-weight functions transform them into risk-weighted assets.

- The **IRB Approach** also includes two variants:
  - ***Foundation IRB***: Banks provide their own estimates of PD and rely on supervisory estimates for other risk components.
  - ***Advanced IRB***: Banks provide more of their own estimates of PD, LGD and EAD.

Introduction

Exposure	IRB - FOUNDATION		IRB - ADVANCED	
	Internal estimates	Supervisory estimates	Internal estimates	Supervisory estimates
Corporate, sovereign and bank exposures	PD	LGD, EAD, M	PD, LGD, EAD, M	-
Retail exposures	PD, LGD, EAD	M	PD, LGD, EAD	M
Equity exposures	Potential loss (*market-based approach*) PD, LGD (*PD/LGD approach*)			

- The estimation of **PD** is a crucial credit component for the development of internal credit risk models (BCBS, 2004, pgh. 391).
- To use internal PD estimations, there are some requirements :
  - ***Corporate, Sovereign, and bank exposures***: Banks may use appropiate techniques when they estimate PD : **internal default experience**, **mapping to external data**, **and statistical default models**
  - ***Retail exposures***: Internal data must be considered the primary source of information for estimating PD.

Introduction

- **PD estimates for Corporate, Sovereign and Bank Exposures:** (BCBS, 2004; Martin-Marin and Trujillo-Ponce, 2004):

INTERNAL DEFAULT EXPERIENCE:
Description: Historical database on banks' clients.
Advantages: It provides a specialised PD estimation.
Limitations: A large database is required. Historical records are usually not complete.

SHORT-TERM SOLUTION

MAPPING TO EXTERNAL DATA:
Description: Banks map their internal grades to the scale used by an external credit assessment institution. Defaults rates for the external grades are attributed to the banks' grades.
Advantages: Easy implementation. Banks benefit of external agencies' experience.
Limitations: Small number of external qualified companies. Mapping must be based on a comparison of internal and external rating criteria.

STATISTICAL DEFAULT MODELS:
Description: Banks are allowed to use a simple average of default-probabilities for individual borrowers in a given grade, drawn from statistical default predicion models.
Advantages & Limitations: Directly derived of the selected default models
This alternative really supports previous approaches.

Introduction

- **Mapping to External Data:**
  - **Multiple concordance points between IRB and external ratings determinants** (BCBS, 1999, 2000; Brunner et al., 2000; Crouhy et al., 2001; Grunert et al., 2005).
  - **Many recent literature attempts to understand rating determinants for industrial companies and banks** (Blume et al., 1998; Tabakis and Vinci, 2002; Estrella et al., 1999; Amato and Furfine, 2004; Altman and Rijken, 2004).
  - **Scarce analysis on insurance fims, BUT…**
    - Insurance firms are key agents on financial markets.
    - The more insurers expands their business, the more important becomes the evaluation of their creditworthiness by independent agencies (Chorafas, 2004).
    - The importance and specificity of the insurance industry affects external rating processes (Buckley, 1997).
    - New regulatory environments affect insurance firms, particularly in the European Union: SOLVENCY II Project.

Introduction

- The **Solvency II (2002-2003; 2003-2006) project** proposes advanced measures in order to deep reform the solvency system of European insurance companies (KPMG 2003, IMSDG, 2005).
- **External ratings** (financial strength ratings) are considered for assisting supervisors in prudential supervision.
- Nevertheless, their potential assistance is conditioned to the comprehensibility of methodologies and ratings determinants employed by external rating agencies.

Considering both Basel II and Solvency II requirements, it is convenient a better understanding of the insurance rating process by modelling financial strength rating determinants

Overview

Researches on external credit rating determinants: a literature review

- **Industrial corporations and banks:**
  - **Explanatory variables:** Profitability, capitalization, liquidity, interest coverage, debt status, and industry indicators.
  - **Techniques:** Linear regression, MDA, multinomial logit, ordered logit, probit, ordered probit, rule-based systems, neural networks, other Machine Learning techniques.

- **Blume et al (1998),** US large and small corporations; ordered probit; accounting and market data; predictive accuracy: 57%, 97% (±1 notch).

- **Estrella et al. (1999),** US banks; logit model; capital ratios.

- **Tabakis and Vinci (2002),** European banks; ordered-probit; profitabitlity, efficiency, capital adequacy, liquidity, loan intensity ratios; adjusted-R^2:59.41%.

- **Amato and Furfine (2004),** US corporations; ordered probit; business risk, financial risk and macroeconomical indicators; predictive accuracy: 53%.

- **Altman and Rijken (2004),** US corporations; logit and ordered logit; financial ratios, market data, size and age information; pseudo-R^2: 21-23%.

Researches on external credit rating determinants: a literature review

- **Insurance firms**
  - **Explanatory variables:** Capitalization factors, asset and liability measures, liquidity, size, growth, diversification, profitability, reinsurance usage.
  - **But very limited literature on the determinants of financial strenght ratings**

- **Pottier (1997),** US life insurers; ordered logit and naïve model; liquidity risk, investment risk, operating risk and financial risk measures.

- **Pottier and Sommer (1999),** 1,687 US property-liability insurer ratings; ordered-probit; capital, liquidity, investment risk, reinsurance, size, leverage, growth, profitability, percentage of business in long-tail lines, geographical diversification, line-of-business diversification ratios; differences between agencies' determinants.

- **Burton et al. (2003),** 335 UK insurers; ordered-probit; profitability, liquidity, organisational form (A.M. Best); leverage, profitability, liquidity (S&P).

- **Gaver and Pottier (2005),** 80 property-liability US insurers; ordered logit; capitalization, liquidity, profitability, size.

Researches on external credit rating determinants: a literature review

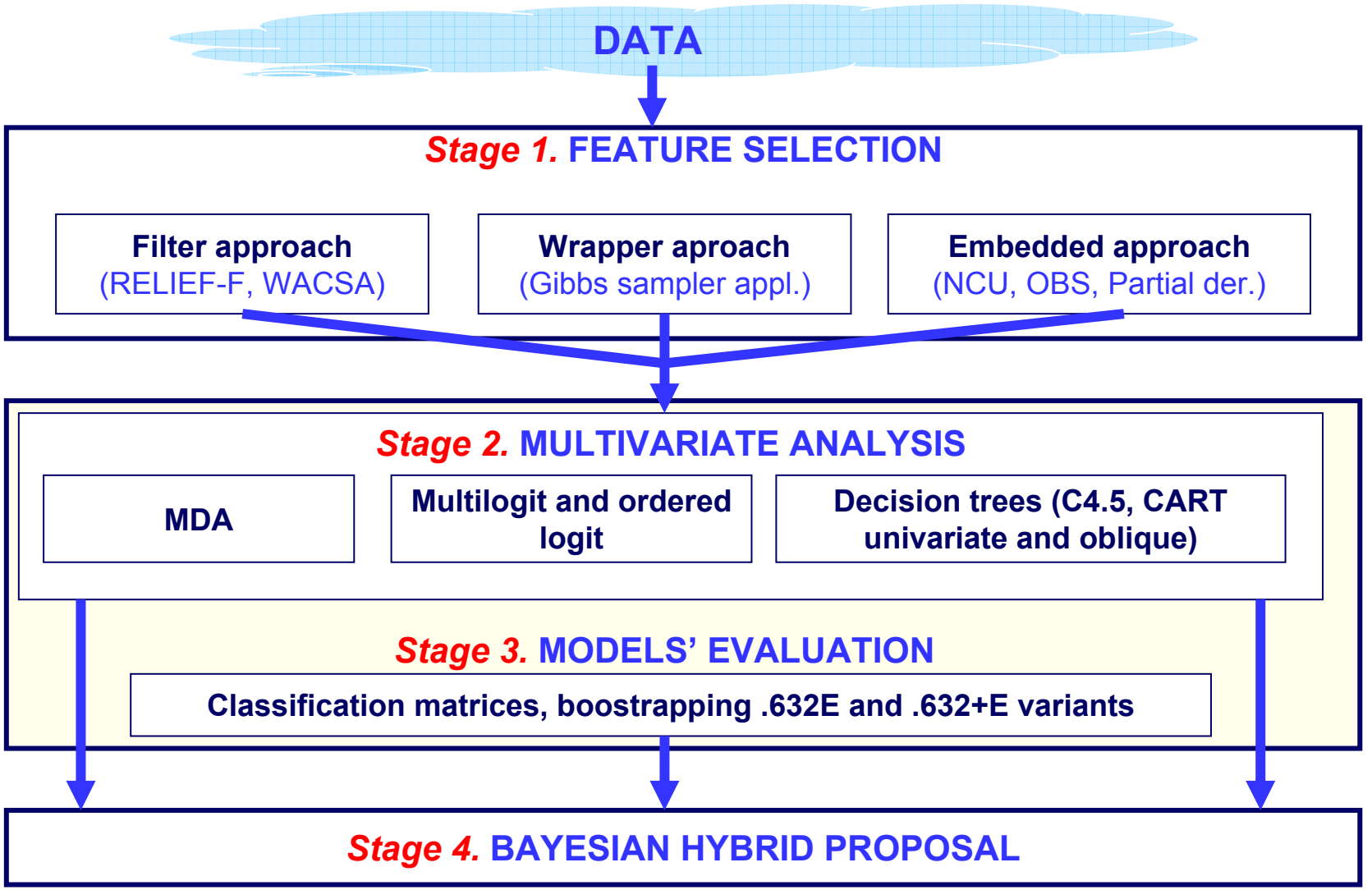
- **Insurance firms**

A.M. Best	Moody's	Standard & Poor's
Business profile, Management and strategy, Operating performance, Investment portfolio, Capitalization	Competitive situation, Regulatory trends, Adequacy of equity capital, Investment risk, Profitability, Liquidity, Group interrelationships, Products and distribution channels, Quality of management and organization, Others (spread of risk, loss reserves, solvency margin, reinsurance program).	Industry risk, Management and corporate strategy, Business review, Results from underwriting, Investment policy and results, Interest rate risk management, Capitalization, Liquidity, Capital and capital requirements

Overview

A proposal for modelling rating determinants

- A four-stage model:



A proposal for modelling rating determinants

Stage 1. FEATURE SELECTION

- **Target:** Finding a strongly relevant subset of the original potential determinants that ideally is necessary and suficient to describe the financial strength rating process.

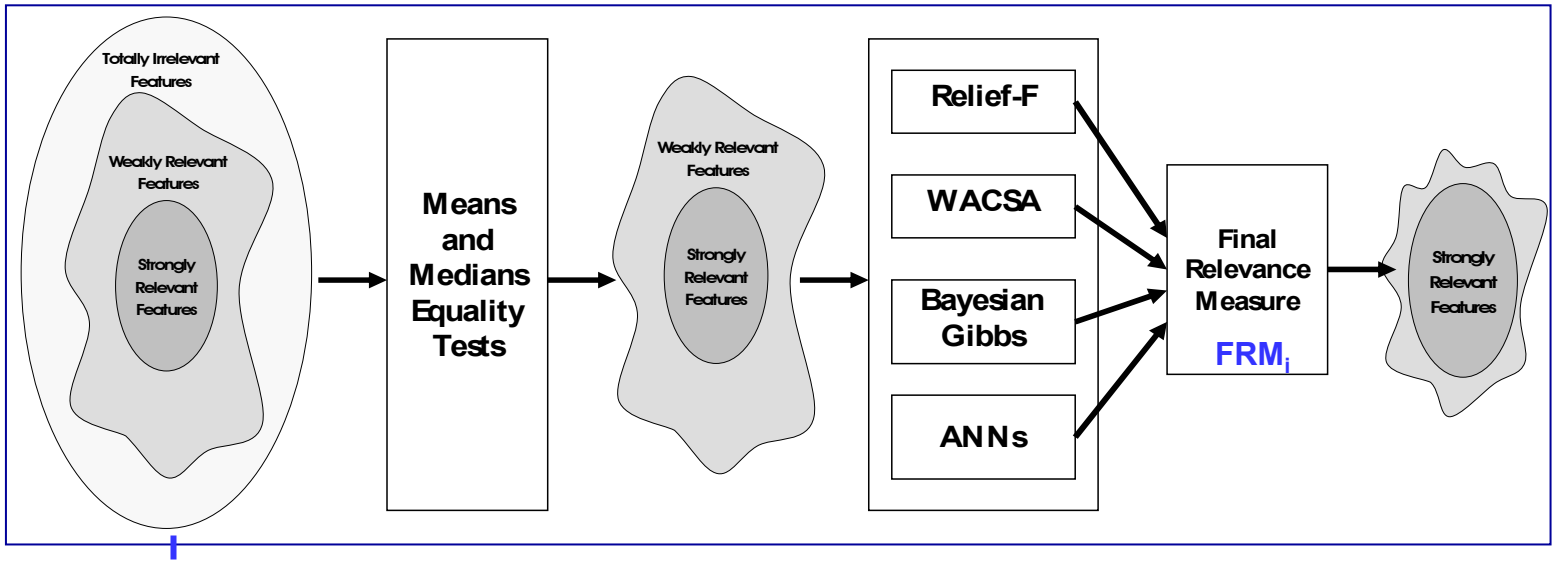
- **Statistical techniques for feature selection:** ANOVA analysis, tests on equality of means and medians, Principal Component Analysis, cluster analysis, etc. (Duran and Odell, 1974; Funaga, 1990; Ben-Bassat, 1992; Hair et al., 1998).

- **Machine Learning techniques for feature selection:** (John et al., 1994;Dash and Liu, 1997)
  - **FILTER METHODS**: Independence of classifiers
    - **Relief-F**
    - **WACSA**
  - **WRAPPER METHODS**: Dependence of classifiers
    - **Bayesian Gibbs sampler MCMC approaches**
  - **EMBEDDED METHODS**: Integration within classifiers. Some **ANN** models:
    - **Non-Contributing Units**
    - **Optimal Brain Surgeon**
    - **Partial Derivative Approximation**

A proposal for modelling rating determinants

Stage 1. FEATURE SELECTION

- **Proposal:** A hybrid selection procedure on five steps



Step 1.5. An ordered list of FRM_i measures is obtained. The final subset of attributes is selected using the cut-point $FRM_i \geq k.FRM_{i+1}$ ($k=2$ *or* $k=3$).

A proposal for modelling rating determinants

Stage 2. MULTIVARIATE ANALYSIS

- **Multiple Discriminant Analysis:**

$$Z_{im=} = \alpha_m + \sum_{k=1}^{K} \beta_{mk} X_{ik}, \quad i = 1,..,N, \quad m = 1...,M-1$$

- **Multinomial Logit Regression:**

$$\ln\left(\frac{\pi_{im}}{\pi_{i1}}\right) = \alpha_m + \sum_{k=1}^{K} \beta_{mk} X_{ik} = Z_{im}, \quad i = 1,..,N$$

where:

π_{im} is $P(Y = m|X)$

$$P(Y_i = m) = \frac{\exp(Z_{im})}{1 + \sum_{h=2}^{M} \exp(Z_{ih})}$$

$$P(Y_i = 1) = \frac{1}{1 + \sum_{h=2}^{M} \exp(Z_{ih})}$$

- **Ordered Logit Model:**

$$P(Y_i = 1) = \Lambda(\mu_0 - Z_i) - \Lambda(\mu_1 - Z_i),$$
$$\ldots$$
$$P(Y_i = M) = \Lambda(\mu_{M-1} - Z_i) - \Lambda(\mu_M - Z_i),$$

where:

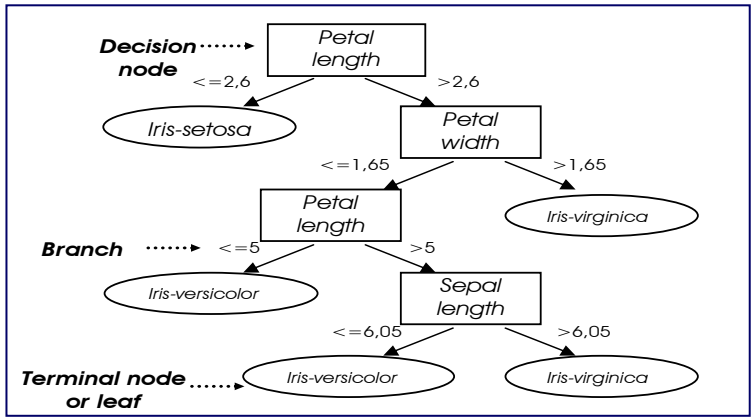
$$Z_i = \sum_{k=1}^{K} \beta_k X_{ik} + \varepsilon_i,$$

Λ is the cumulative logistic function

A proposal for modelling rating determinants

Stage 2. MULTIVARIATE ANALYSIS

- **Decision trees:** Hierarchical and sequential classification models that recursively divide a set of data (X), using univariate or oblique partitions.



- ***CART algorithm*** (Breiman et al., 1984)
  - Both **univariate** and **oblique** partitions can be used.
  - Two different split criteria (**Gini criterion** and **Twoing index**), **pruning strategy**

- ***C4.5 algorithm*** (Quinlan, 1993)
  - **Univariate** partition trees.
  - Two different split criteria (**gain criterion** and **gain ratio criterion**) based on CLS (Hunt, 1962), **stopping criterion based on** χ^2 and **pessimistic-error-based pruning criterion**.

A proposal for modelling rating determinants

Stage 3. MODELS' EVALUATION

- **Several measures have been proposed:** classification matrix; percentage of correctly classified observations; 1-notch, 2-notches or 3-notches difference measures; AIC, adjusted R^2 (Cantor and Packer, 1996; Blume et al., 1998; Hosmer and Lemeshow, 2000; Afonso, 2003).

- **Proposal:**
  - **Classification matrices and percentage of correctly classified observations.**
  - **Soft classification error measure:** Predicted rating minus one (AA, A) or plus one (BBB, BB, B)
  - **Boostrapping estimates** of real classification errors: **.632E** and **.632+E** strategies (Efron, 1983; 1986; Efron and Tibshirani, 1986; 1993; 1995)

$$\hat{E}rr_{0.632E} = 0.368 \times \overline{err} + 0.632 \times E_0$$

$$\hat{E}rr_{0.632+E} = (1 - w) \times \alpha + w\varepsilon$$

where:

$$R = (\varepsilon - \alpha)/(\gamma - \alpha)$$

$$\gamma = \sum_l p_l(1 - q_l)$$

$$w = 0.632/(1 - 0.368R)$$

where $\overline{err}$ is the error on the training set (resubstitution error) and E_0 is the bootstrap's average error on training vectors not included in each bootstrap sample.

A proposal for modelling rating determinants

Stage 4. BAYESIAN HYBRID PROPOSAL

- **Target:** To obtain **synergic effects** from multivariate methods and results.
- **Justification:** Due to different techniques capture diverse characteristics and variability from data, **their combination on a unique net** would offer a hybrid solution that could outperform individual results.
- **State:** Currently **in progress** (*not included in this paper*).

Overview

Empirical Application

Data and Sample

- **Data on European non-life insurers' financial strength ratings**
- **Standard and Poor's 'pi' ratings**
  - Based on the analysis of **published financial information** and additional public infomation.
  - They **do not reflect in-depth meetings** with insurer's managers.
  - They **are not applied** for companies.
  - **Annually reviewed** based on a new year's financial statements (or shortly if major events).
  - A principle of **moderate conservatism** applies.
  - **Analytical methods nearly identical** to those applied for interactive ratings.
- **Final sample:**
  - **257 European non-life insurance companies** from 14 EU countries.
  - All companies were regulated by the **92/49/EEC Directive**.
  - All companies were **'pi' rated** on **Jan'99 – Aug'00** from the last financial statements.
  - All companies have available **individual financial data for the four last years** before rating.
  - Each company is included **only one time** in the dataset (last 'pi' rating).
  - **Five rating grades** were considered: **AA, A, BBB, BB, and B.**

Empirical Application

Data and Sample

- **Independent variables (potential rating determinants)** (based on both previous literature and S&P public information on ratings' methodology):

AREA	EXPLANATION	N. FEATURES
Exposure (EXP)	Capital adequacy requirements	8 ratios
Reassurance (REA)	Use of reassurance	2 ratios
Liquidity (LIQ)	Assets structure and firms' solvency	11 ratios
Profitability (PROF)	Firm's profit and equity structure	21 ratios
Reserves (RES)	Technical provisions and reserves	7 ratios
Debt structure (DEB)	Debt composition	2 ratios
Size (DIM)	Firm's dimension	3 ratios
Industry (IND)	Each country's particular characteristics for the insurance industry	8 ratios
TOTAL		**62 ratios**

- Ratios were calculated on a **1-year basis**, **2-year average basis** and **3-year average basis: 186 potential features**

Empirical Application

Feature Selection

- **Step 1.1.:** From **186 features**, parametric and non parametric tests on equality of means and medians: 42 rejected features
- **Step 1.2.:** From **144 features**, Relief-F (R), WACSA (W), Bayesian-Gibbs (B) and ANN (A) methods were applied.
- **Step 1.3.** Each relevance indicator was **normalised** in [0,1]
- **Step 1.4.** A **Final Relevance Measure (FRM)** index was obtained for each feature
- **Step 1.5.** A **cut-point** was selected (***k=1.067***)

7 SELECTED FEATURES

2 Firm's dimension ratios (DIM2 and $DIM2_{2\text{-Years average}}$)

1 Liquidity ratio ($LIQ1_{2\text{-Years average}}$)

4 Profitability ratios ($PROF19_{3\text{-Years average}}$, PROF8, $PROF8_{2\text{-Years average}}$, $PROF10_{3\text{-years average}}$)

(see text for details)

- **DIM, PROF** and **LIQ** ratios are the most explanative attributes
- No **EXP, REA, RES, DEB** or **IND** ratios are selected.
- **Multicollinearly effects** between some pairs of attributes, such as **DIM2 and 2-DIM2** (near 0.9), **PROF8 and 2-PROF8** (near 0.9), **PROF8 and 3-PROF19** (near 0.8), **3-PROF10 and 3-PROF19** (near 0.6)

Empirical Application

Multivariate Methods: Results and Comparison

MDA:

- Statistically significant features: 2-LIQ1 (-), 2-DIM2 (+), 3-PROF10 (+), 3-PROF19 (+)
- Classification matrix hit ratio: 59% (vs. 31% by chance).
- Rating grades´ size impact: A and BBB are better predicted than AA or BB
- Soft classification error measure (+1 or -1 notch): 94%
- .632E and .632+ bootstrapping true hit ratio estimates: 56% (0.52,0.60)

MULTINOMIAL LOGIT:

- Statistically significant features: 3-PROF10 (+), 3-PROF19 (+) for all functions; 2-LIQ1 (-) for functions 2, 3, 4.
- R^2 Cox-Snell=60%; R^2 Nagelkerke=64%; R^2 McFadden=34%
- Classification matrix hit ratio: 60%
- Rating grades´ size impact: A and BBB are better predicted than AA or BB
- Soft classification error measure (+1 or -1 notch): 95%
- .632E and .632+ bootstrapping true hit ratio estimates: 53%-54% (0.45,0.63)

Empirical Application

Multivariate Methods: Results and Comparison

ORDERED LOGIT:

- Statistically significant features: 2-LIQ1 (-)*, 3-PROF10 (+)*, 3-PROF19 (+)*
- R^2 Cox-Snell=54%; R^2 Nagelkerke=58%; R^2 McFadden=29%
- Classification matrix hit ratio: 58%
- Rating grades´ size impact: More equilibrium between categories
- Soft classification error measure (+1 or -1 notch): 91%
- .632E and .632+ bootstrapping true hit ratio estimates: 55% (0.44,0.65)

C4.5 Decision Tree:

- N. decision rules: 11 (AA:0; A:3; BBB:6; BB:1; B:1)
- Maximum depth: 6 levels.
- Significant features: All of them, and specially 2-PROF8, 3-PROF10, 2-LIQ1
- Classification matrix hit ratio: 66%
- Rating grades´ size impact: Both on accuracy and nodes
- Soft classification error measure (+1 or -1 notch): 98%
- .632E and .632+ bootstrapping true hit ratio estimates: 53%-55% (0.50,0.61)

Empirical Application

Multivariate Methods: Results and Comparison

CART – Gini univariate:

- N. decision rules: 12 (AA:0; A:5; BBB:5; BB:1; B:1)
- Maximum depth: 6 levels
- Significant features: 6 features (*except* 2-LIQ1), spec. 2-PROF8, 3-PROF10, DIM2
- Classification matrix hit ratio: 68%
- Rating grades´ size impact: Both on accuracy and nodes
- Soft classification error measure (+1 or -1 notch): 92%
- .632E and .632+ bootstrapping true hit ratio estimates: 56%-58% (0.48,0.68)

CART – Gini oblique:

- N. decision rules: 9 (AA:0; A:2; BBB:4; BB:2; B:1)
- Maximum depth: 5 levels
- Significant features: All of them (7 features).
- Classification matrix hit ratio: 74%
- Rating grades´ size impact: Both on accuracy and nodes
- Soft classification error measure (+1 or -1 notch): 96%
- .632E and .632+ bootstrapping true hit ratio estimates: 59-61% (0.49,0.66)

Overview

Conclusions

- According Basel II, the **Internal Ratings-Based (IRB)** approach for credit risk management and control is based on the estimate of the probability of default (PD) of borrowers.

- For corporate, sovereign and bank exposures, the estimate of PD may be obtained through the **mapping of internal grades to external data**, based on a comparison of internal and external rating determinants.

- Additionally, the **Solvency II** Project uses external ratings for assisting supervisors, conditioned to the comprehensibility of methodologies and ratings' determinants.

- In the literature, ratings' deteminants have been analysed for industrial corporations and banks, but **researches on insurance firms** are much more limited.

- In this paper, we proposed a **four-stage model**, based on statistical techniques and Machine Learning models, to analyse rating determinants, that was applied to a sample of European insurance firms.

Conclusions

- The initial **feature selection process** selected seven financial ratios as relevant attributes, concerning firms' liquidity, profitabilty and size.

- Several **multivariate models** were applied, including MDA, multivariate logit regression, ordinal logit, and decision trees as C4.5, CART-Gini univariate and CART-Gini oblique.

- Previous models were compared through their **comprehensibility and accuracy**, using .632E and .632E+ bootstrap variants.

- The **CART-Gini oblique tree** with nine non-simple decision rules outperformed all other models, and could be a satisfactory solution for the problem of financial strength ratings' determinants.

- ***In the future***, an out-of-time validation test must be performed, together to the full development of the Bayesian hybrid proposal. Also, other Machine Learning techniques should be tested.