Data analytics and decision aiding models for credit risk modeling and banking

Michalis Doumpos
Technical University of Crete
School of Production Engineering and Management
Financial Engineering Laboratory

5th Winter School on Data Analytics

Outline

- Introduction to the context of financial decisions
  - Example context: credit scoring & rating
- Overview of data analytics methods
- Applications in corporate failure prediction (SMEs & banks)
- Conclusions & perspectives

The nature of financial decisions

- Main areas of interest
  - Corporate finance
  - Investments
  - Valuation
  - Risk management
  - Financial engineering
- Fundamental works that shaped the field
  - Markowitz (1950s): portfolio theory
  - Black-Scholes-Merton (1970s): contingent valuation (option pricing)
- Financial decision making has become highly analytical with a strong level of modeling and methodological sophistication

Types of models in finance

- **Normative:** How financial decision makers *should* take *rational* decisions?
- **Descriptive:** What factors drive financial decisions and the financial environment?
- **Predictive:** What is expected to happen?
- **Prescriptive:** How should we act in a specific situations?

Some characteristics

- Dynamic environment with constant changes and deep uncertainties
  - **Implication:** data-driven validation and robustness checks
- Large-scale historical data (real-time data, in some cases)
  - **Implication:** scalable methodologies suitable for large data
- Heavy regulation and tight supervision
  - Basel Committee accords for capital requirements, IFRS accounting/reporting standards, etc.
  - **Implication:** models should comply with regulatory requirements

Credit risk

- Credit risk refers to the probability that borrowers will fail to meet their debt obligations (failure/default)
  - Timely issue (global financial crisis of 2007-08, insolvencies due to the COVID-19 pandemic, …)
- Credit risk is not only relevant for financial institutions
  - Non-financial firms (capital structure decisions, customer relationship management, management of suppliers, etc.)
  - Investors
  - Electronic platforms for online transactions, social lending, crowdfunding, etc.

Credit risk management

- Estimation of the expected loss $\mathbb{E}(L)$ over a specified period

$$\mathbb{E}(L) = PD \times EAD \times LGD$$

  - PD = Probability of default
  - EAD = Exposure at default
  - LGD = Loss given default (% of EAD)
- Each of the three elements (PD, EAD, LGD) is modeled using different approaches
  - International Journal of Forecasting 28(1), 2012
- Credit decision-making considers the trade-off between expected losses in case of default and the revenues from granting "good" loans

Oliver, R.M. (2013), "Financial performance measures in credit scoring", *EURO Journal on Decision Processes* 1(3), 169-185

Credit scoring & rating models

- Credit scoring vs credit rating (Van Gestel & Baesens, 2009; Doumpos et al., 2018)
  - Credit scores are expressed in numeric form and they usually refer to consumers
  - Credit ratings are expressed in symbolic form and they usually refer to corporate and sovereign debt
- Models that evaluate the creditworthiness of a borrower, estimate the probabilities of default and classify the borrowers into risk groups
- Brief history
  - Major rating agencies (Moody’s, S&P, Fitch) established in early 1900s
  - Consumer credit scores introduced in the 1950s by Fair Isaac
  - US Fair Credit Reporting Act (1970)
  - Rapid expansion mainly after the 1980s (FICO score, 1989)

Examples of credit ratings

Moody’s	S&P, Fitch	Credit worthiness
Aaa	AAA	An obligor has **extremely strong** capacity to meet its financial commitments.
Aa1, Aa2, Aa3	AA+, AA, AA–	An obligor has **very strong** capacity to meet its financial commitments. It differs from the highest-rated obligors only to a small degree.
A1, A2, A3	A+, A, A–	An obligor has **strong** capacity to meet its financial commitments but is somewhat more susceptible to the adverse effects of changes in circumstances and economic conditions than obligors in higher-rated categories.
Baa1, Baa2, Baa3	BBB+, BBB, BBB–	An obligor has **adequate** capacity to meet its financial commitments. However, adverse economic conditions or changing circumstances are more likely to lead to a weakened capacity of the obligor to meet its financial commitments.
Ba1, Ba2, Ba3	BB+, BB, BB–	An obligor is **less vulnerable** in the near term than other lower-rated obligors. However, it faces major ongoing uncertainties and exposure to adverse business, financial, or economic conditions which could lead to the obligor's inadequate capacity to meet its financial commitments.
B1, B2, B3	B+, B, B–	An obligor is **more vulnerable** than the obligors rated 'BB', but the obligor currently has the capacity to meet its financial commitments. Adverse business, financial, or economic conditions will likely impair the obligor's capacity or willingness to meet its financial commitments.
Caa	CCC	An obligor is **currently vulnerable**, and is dependent upon favorable business, financial, and economic conditions to meet its financial commitments.
Ca	CC, C	An obligor is **currently highly vulnerable**.
C	D	An obligor has **failed** to pay one or more of its financial obligations when it became due.

Three main types of approaches

- Judgmental models
  - Applicable when there is a lack of historical data
  - Expert judgments and experience of the credit analysts
  - Elaborate structure providing rich information on all aspects of assessment process
- Empirical analytical models
  - Linear or non-linear models
  - Reliance of historical databases (internal & external)
  - Quantitative and qualitative data
- Financial models
  - Based on theories like option pricing and market data (equities, bonds, CDSs)
- Each approach has its pros and cons
  - Empirical and financial models dominate the industry, unless data are lacking

Methods for analytical modeling

- Statistical techniques
  - Discriminant analysis, logistic regression, hazard models
- Data mining / machine learning
  - Neural networks, kernel methods, decision trees, ensembles
- Some reviews and comparative studies
  - Abdou & Pointon (2011)
  - Lessmann et al. (2015)
  - Papageorgiou et al. (2008)

Statistical approaches

- Discriminant analysis (linear & quadratic)

$$f(\mathbf{x}) = a + \mathbf{b}^\top \mathbf{x} + \mathbf{x}^\top \mathbf{D}\mathbf{x}$$

  - Estimation of the parameters based on the assumption of multivariate normality

- Logistic regression (LR)

$$\Pr(1|\mathbf{x}) = \frac{1}{1 + \exp[-(a + \mathbf{b}^\top \mathbf{x})]} \Rightarrow \ln \frac{\Pr(1|\mathbf{x})}{1 - \Pr(1|\mathbf{x})} = a + \mathbf{b}^\top \mathbf{x}$$

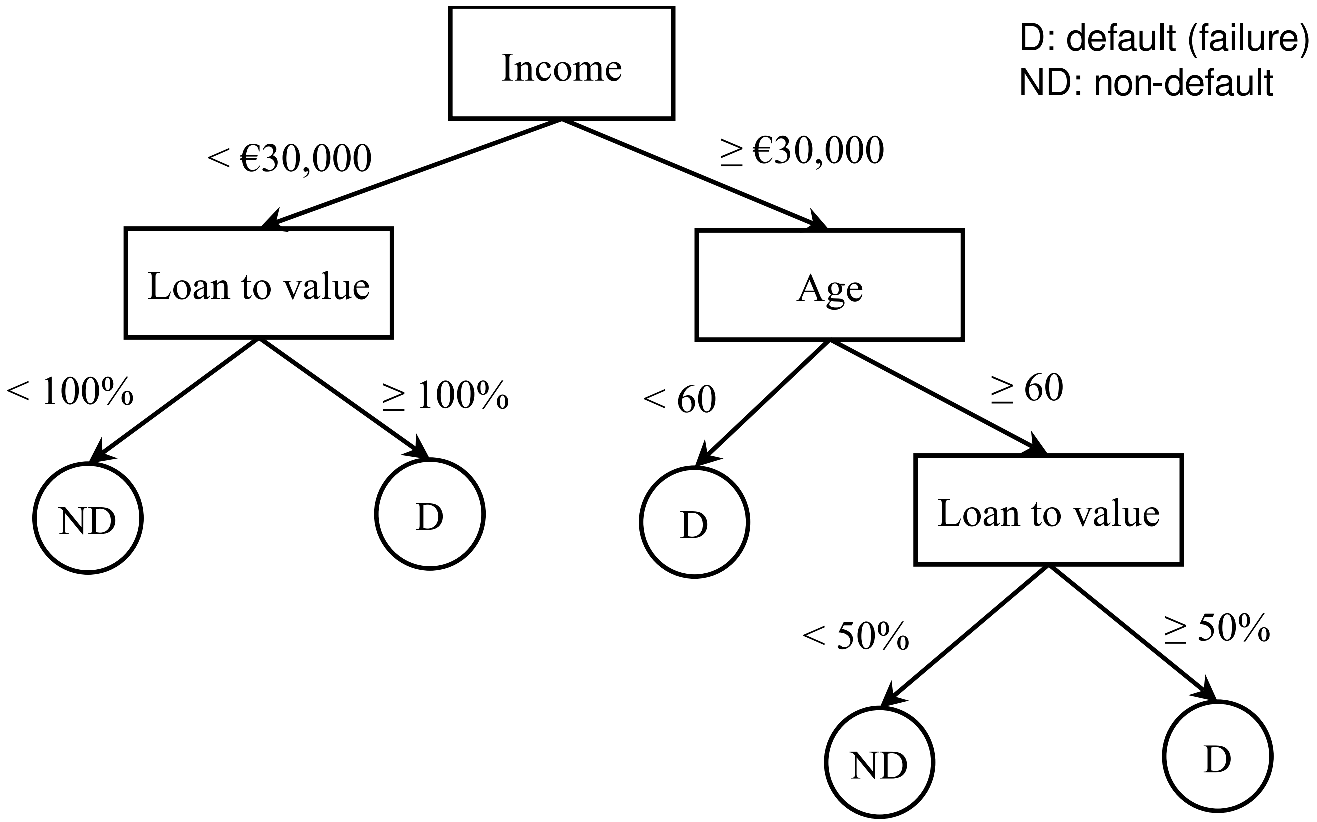
  - Estimation by maximum likelihood
  - Allows for statistical inferences
  - Industry standard

Machine learning methodologies

- Standalone approaches
  - Classification trees & rule-based models
  - Neural networks
  - Kernel methods (e.g., support vector machines)
- Ensembles
  - Combination of multiple classifiers developed with a single method or multiple methods
- Hybrid systems
  - Combination of different methodologies for feature selection and model construction

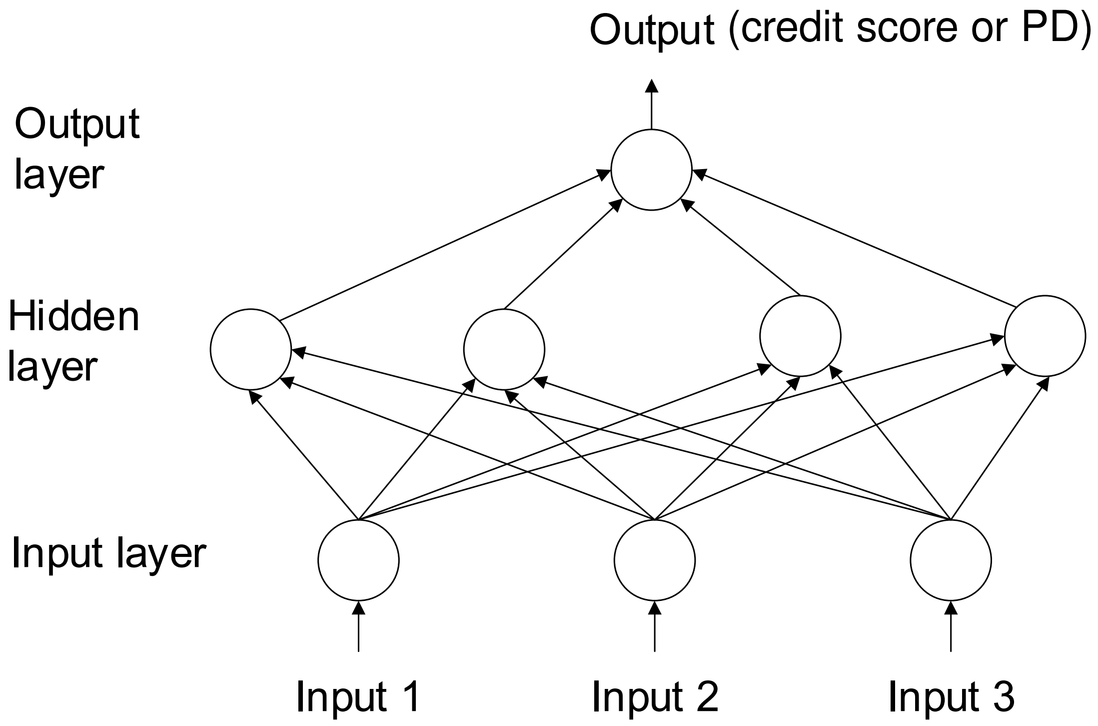
Hastie, T., Tibshirani, R., and Friedman, J. (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.)*, Springer, New York.

Classification trees



- Fast model construction, easy to understand (if the tree is small)
- Not ideal for scoring and PD estimation

Neural networks



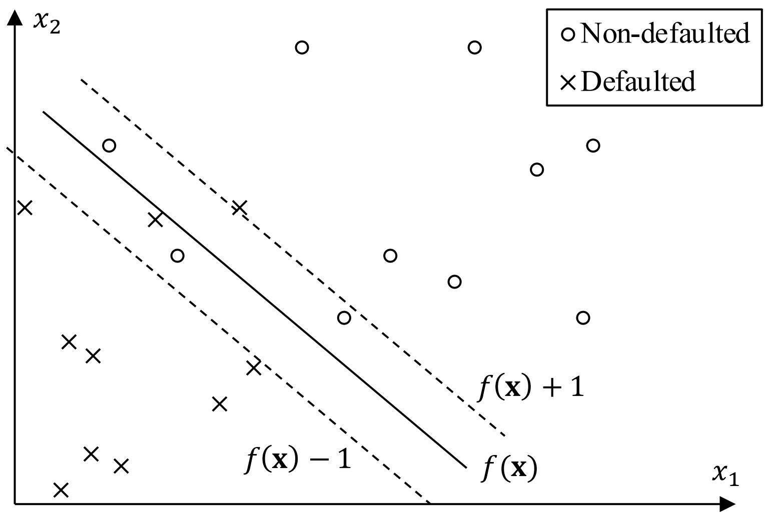
- General model, providing a lot of flexibility
- The network's architecture and fitting could be issues
- State-of-the-art deep learning systems allow the modeling of highly complex data described by many features (Sirignano et al., 2018)

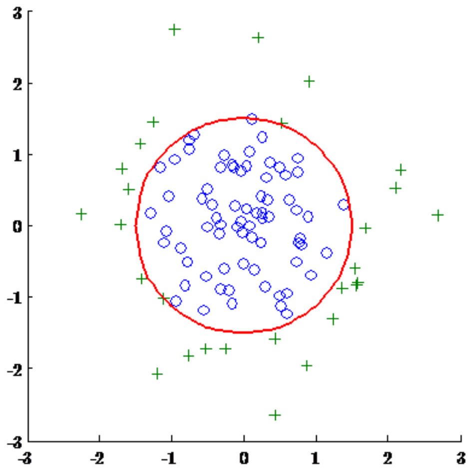
Support vector machines

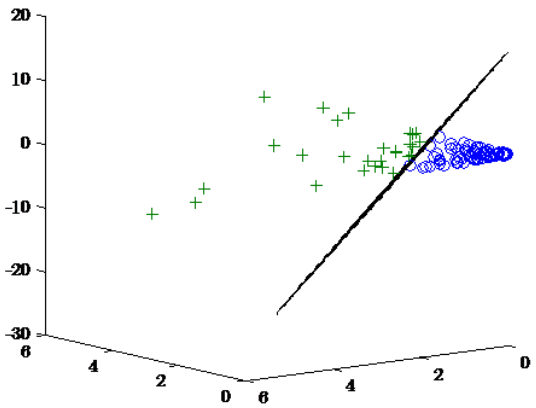
- Margin maximization principle
- Non-linear separation by mapping the data to a higher dimensional space though a kernel function

$$f(\mathbf{x}) = a + \sum_{i=1}^{m} b_i y_i K(\mathbf{x}, \mathbf{x}_i)$$

- Quadratic optimization based on the regularization principle (penalize complexity)
- Non-linear model fitting is computationally intensive for large data sets







Ensembles

- Bagging (bootstrap aggregation; Breiman, 1996)
  - Combine multiple base models derived by the application of a specific method to different bootstrap samples of a data set
  - Random forest (Breiman, 2001): impose independence by sampling both observations and features
- Boosting (Schapire, 1990)
  - Combine multiple base models developed iteratively, gradually focusing on more difficult instances
- Stacking (Wolpert, 1992)
  - Combine models developed with different methods

Multicriteria value function models

- Evaluation of a borrower's creditworthiness through an additive value function

$$V(\mathbf{x}) = w_1 v_1(x_1) + w_2 v_2(x_2) + \cdots + w_n v_n(x_n)$$

  - $w_1, \ldots, w_n$: non-negative weights of the attributes that sum up to 1
  - $v_1, \ldots, v_n$: monotone marginal value functions (usually defined in [0, 1])
- Many rating models have this form (Krahnen & Weber, 2001)
- Advantages / disadvantages
  - Comprehensibility, easy to use and construct
  - The additivity assumption may be strong in some cases
- Model fitting with linear/quadratic programming (Doumpos & Zopounidis, 2002, 2014)

Comprehensibility vs predictive power

- Many recent data-driven approaches rely on complex machine learning approaches
  - Often, they lack transparency
  - Reliance on large sets predictor attributes
- The trade-off between high predictive performance and interpretability is relevant (Hand, 2006; Obermann & Waack, 2015)
  - Interpretability provides insights into the operation of the models (do they make economic/business sense?), adds transparency, and builds user's confidence on the model's outputs
- Some comparative studies have shown that simpler models perform well
  - Jones, Johnstone & Wilson (2015, JBF)
  - Obermann & Waack (2015, ESWA)

Application - Prediction of SMEs' failures

- Test the performance of state-of-the-art data analytic methods for predicting the failure of small and medium-sized enterprises (SMEs) in Europe
- Empirical analysis on a large sample of European SMEs
  - Cross-country setting, not widely studied in the literature

Sample

- European commercial SMEs
  - Employees<250 & (turnover≤€50M or assets≤€43M)
  - Micro enterprises excluded
    - Employees<10 & (turnover≤€2M or assets≤€2M)
- Data sources:
  - Amadeus database (financial data and failure status)
  - IMD World Competitiveness Yearbook (macroeconomic data)
- Time period
  - Failures from 2012q1 to 2017q1
  - Data lagged by one year prior to observation year (2011-2015)
- Country coverage: Countries with at least 100 bankruptcies
  - Finland, France, Germany, Italy, Portugal, Spain
- Goal: develop models to predicting the failure of SMEs

Sample composition

Country	Obs.	Failure rate
Italy	234,949	2.13%
Spain	107,373	0.64%
France	53,670	1.59%
Portugal	48,473	1.25%
Germany	24,307	0.64%
Finland	11,757	1.69%
Total	480,529	1.56%

Training observations weighted by country

Models

- Logistic regression (LR)
- Regularized logistic regression (RegLR)

$$\min_{\mathbf{b}} \|\mathbf{b}\|_1 + C \sum_i \ln\left(1 + e^{-y_i \mathbf{b}^\top \mathbf{x}_i}\right)$$

- Generalized additive model (Hastie & Tibshirani, 1990)

$$\ln \frac{\Pr(1|\mathbf{x})}{1 - \Pr(1|\mathbf{x})} = a + f_1(x_1) + f_2(x_2) + \cdots + f_n(x_n)$$

  - $f_1, \ldots, f_n$ smooth penalized regression splines (MGCV R package)
- Random Forest (RF, Breiman, 2001)
- Gradient Boosting Machine (GBM, Friedman, 2001)
- Extreme Gradient Boosting (XGB, Chen & Guestrin, 2016)

Model fitting and testing setup

- Training on 2011-13, testing on 2014-15
- Model specifications
  - Specification 1: Country-specific models based only on financial ratios
  - Specification 2: Pooled data set consisting of all countries
    - Models based solely on financial ratios
    - Combination of financial ratios with macroeconomic indicators
- Performance metric
  - Area under the receiver operating characteristic curve (AUROC)
    - AUROC represents the probability that a random non-failed firm will be assigned a lower PD score (lower likelihood of failure) than a failed firm
  - Accuracy rates can be problematic
    - Misleading due to major imbalance between the classes
    - They do not consider the distribution of PD scores

Attributes

Financial ratios
Profitability: Profit before tax / Assets
Solvency: Equity / Assets
Debt turnover: Debt / Turnover
Liquidity: Cash / Current liabilities
Management: Cost of employees / Turnover

Macroeconomic environment
Investment risk
Unemployment rate
Corporate tax rate on profit
Impact of taxation on entrepreneurial activity
Investment incentives
Competition legislation
Ease of doing business
Labor regulations
Efficiency of SMEs
Banking and financial services
Access to credit
Bureaucracy

AUROC - Models based on financial ratios

Country-specific models

	FIN	FRA	DEU	ITA	PRT	ESP	*Mean*
LR	**63.8**	77.7	74.6	86.2	82.7	85.1	*78.4* (2)
RegLR	57.9	68.5	75.0	85.9	78.1	85.5	*75.1* (4)
GAM	61.6	**78.1**	**78.0**	87.9	83.9	86.2	***79.3* (1)**
RF	57.9	69.3	70.7	84.2	77.9	72.4	*72.1* (5)
GBM	60.1	76.7	62.1	**88.1**	83.9	85.5	*76.1* (3)
XGB	62.1	78.1	71.8	87.7	**84.5**	**86.4**	*78.4* (2)
Mean	*60.6*	*74.7*	*72.0*	*86.7*	*81.8*	*83.5*	*76.6*

Pooled data

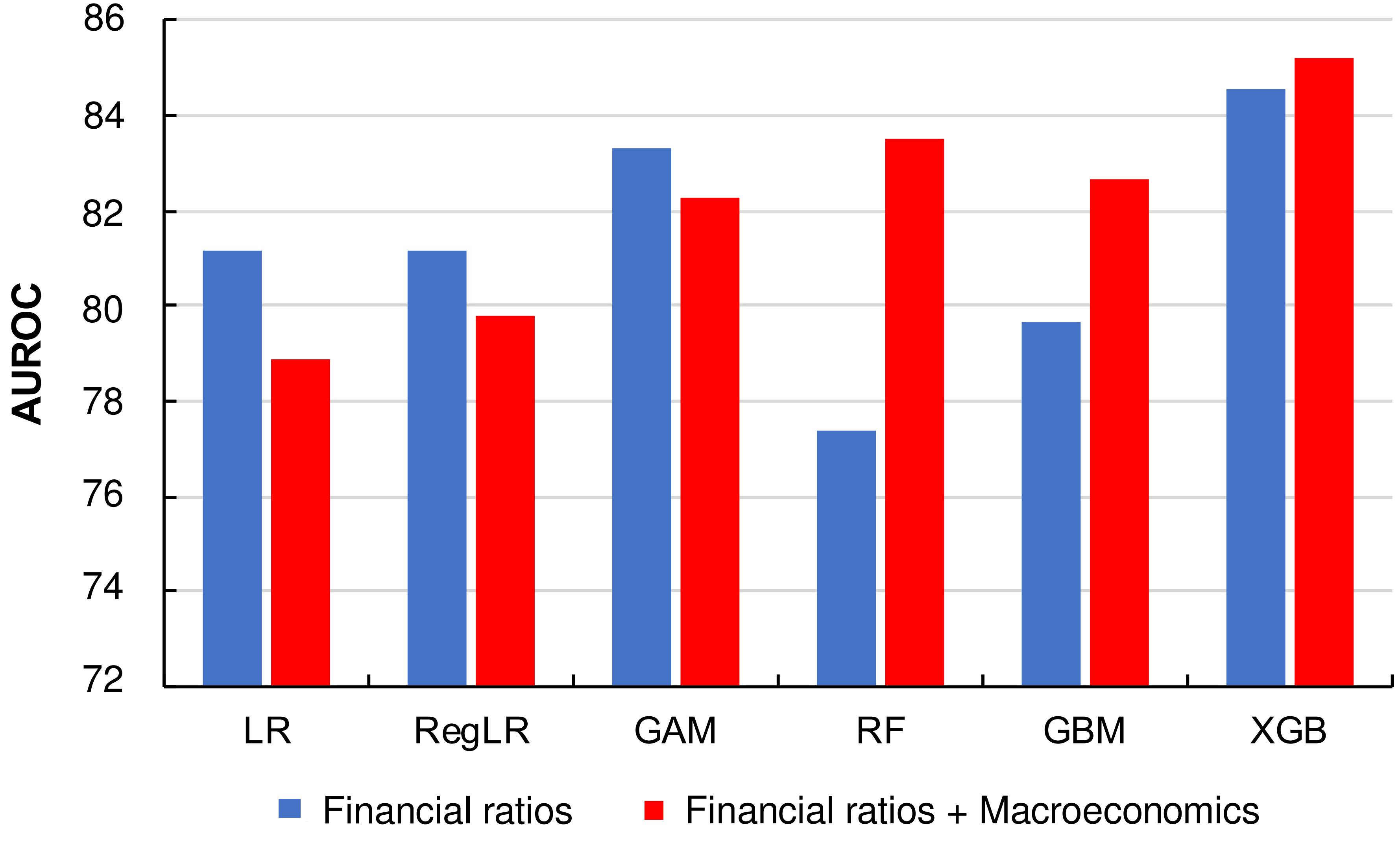
	FIN	FRA	DEU	ITA	PRT	ESP	*Mean*	*Full sample*
LR	62.0	76.0	74.8	85.8	79.2	85.3	*77.2*	*81.2* (3)
RegLR	62.0	76.0	74.8	85.8	79.2	85.3	*77.2*	*81.2* (3)
GAM	**65.1**	78.1	76.5	87.5	82.4	87.1	*79.5*	*83.3* (2)
RF	59.8	67.5	70.2	82.2	79.0	78.9	*72.9*	*77.4* (5)
GBM	62.5	75.2	73.3	83.2	78.5	83.1	*75.9*	*79.7* (4)
XGB	64.6	**78.9**	**78.6**	**88.7**	**85.6**	**87.9**	***80.7***	***84.5* (1)**
Mean	*62.7*	*75.3*	*74.7*	*85.5*	*80.7*	*84.6*	*77.2*	

AUROC – Full models

	FIN	FRA	DEU	ITA	PRT	ESP	*Full sample*
LR	62.1	76.2	74.6	86.5	77.7	85.9	*78.9* (5)
RegLR	61.6	76.7	73.5	86.9	80.4	86.4	*79.8* (6)
GAM	66.2	78.0	76.0	88.3	81.0	**87.8**	*82.2* (4)
RF	64.7	76.0	75.9	86.8	84.3	85.4	*83.5* (2)
GBM	**66.2**	77.7	78.1	88.0	82.8	87.1	*82.6* (3)
XGB	63.4	**78.6**	**79.8**	**88.4**	**84.8**	87.2	***85.2* (1)**

With the enrichment of the set of attributes (financial ratios + macro-economic indicators), the machine learning models improve significantly

Comparison of models



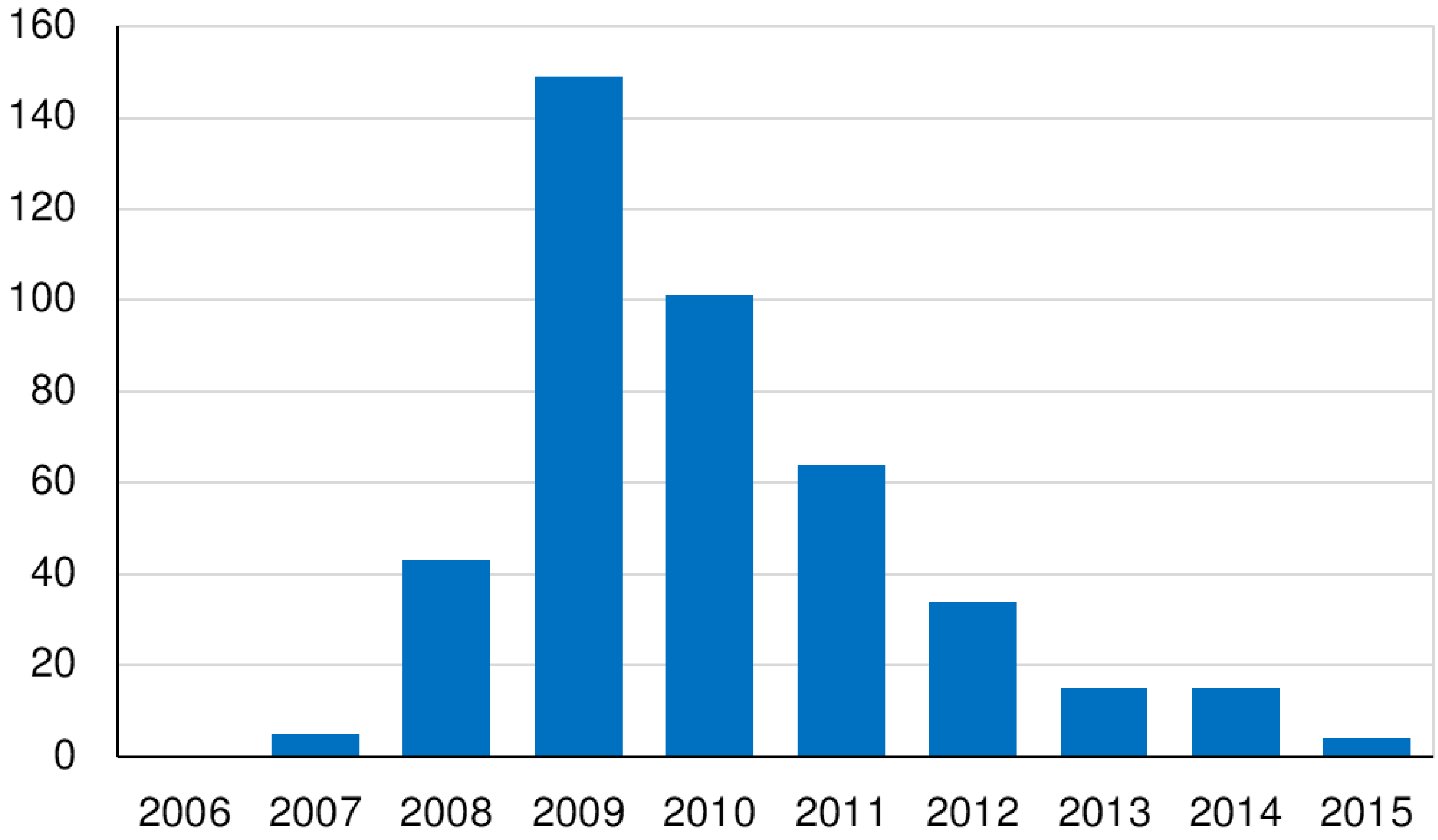
Application on bank failure prediction

- Bank defaults have some unique aspects (e.g., very rare, but severe)
- Historically, bank default rates have been very low, but picked up a lot during the 2007-2008 global crisis
- Goal: develop bankruptcy prediction models in a multi-class context considering the time to failure

Data and setting of the analysis

- Annual data for US banks derived from
  - Federal Financial Institutions Examination Council (FFIEC) call reports
  - 430 failed banks from the Federal Deposit Insurance Corporation (FDIC)
- In each year $t = 2005, \dots, 2014$, the banks are classified as failed or non-failed, depending on their status in $t + 1$
  - Failed banks are recorded up to 3 years prior to failure, i.e., for a bank that failed in year t we collect data for $t - 1, t - 2, t - 3$
- 58140 non-failed & 1207 failed bank-year observations

Number of failures by year



Bootstrap testing (100 runs)

Sampling with replacement from the unique bank IDs for avoid training-testing overall

Attributes - Financial ratios

C	Equity / Assets (EQTA)
A	Loan loss allowances / Total loans (LLATL)
M	Efficiency ratio (expenses/revenues, EFF)
E	Return on assets (ROA)
	Net interest margin (NIM)
L	Loans / Deposits (LTD)
S	Off-balance sheet risk-weighted asses / Total risk-weighted asses (OBS)
Size	Ln(Total assets) (LNTA)

CAMELS

C: capital, A: asset quality, M: management capacity, E: earnings power,
L: liquidity, S: sensitivity to market risks

Attributes - Diversification indices

Items	Components
Income	Interest vs non-interest income (DIV_{income})
Expenses	Interest vs non-interest expenses ($DIV_{expenses}$)
Interest income	Loans & leases, investments, interest bearing bank balances, federal funds sold and security resales, trading account assets ($DIV_{int.\ income}$)
Earning assets	Net loans vs total investments (DIV_{assets})
Loan portfolio	Real estate loans, agricultural loans, officer shareholders loans, individual loans, commercial loans, other loans (DIV_{loans})
Deposits	Individuals & corporations, US Government, US banks and depository institutions, Foreign banks, Foreign governments ($DIV_{deposits}$)

Indicators' AUROC

Indicators	Full sample	$t-1$	$t-2$	$t-3$	$\Delta(t-3, t-1)$
EQTA	75.6	90.1	72.4	63.0	–27.1
LLATL	68.7	87.0	66.1	51.5	–35.5
EFF	66.8	81.4	65.0	52.7	–28.7
ROA	77.0	93.3	76.5	59.7	–33.6
NIM	63.0	74.0	62.5	51.4	–22.6
LTD	66.7	57.3	70.6	73.0	15.7
OBS	52.8	47.0	54.2	57.6	10.6
LNTA	59.5	58.7	60.1	59.7	1.0
$DIV_{int. income}$	64.5	57.6	65.2	71.4	13.8
DIV_{income}	71.0	70.8	71.4	70.7	–0.1
$DIV_{expenses}$	69.1	64.0	70.8	72.8	8.8
DIV_{assets}	71.8	68.2	73.6	73.9	5.7
DIV_{loans}	71.5	73.5	71.3	69.5	–4.0
$DIV_{deposits}$	52.7	50.5	52.7	55.0	4.5

Strong predictors close to failure, weak long-term power

Robust predictive power (short & long-term)

Binary and multi-class classifiers

- Credit risk and bankruptcy prediction models are usually considered in a binary classification setting
- Different modeling specifications tested in the analysis
  - BIN-T1: binary classifiers fitted only on the data for $t-1$
  - BIN-ALL: binary classifiers fitted on all data
  - MULTICLASS: four ordinal classes

    $\mathcal{C}_1$: non-failed banks (low risk)

    $\mathcal{C}_2$: failed banks 3 years prior to failure (medium risk)

    $\mathcal{C}_3$: failed banks 2 years prior to failure (high risk)

    $\mathcal{C}_4$: failed banks 1 years prior to failure (very high risk)

Error-correcting output codes (ECOC)

- Ensemble scheme for creating multi-class models by combining binary classifiers (Dietterich & Bakiri, 1995)

Schemes	Binary partitions of the classes
ECOC1	$\{C_1, C_2\}$, $\{C_2, C_3\}$, $\{C_3, C_4\}$
ECOC2	$\{C_1, C_2\}$, $\{C_1, C_3\}$, $\{C_1, C_4\}$, $\{C_2, C_3\}$, $\{C_2, C_4\}$, $\{C_3, C_4\}$
ECOC3	$\{C_1, \{C_2, C_3, C_4\}\}, \{C_2, \{C_3, C_4\}\}, \{C_3, C_4\}$
ECOC4	$\{C_1, \{C_2, C_3, C_4\}\}, \{\{C_1, C_2\}, \{C_3, C_4\}\}, \{\{C_1, C_2, C_3\}, C_4\}$
ECOC5	$\{C_1, C_2\}$, $\{C_1, C_3\}$, $\{C_1, C_4\}$, $\{C_2, C_3\}$, $\{C_2, C_4\}$, $\{C_3, C_4\}$, $\{C_1, \{C_2, C_3\}\}$, $\{C_1, \{C_2, C_4\}\}$, $\{C_1, \{C_3, C_4\}\}$, $\{C_2, \{C_3, C_4\}\}$, $\{\{C_1, C_2\}, C_3\}$, $\{\{C_1, C_2\}, C_4\}$, $\{\{C_1, C_3\}, C_4\}$, $\{\{C_2, C_3\}, C_4\}, \{\{C_1, C_2\}, \{C_3, C_4\}\}$, $\{C_1, \{C_2, C_3, C_4\}\}, \{\{C_1, C_2, C_3\}, C_4\}$

ECOC model combination

- Combination of binary classifiers $g_1(\mathbf{x}), g_2(\mathbf{x})$, ..., into a stacked meta-model fitted on the multi-class labels

$$f(\mathbf{x}) = w_1 g_1(\mathbf{x}) + w_2 g_2(\mathbf{x}) + \cdots$$

  - Combination with ordinal LR
  - Optimization-based combination

$$\begin{aligned}
\min \quad & \lambda \left(\sum_{\ell=1}^{3} \frac{1}{m_\ell} \sum_{i \in C_\ell} \sigma_i^2 + \sum_{\ell=2}^{4} \frac{1}{m_\ell} \sum_{i \in C_\ell} \varepsilon_i^2 \right) + \frac{1}{2} \|\mathbf{w}\|^2 \\
\mathrm{s.t.:} \quad & \sum_k w_k g_k(\mathbf{x}_i) + \sigma_i \geq t_\ell + 1 \qquad \forall i \in C_\ell, \ell = 1, 2, 3 \\
& \sum_k w_k g_k(\mathbf{x}_i) - \varepsilon_i \leq t_{\ell-1} - 1 \qquad \forall i \in C_\ell, \ell = 2, 3, 4 \\
& w_k, t_\ell, \sigma_i, \varepsilon_i \geq 0
\end{aligned}$$

Classifiers

- LR and ordinal LR
- Support vector machines (SVMs) with exponential kernel
- RUSBOOST (Seiffert et al., 2010)
- Multicriteria additive value model (MCDA)

Performance measure

- Area Under Receiver Operating Characteristic curve (AUROC)
- Extensions to ordinal classification (Waegeman et al., 2008)
  - Average of AUROCs for all pairs of classes $\{\mathcal{C}_k, \mathcal{C}_\ell\}$, $\forall \ell > k$

$$AUROC_{\text{ovo}} = \frac{2}{q(q-1)} \sum_{k} \sum_{\ell > k} AUROC_{k\ell}$$

  - Average of AUROCs for $q-1$ binary partitions between a set of classes $\{\mathcal{C}_1, \dots, \mathcal{C}_k\}$ versus the set of classes $\{\mathcal{C}_{k+1}, \dots, \mathcal{C}_q\}$

$$AUROC_{\text{cons}} = \frac{1}{q-1} \sum_{k=1}^{q-1} AUROC_k$$

AUROC results for standalone models

	BIN-ALL				BIN-T1				MULTICLASS	
	MCDA	LR	RUSBOOST	SVM	MCDA	LR	RUSBOOST	SVM	MCDA	OLR
Cons	94.59	93.50	94.60	**94.79**	94.02	93.13	93.40	94.20	94.66	93.47
Ovo	83.23	81.98	83.36	81.05	84.21	83.30	83.45	83.54	**84.62**	82.05
Full binary	92.61	91.45	92.44	**93.24**	90.89	89.66	90.19	91.33	92.17	91.19
$t-1$	96.89	95.97	97.04	96.74	97.34	97.01	96.80	97.34	**97.36**	96.17
$t-2$	92.32	90.87	92.21	**92.93**	90.80	88.83	90.07	91.08	92.07	90.55
$t-3$	88.22	87.10	87.64	**89.74**	83.93	82.48	83.08	85.02	86.60	86.41

AUROC of ECOC schemes (averaged over all classification methods)

	Stacking with MCDA						Stacking with OLR					
	ECOC1	ECOC2	ECOC3	ECOC4	ECOC5	Mean	ECOC1	ECOC2	ECOC3	ECOC4	ECOC5	Mean
Cons	93.61	94.63	**94.66**	94.41	94.62	94.39	94.21	94.53	94.53	94.61	94.65	94.50
Ovo	83.26	84.32	83.91	84.09	**84.33**	83.98	82.24	82.97	82.93	82.82	82.85	82.76
Full binary	90.63	92.16	92.37	91.79	92.13	91.81	92.13	92.49	92.52	92.63	**92.71**	92.49
$t-1$	96.87	97.37	97.25	97.28	**97.37**	97.23	96.63	96.92	96.89	96.92	96.91	96.85
$t-2$	90.27	91.89	92.09	91.54	91.88	91.53	91.82	92.12	92.16	92.33	**92.41**	92.17
$t-3$	84.18	86.73	87.30	86.03	86.67	86.18	87.52	88.02	88.10	88.23	**88.42**	88.06

Classifiers performance (averaged over all ECOC schemes)

Stacking approach	MCDA				OLR			
Individual classifiers	MCDA	LR	RUSBOOST	SVM	MCDA	LR	RUSBOOST	SVM
Cons	94.73	94.04	93.72	**95.06**	94.59	93.97	94.48	94.99*
Ovo	**84.62**	83.72	83.48	84.10	83.40	82.90	82.19	82.55
Full binary	92.26	91.46	90.68	92.85	92.54	91.78	92.43	**93.21**
$t-1$	97.41	96.98	96.97	**97.54**	96.92	96.56	96.83	97.10
$t-2$	92.14	90.86	90.52	92.61	92.30	91.18	92.25	**92.94**
$t-3$	86.74	86.07	83.96	87.97	88.01	87.17	87.79	**89.24**

ECOC5 vs standalone binary/multi-class models

	ECOC5 versus BIN-ALL								ECOC5 versus MULTICLASS			
Stacking model	MCDA				OLR				MCDA		OLR	
Indiv. models	MCDA	LR	RUSBOOST	SVM	MCDA	LR	RUSBOOST	SVM	MCDA	LR	MCDA	LR
Cons	+	+	−	+	●	+	+	+	+	+	●	+
Ovo	+	+	+	+	+	+	−	+	+	+	−	+
Full binary	−	●	−	−	●	+	+	+	+	+	+	+
$t-1$	+	+	+	+	+	+	−	+	+	+	−	+
$t-2$	−	●	−	−	●	+	+	+	+	+	+	+
$t-3$	−	−	−	−	−	+	+	−	+	−	+	+

Conclusions & perspectives

- Financial decision problems have a strong analytical aspects and they have become data-intensive
- Broad range of problems for data analytics
  - Supervised learning (classification & regression)
  - Unsupervised learning (clustering)
- Perspectives
  - Regulatory compliance (where relevant)
  - Comprehensibility versus performance
  - Incorporation of expert (domain) knowledge
  - Financial materiality versus statistical performance

References

Abdou, H. A. and Pointon, J. (2011), “Credit scoring, statistical techniques and evaluation criteria: A review of the literature”, *Intelligent Systems in Accounting, Finance and Management*, 18(2–3), 59–88.

Breiman, L. (1996), “Bagging predictors”, *Machine Learning*, 24, 123-140.

Breiman, L. (2001), “Random forest”, *Machine Learning*, 45, 5-32.

Chen, T. and Guestrin, C. (2016), “XGBoost”, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’16*, pp. 785–794.

Doumpos, M. and Zopounidis, C. (2002), *Multicriteria Decision Aid Classification Methods*, Kluwer Academic Publishers, Dordrecht.

Doumpos, M. and Zopounidis, C. (2014), *Multicriteria Analysis in Finance*. Springer, Berlin-Heidelberg.

Doumpos, M., Lemonakis, C., Niklis, D., and Zopounidis, C. (2018), *Analytical Techniques in the Assessment of Credit Risk*, Springer, Berlin-Heidelberg.

Friedman, J. H. (2001), “Greedy function approximation: A gradient boosting machine”, *Annals of Statistics*, 29(5), 1189–1232.

Jones, S., Johnstone, D., and Wilson, R. (2015), “An empirical evaluation of the performance of binary classifiers in the prediction of credit ratings changes”, *Journal of Banking & Finance*, 56, 72–85.

Hand, D.J. (2006), “Classifier technology and the illusion of progress”, *Statistical Science*, 21(1), 1-14.

Hastie, T. J. and Tibshirani, R. J. (1990), *Generalized Additive Models*, Chapman & Hall/CRC, New York/Boca Raton.

Krahnen, J.P. and Weber, M. (2001), “Generally accepted rating principles: A primer”, *Journal of Banking and Finance*, 25, 3–23.

Lessmann, S., Baesens, B., Seow, H.-V., and Thomas, L. C. (2015), “Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research”, *European Journal of Operational Research*, 247(1), 124–136.

Obermann, L. and Waack, S. (2015), “Demonstrating non-inferiority of easy interpretable methods for insolvency prediction”, *Expert Systems with Applications*, 42(23), 9117-9128.

Papageorgiou, D., Doumpos, M., Zopounidis, C., and Pardalos, P. M. (2008), “Credit rating systems: Regulatory framework and comparative evaluation of existing methods”, in: C. Zopounidis, M. Doumpos, & P. M. Pardalos (Eds.), *Handbook of Financial Engineering* (pp. 457–488), Springer, New York.

Schapire, R. (1990), “The strength of weak learnability”, *Machine Learning*, 5, 197-227.

Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J., and Napolitano, A. (2010), “RUSBoost: A hybrid approach to alleviating class imbalance”, *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 40(1), 185–197.

Sirignano, J., Sadhwani, A. and Giesecke, K. (2018), “Deep learning for mortgage risk” (https://arxiv.org/abs/1607.02470).

Waegeman, W., De Baets, B., and Boullart, L. (2008), “ROC analysis in ordinal regression learning”, *Pattern Recognition Letters*, 29(1), 1–9.

Wolpert, D.H. (1992), “Stacked Generalization”, *Neural Networks*, 5, 241–259.

Van Gestel, T., and Baesens, B. (2009), *Credit Risk Management - Basic Concepts*, Oxford University Press, Oxford.