Bankruptcy scoring using the SAS Enterprise Miner

Michael Wetzel
Systematika Informationssysteme AG
Switzerland

seugi 2002 paris
11-13 June 2002

systematika

Systematika

Systematika- Specialist for business intelligence in the finance industry: credit risk, CRM, scoring, optimization etc.

1983 Systematika founded in Heidelberg/Germany

1995 Systematika founded in Zürich/Switzerland
with offices in Zürichand Geneva

1997 Foundation of the Subsidiary
Math Consulting Group AG, specialized in
Optimization in Asset and Liability Management

Partnerships: SAS Institute, etc.

Empirial facts

In the US, 300 companies give up every week

Worldwide, 100 bond issuing companies went bankrupt on $36bn of long-term, publicly held corporate debt in 1999.

Credit loss increases dramatically for lower rating categories.

Company defaults and economic prosperity are significantly correlated

The overall economic loss is substatial

The banks‘ data situation on defaults is not satisfactory
information comes too late, limited view, lack of consistency

Prediction has been attempted since 30 years and is still a challenge.

History of default prediction

Until 1968 no systematic approach, rather subjective assessment, company valuation

1968 Altman‘s first model: discriminant analysis of 5 financial ratios

70s, 80s: regression based models, criticised for their assumptions

80s, 90s: neural networks, criticised for their shortcomings and maximum-likelihood logit or probit models

Consensus that traditional tools are too expensive and subjective.

>90s: a lot of empirical research but no clear picture of what data and what methodology is best

Today, models take the complexity into account: multiple approaches
- to data: different sets
- to methods: linear and nonlinear

The Basel II requirements by the BIS make the development of better prediction systems business-critical.

Some difficulties in default prediction

Operational definition of „default“

Lack of historical default data

A lot of qualitative knowledge, but few quantitative analyses

Inconsistency of data, every bank has it‘s own view and data

Formalization of qualitative information

Models often do not give a clear result.

Technical, modelling problems:

statistical assumptions (normality) are violated, e.g. risk factors have log-normal distribution

Many scoring systems can behave strangely

Adequacy of the models is not always guaranteed

SYSTEMATIKA

Variables for a prediction model

Static/dynamic information: Single point in time/trend data

Account data: TRX, volumes, balances, and derived indicators

Management quality: planfulness; company quality: products, etc

Business environment: economy, region, sector, competitors

Benchmark deviations: comparison with competitors (means, trends)

Assessments: ratings, scores and their shifts (rating migration)

Financial well-being: credit exposure, assets, guarantees

Financial ratios: balance sheet analysis, company valuation

Risk measures: indicators of the company‘s operations/performance

Financial data: debt capacity, return on equity, risk measures

Behavioural indicators: corporate governance

Soft factors: reputation, classifications (e.g.„underperformer“)

Methods, models

Financial statement information; financial ratios

A long tradition of accounting and corporate finance models; company valuation
Problem: latency, limitations, randomness of ratios, uncertainty which predictors are meaningful; simplistic approach

Approaches based on Option pricing models

Only usable for publicly listed companies

Rating

Subjective, structured assessment of a company's health
Problem: lack of objectivity, lengthy and costly process

Methods, models

Scoring models

Statistical models using stepwise discriminant analysis and logistic regression
e.g. CreditMetrics; comparable to credit scoring
Problem: predictors, assumptions
Neural network approach: non-linear relationships
e.g. S&P CreditModel
Problem: model acceptance
Advantage: reliability can be evaluated statistically; faster and less costly than traditional methods

Other approaches

Simulation (RiskPro)
OLAP (RiskPro)
Visualisation (Skyon)
Mining: Decision tree (SAS Enterprise Miner)

Empirical study

Customer: one of the world‘s leading financial institutions

Data: small & medium enterprises (SME): N=100.000, non-listed

„Critical“ customers: N=5.000, Real defaults: 500

Data history: 3-5 years

Predictors

Phase 1 bank internal
account balance and turnover
ratings, scores, segment, typology
financial behaviour, excesses
static information, age
company: size, turnover, sector
Phase 2 external data
competitive status
financial ratios, accounting information
market data

Criterion

„Default“="corporate distress"
factor score named
"Customer in critical condition"
averse rating migration
writeoffs, provisions
legal measures by the bank
nonperforming loans
bankruptcy

Data mining process

1. Analysis of predictors
- predictive power assessed individually: (non)parametric tests
- correlation and factor analyses - exploration of the dimensionality

2. Analysis of the criterion and the value function
- "critical situation"
- loss incurred by the bank

Factor analyses, canonical correlation with predictors

3. Decision decision tree / CHAID analysis "just to see"

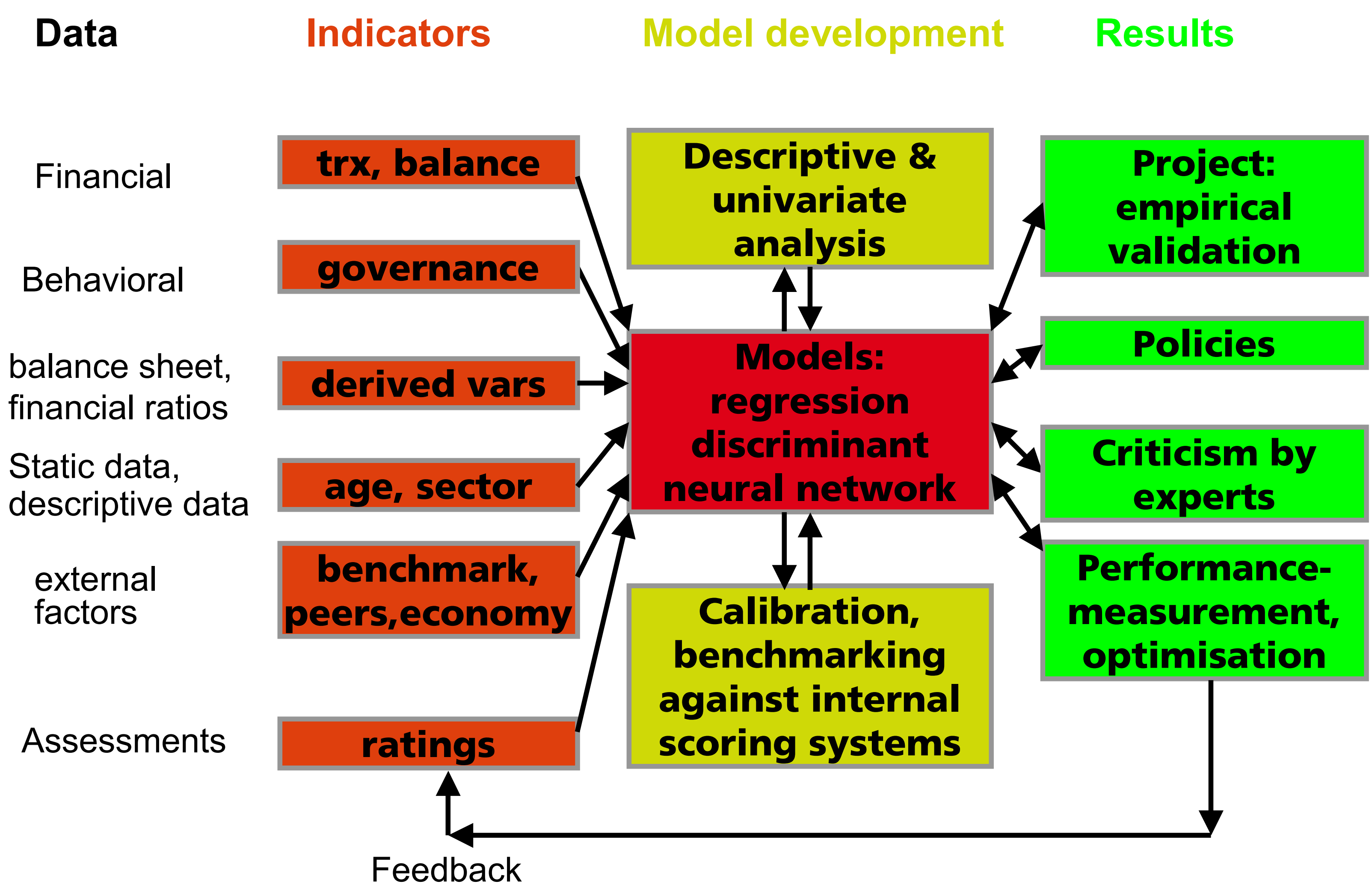
4. Clustering to evaluate if the groups found make sense.

5. Regression model & neural network model for the real analysis

6. Attempt to improve the model performance by segmentation

7. Validation of the model (holdback sample, Jacknife, another set of data)

Model



Data
Indicators
Model development
Results
Financial
Behavioral
balance sheet, financial ratios
Static data, descriptive data
external factors
Assessments
trx, balance
governance
derived vars
age, sector
benchmark, peers,economy
ratings
Descriptive & univariate analysis
Models: regression discriminant neural network
Calibration, benchmarking against internal scoring systems
Project: empirical validation
Policies
Criticism by experts
Performance-measurement, optimisation
Feedback

Results

Most significant predictors

- age of the company and age of the business relationship
- avg assets in recent months / avg assets over 1 year
- avg balance summed over all accounts
- sum of earnings / sum of expenditures
- sum of earnings / sum of loan facilities
- number of days in overdraft or excess mode
- intensity of the relationship with the bank
- delinquency events

Different models are required for

- industry sectors,
- ownership/management setups
- publicly traded/big companies

More results

Data history

In most cases a data history of 2-3 years is sufficient because changes in corporates introduce a lot of noise into the data

Prediction horizon

A prediction horizon of 1 year is ambitious but can be achieved.

Prediction criteria / target function

Instead of "critical situations": avoidance of bad debt & losses

More results

Human reasoning

The scoring differentiates well between perfectly good and bad cases, but it leaves 20-30% of cases in a gray zone

The prediction system must not be used as an automatic machine!

The bank's objective
- minimize amount of time needed to monitor undisputed cases
- allocate more time for investigation to uncertain cases

No layoff of qualified personnel in risk management.

Future development
- base the decision on even more data
- realtime decision making compatible with and built into the bank's workflows

Issues and problems

Will the model be constant over time?

- no, it has to be re-run every 1-2 years

Can the model predict a company's trends?

- if the model is validated against ratings, which contain prospective views, it automatically reflects the assessment of the future/trend

Which methodology is best?

- we prefer a multiplicity of variables, statistical&mining procedures;

A hybrid approach is recommended.

SAS / Enterprise Miner

All analyses were run with SAS on a big Unix server.

The warehouse contains a 3-6 years‘ history of business relevant data

The SAS Enterprise Miner facilitates the labour intensive research

A hybrid approach consisting of statistics and mining is required

Resumé

Default prediction is an academic playground - only partly usable.

Company defaults will always happen, some are unpredictable.

We should extend the scope from default to **recovery prediction**.

Managing the **financial turnaround** and the organizational change is a big challenge, and it is rewarding.

The End

Bankruptcy scoring using the Enterprise Miner

michael.wetzel@systematika.com