Challenges in Credit Risk Analytics

Prof. dr. Bart Baesens

Department of Decision Sciences and Information Management
K.U.Leuven (Belgium)

Vlerick Leuven Ghent Management School (Belgium)

School of Management
University of Southampton (United Kingdom)
Bart.Baesens@econ.kuleuven.be
Twitter: DataMiningApps
Facebook: Data Mining with Bart

Overview

- Introduction
- Data Quality
- Model requirements
- Model discrimination versus calibration
- Model validation

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
  - fraud risk models
  - insurance risk models
  - model risk metamodels (?)
  - …
- Analytics typically used to build all these models!
- Often subject to regulation (e.g. Basel II/Basel III, Solvency II, …)!
- Model errors directly affect profitability, solvency, shareholder value, macro-economy, …, society as a whole!

Traditional Analytics: Performance benchmarks

Context	Number of Characteristics	AUC ranges
Application Credit Scoring	10-15	70%-85%
Behavioural Credit Scoring	10-15	80%-90%
Fraud detection (insurance)	10-15	70%-90%
Churn detection (Telco)	6-10	60%-80%

BAESENS B., VAN GESTEL T., VIAENE S., STEPANOVA M., SUYKENS J., VANTHIENEN J., Benchmarking State of the Art Classification Algorithms for Credit Scoring, *Journal of the Operational Research Society*, 2003.

VERBEKE W., DEJAEGER K, MARTENS D., HUR J., BAESENS B., New insights into churn prediction in the telecommunication sector: a profit driven data mining approach, *European Journal of Operational Research*, 2011.

Improving Traditional Analytics: 2 strategies

- **Strategy 1**: Use complex modeling techniques
  - E.g. neural networks, support vector machines, random forests, …
  - Pro: powerful models (e.g. universal approximation)
  - Con: loss of interpretability, marginal performance gains
- **Strategy 2:** Enrich your data
  - External data (FICO score, bureau data, …)
  - Data quality!
  - Pro: model still interpretable
  - Con: additional resources needed (ICT)

Data Quality

- GIGO principle
  - Garbage in, Garbage out; messy data gives messy models
- In many cases, simple analytical models perform well, so biggest performance increase comes from the data!
  - Baesens et al., 2003; Van Gestel, Baesens et al., 2004
  - Holte, 1993
- Importance of Master Data Management and Data quality programmes!
- " The best way to improve the performance of an analytical model is not to look for fancy tools or techniques, but to improve DATA QUALITY first"
- Baesens B., It's the data, you stupid!, Data News, 2007.

Example data quality criteria

- Data accuracy
  - E.g., outliers
  - Age is 300 years versus Income is 1.000.000 Euro (not the same!)
- Data completeness
  - Are missing values important?
- Data bias and sampling
  - Try to minimise, but can never totally get rid of
- Data definition
  - Variables: what is the meaning of 0?
  - Target: fraud, churn, default, customer lifetime value (CLV), ...
- Data recency/latency
  - Refresh frequency

Data Quality Criteria (Moges, Lemahieu, Baesens, 2011)

Cat.	DQ dimensions	Definitions
Intrinsic	Accuracy (AC)	The extent to which data are certified, error-free, correct, flawless and reliable
	Objectivity (OBJ)	The extent to which data are unbiased, unprejudiced, based on facts and impartial
	Reputation (REP)	The extent to which data are highly regarded in terms of its sources or content
Contextual	Completeness (COM)	The extent to which data are not missing and covers the needs of the tasks and is of sufficient breadth and depth of the task at hand
	Appropriate-amount (APM)	The extent to which the volume of information is appropriate for the task at hand
	Value-added (VAD)	The extent to which data are beneficial and provides advantages from its use
	Relevance (REL)	The extent to which data are applicable and helpful for the task at hand
	Timeliness (TIM)	The extent to which data are sufficiently up-to-date for the task at hand
	Actionable (ACT)	The extent to which data is ready for use
Representation	Interpretable (INT)	The extent to which data are in appropriate languages, symbols, and the definitions are clear
	Easily-understandable (EU)	The extent to which data are easily comprehended
	Representational-consistent (RC)	The extent to which data are continuously presented in same format
	Concisely-represented (CR)	The extent to which data is compactly represented, well-presented, well-organized, and well-formatted
	Alignment (AL)	The extent to which data is reconcilable
Access	Accessibility (ACC)	The extent to which data is available, or easily and swiftly retrievable
	Security (SEC)	The extent to which access to data is restricted appropriately to maintain its security
	Traceability (TRA)	The extent to which data is traceable to the source

Survey: data quality for credit risk analytics

- 50+ banks participating world-wide
- Focus on credit risk analytics
- Initial findings:
  - Most banks indicated that between **10-20 percent** of their data suffer from data quality problems
  - **Manual data entry** one of the key problems
  - Diversity of data sources and **consisent** corporate wide data representation main challenge for data quality
  - **Regulatory compliance** key motive to improve data quality
- Moges, Lemahieu, Baesens, 2011

Data quality: short term versus long term impact

- No short term solution
  - Deal with in a statistical way using e.g. data transformations
    - Outlier truncation, missing value imputation, data enhancement
  - Buy external data (data poolers!)
- Structural solutions in the long term
  - Re-design data entry processes
  - Master data management

Analytic Model Requirements

- **Statistical performance**
  - Lift curve, ROC curve, Gini coefficient, …
  - R-squared, MSE, …
- **Interpretability + Justifiability**
  - Very subjective, but CRUCIAL!
  - Often need to be balanced against statistical performance
- **Operational efficiency**
  - How much effort is needed to evaluate/monitor/re-train the model(s)?
- **Economical cost**
  - What is the cost to gather the model inputs and evaluate the model?
  - Is it worthwhile buying external data and/or models (e.g. BKR score)?
- **Regulatory compliance**
  - In accordance with regulation and legislation
  - E.g., Basel II\Basel III, Solvency II

Model discrimination versus Model calibration

- **Model discrimination**
  - Rank order (score) entities with respect to likelihood of event occurring
  - Examples
    - Rank order customers in terms of likelihood to default on their obligation
    - Bart is more risky to default than Victor!
  - However, despite traditional focus in data mining, this is no longer sufficient!
  - We need to know the **EXACT** probability of the event occurring!
- **Model calibration**
  - Provide well-calibrated and accurate projected probabilities based on
    - Historical data
    - Expectations with respect to the future (e.g. GDP contraction versus expansion)
  - Losses only make sense in an **ABSOLUTE** way!
  - Example
    - P(Bart defaults)=0.90; P(Victor defaults)=0.75

BRING THE MACRO-ECONOMY INTO THE MODEL!

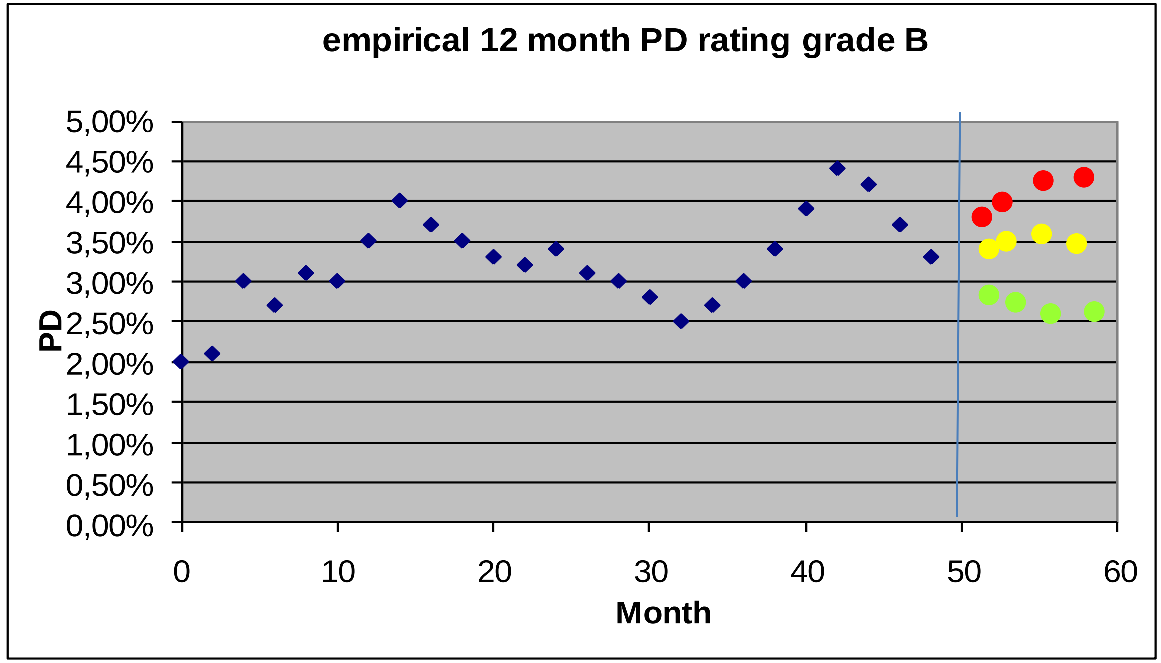
Model discrimination versus Model calibration

Model Discrimination

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

Example application scorecard

Model Calibration

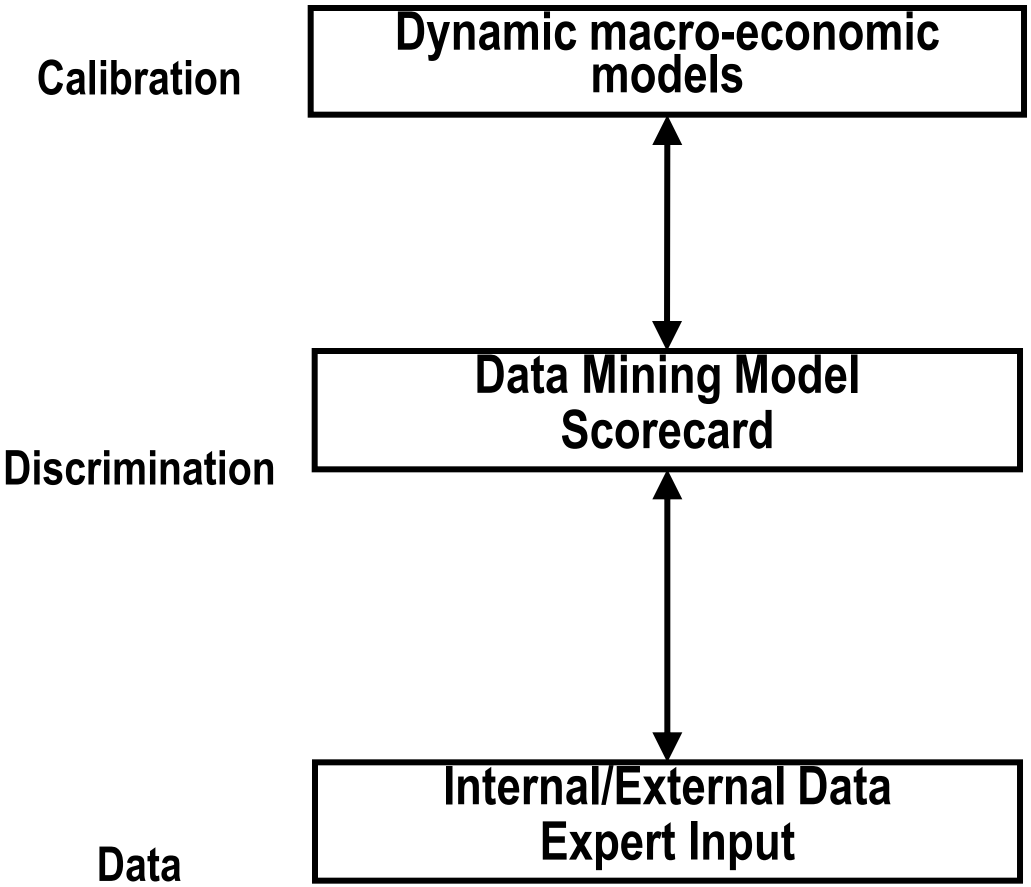


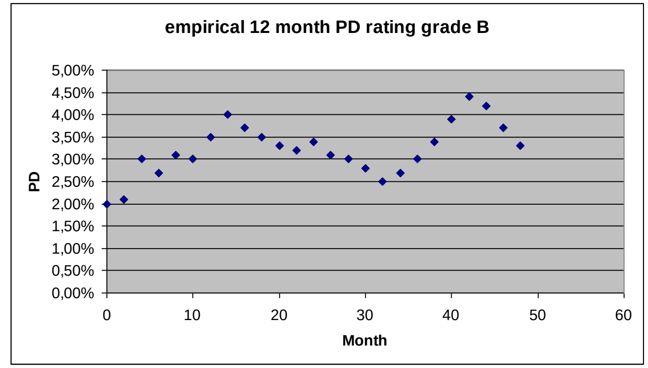
Historical probability of default (PD) calibration for customer segment B!

Model Calibration: example approach

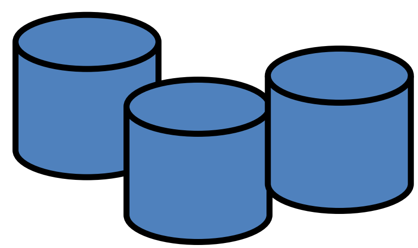
- Analytical models typically built using a snapshot at a given period in time!
- Cluster data mining model outputs (e.g. scores) into pools
  - Scores are too fine granular anyway!
  - Essentially, a semi-supervised learning exercise
  - Score 200-300: pool A; score 301-500: pool B, score 501-650: pool C, …
- For each pool, calibrate event probability using
  - Time series analysis techniques (ARIMA, VAR, …)
  - Dynamic models/Markov Chains
  - Simulations
  - Projected macro-economic scenarios
- 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) analytical model?

Summarising: Model architecture





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



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
  - E.g. assume all credit scores decrease by 5%
- Scenario analysis
  - Historical or hypothetical
  - E.g. 3 successive years of GDP contraction, house prices drop by 5%, …
  - Could be a 1/25 years event (e.g. in the United Kingdom)
- Common challenges/problems:
  - Lack of historical data
  - Correlations break down during stress (need to have data on downturn periods)
  - Integrate risks
  - What is stress??
  - What to do with the results? Strategic impact?

Model Risk

- "Essentially, all models are wrong, but some are useful" (George E. P. Box, 1987)
- Models are not perfect, some are actually VERY bad, but what's the alternative???
  - Default risk/fraud prediction: **good** performance (Gini coefficients around 50 to 80%)!
  - Loss/LGD prediction: **awful** performance (R^2 of 0 .30 already great!)
- Model imperfection is typically dealt with by
  - Conservative parameter calibration
    - aka **economic downturn** calibration
      - E.g. assume statistically estimated probability of default is 3%.
      - Use 5% for strategic decisions to capture model risk!
  - Create equity buffer/provisions for model risk
    - Hard to quantify!

Model monitoring

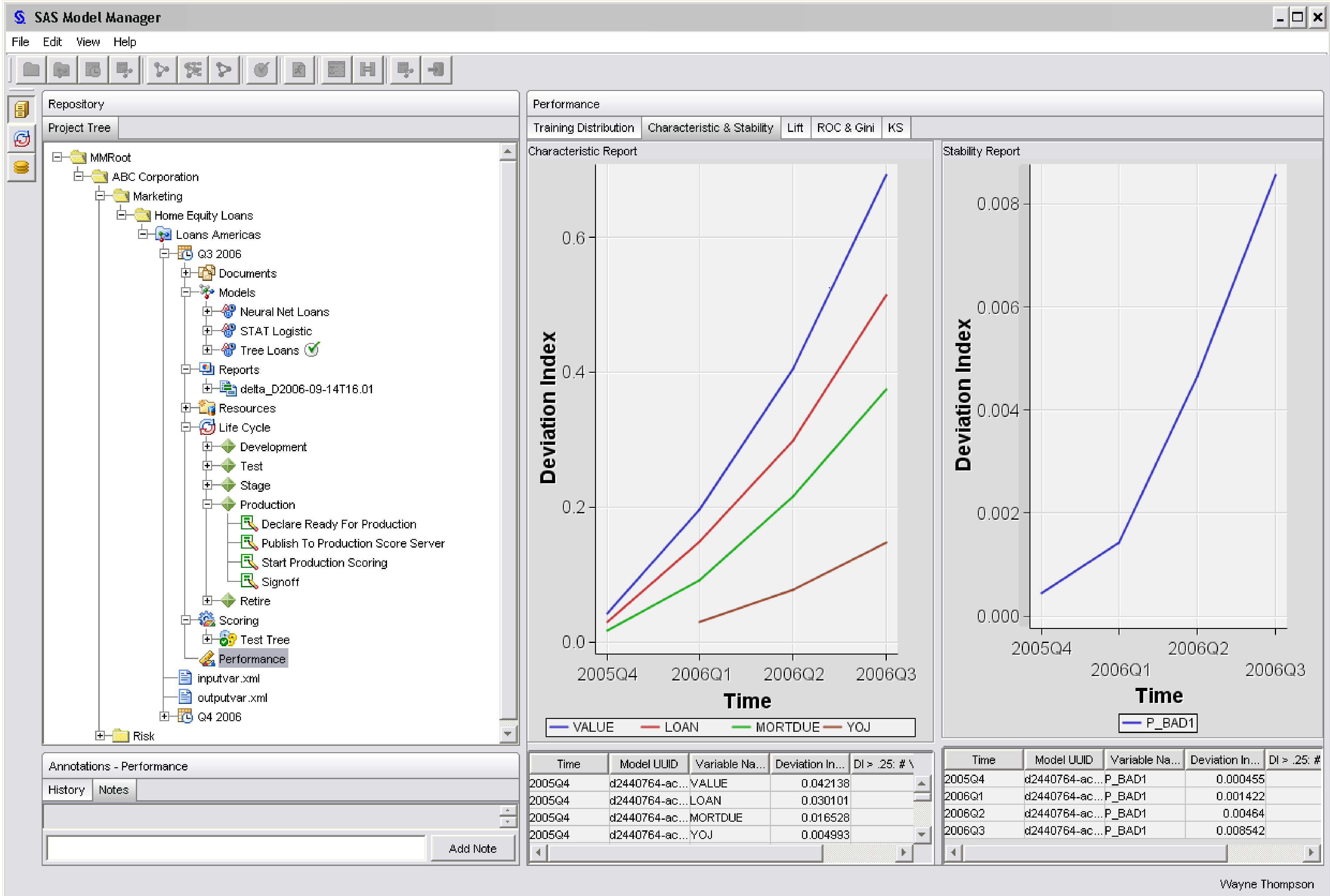
- Why data mining models may degrade in performance?
  - Sample effects (models estimated on limited samples)
  - Macro-economy (downturn versus upturn)
  - Internal effects (e.g. strategy change, population drift, M&A)
  - In reality: a very nice (?) mixture of these!
- Need to constantly monitor outcomes of models
- Crucial since models more and more steer strategic decisions of the firm (cf. supra)
  - E.g. equity calculation in a Basel II/Solvency II environment
  - Risk based pricing
- Quantitative versus Qualitative validation

Model validation

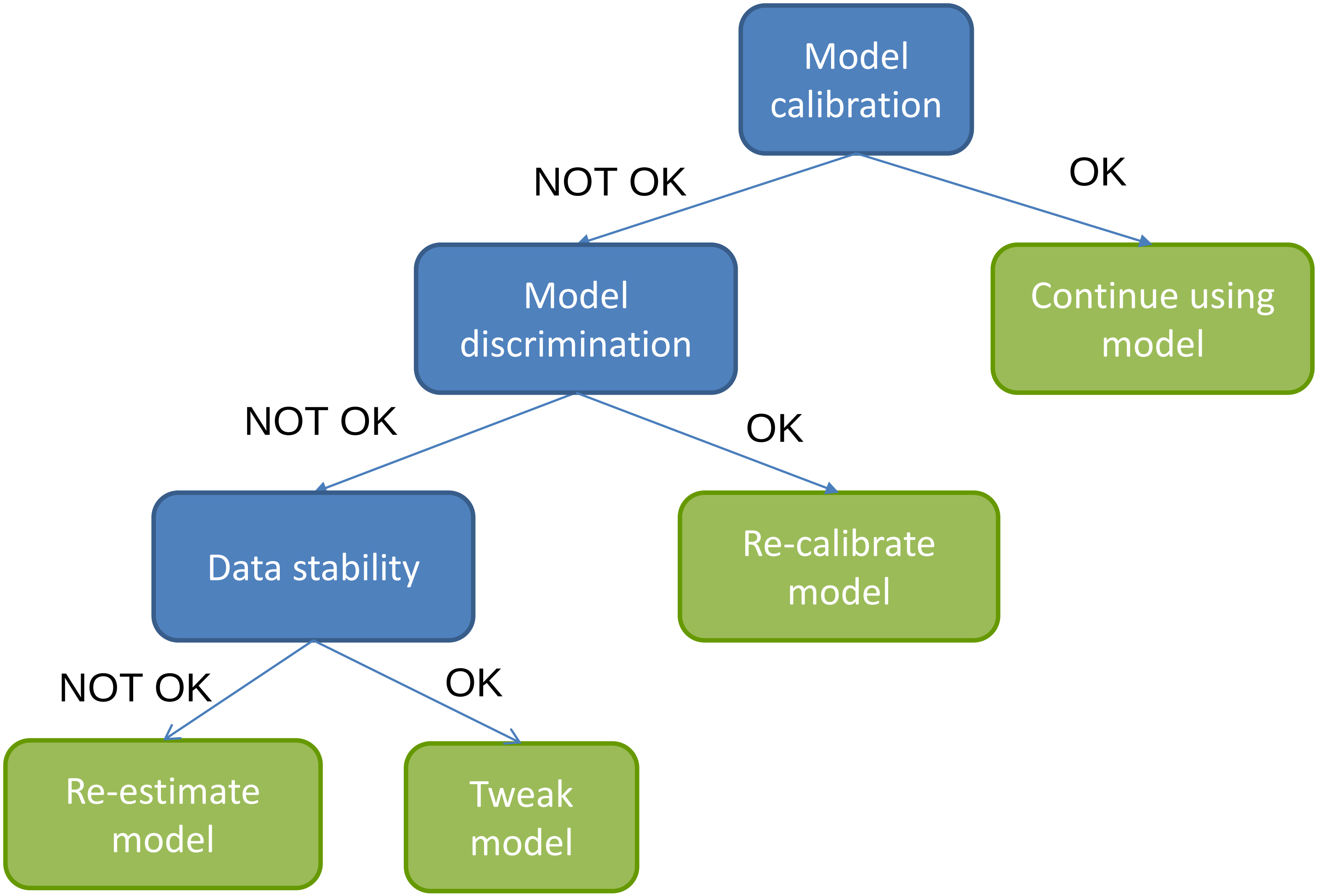
- Quantitative validation
  - Backtesting
  - Benchmarking
- Qualitative validation
  - Data quality
  - Model design
  - Documentation
  - Corporate governance and management oversight

Backtesting

- Contrasting ex-post realised numbers with ex-ante predictions
- Using statistical tests and performance measures
- Examples
  - Use binomial test for comparing default/fraud rates
  - Monitor decrease in AUC (Gini) over time
- Challenges
  - Which test statistics to use?
  - Which confidence levels to adopt?
  - How to deal with correlated behavior (portfolio effects)?
  - When to take action and what action?


SAS Model Manager
File Edit View Help
Repository
Project Tree
MMRoot
ABC Corporation
Marketing
Home Equity Loans
Loans Americas
Q3 2006
Documents
Models
Neural Net Loans
STAT Logistic
Tree Loans
Reports
delta_D2006-09-14T16.01
Resources
Life Cycle
Development
Test
Stage
Production
Declare Ready For Production
Publish To Production Score Server
Start Production Scoring
Signoff
Retire
Scoring
Test Tree
Performance
inputvar.xml
outputvar.xml
Q4 2006
Risk
Annotations - Performance
History
Notes
Add Note
Performance
Training Distribution
Characteristic & Stability
Lift
ROC & Gini
KS
Characteristic Report
Deviation Index
0.0
0.2
0.4
0.6
2005Q4
2006Q1
2006Q2
2006Q3
Time
VALUE
LOAN
MORTDUE
YOJ
Stability Report
Deviation Index
0.000
0.002
0.004
0.006
0.008
2005Q4
2006Q1
2006Q2
2006Q3
Time
P_BAD1
Time | Model UUID | Variable Na... | Deviation In... | DI > .25: # \
2005Q4 | d2440764-ac... | VALUE | 0.042138
2005Q4 | d2440764-ac... | LOAN | 0.030101
2005Q4 | d2440764-ac... | MORTDUE | 0.016528
2005Q4 | d2440764-ac... | YOJ | 0.004993
Time | Model UUID | Variable Na... | Deviation In... | DI > .25: #
2005Q4 | d2440764-ac... | P_BAD1 | 0.000455
2006Q1 | d2440764-ac... | P_BAD1 | 0.001422
2006Q2 | d2440764-ac... | P_BAD1 | 0.00464
2006Q3 | d2440764-ac... | P_BAD1 | 0.008542
Wayne Thompson

Action plans



Key lessons learnt

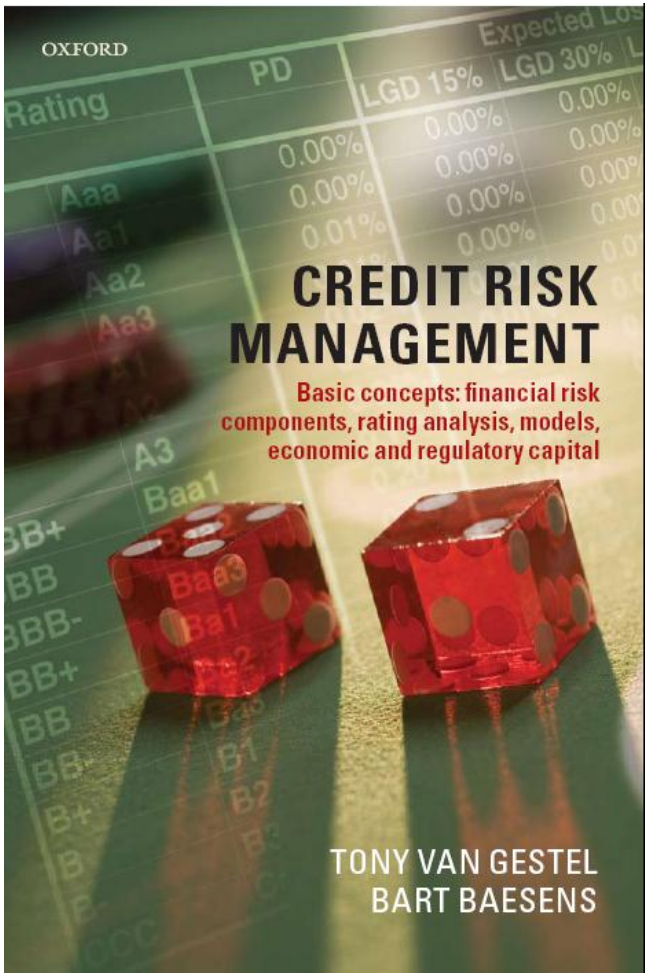
- The best way to improve the performance of an analytical model is to improve **data quality** first
- A good model does more than giving good statistical performance (**model requirements**)!
- Discrimination versus calibration: bring the **macro-economy** into the model!
- Introduced the idea of **model risk**
- Discussed the need for **model validation** and **action plans**!

References

- Moges H.T., Dejaeger K., Lemahieu W., Baesens B., A Total Data Quality Management for Credit Risk: New insights and challenges, *International Journal of Information Quality*, forthcoming, 2012.
- Verbraken T., Verbeke W. Baesens B., A Novel Profit Maximizing Metric for Measuring Classification Performance of Customer Churn Prediction Models, *IEEE Transactions on Knowledge and Data Engineering*, forthcoming, 2012.
- BAESENS B., MARTENS D., SETIONO R., ZURADA J., White Box Nonlinear Prediction Models, editorial special issue, *IEEE Transactions on Neural Networks*, Volume 22, Number 12, pp. 2406-2408, 2011.
- Baesens B., Mues C., Martens D., Vanthienen J., 50 years of Data Mining and OR: upcoming trends and challenges, *Journal of the Operational Research Society*, Volume 60, pp. 16-23, 2009.
- Glady N., Croux C., Baesens B., Modeling Churn Using Customer Lifetime Value, *European Journal of Operational Research*, Volume 197 Number 1, pp. 402.411, 2009.
- Martens D., Baesens B., Van Gestel T., Decompositional Rule Extraction from Support Vector Machines by Active Learning, *IEEE Transactions on Knowledge and Data Engineering*, Volume 21, Number 1, pp. 178-191, 2009.
- Setiono R., Baesens B., Mues C. Recursive Neural Network Rule Extraction for Data with Mixed Attributes, *IEEE Transactions on Neural Networks*,19 (2),pp.299-307, 2008.
- Baesens B., Setiono R., Mues C., Vanthienen J., Using Neural Network Rule Extraction and Decision Tables for Credit-Risk Evaluation, *Management Science*, Volume 49, Number 3, pp. 312-329, March 2003

- See www.dataminingapps.com

Bart’s Book



Contents