# Development of Scorecards in Retail Banking

Sabine Detering

Eurobanking 2009

readybank

# readybank in a Nutshell

- Founded in 1929
- Owned by WestLB AG since March 2006
- About 300 employees (including service company)
- Balance sheet total: 400 Mio. Euros
- Headoffice in Berlin, large office in Düsseldorf
- Specialised in consumer lending, main distribution channels:
  - local savings banks (strategic partnership)
  - automotive

# Typical Steps of Scorecard Development

- Choice of method
- Choice of data
- Performance definition
- Data preparation
- Univariate analyses
- Multivariate analyses
- Rejection inference
- Final model

# Choice of Method

- Expert model
- Statistical model (logistic regression, decision trees,...)
- Artificial intelligence (neural networks, genetic programming,...)

→ Standard approach: expert based, categorial, logistic regression

# Choice of Data: Data History (I)

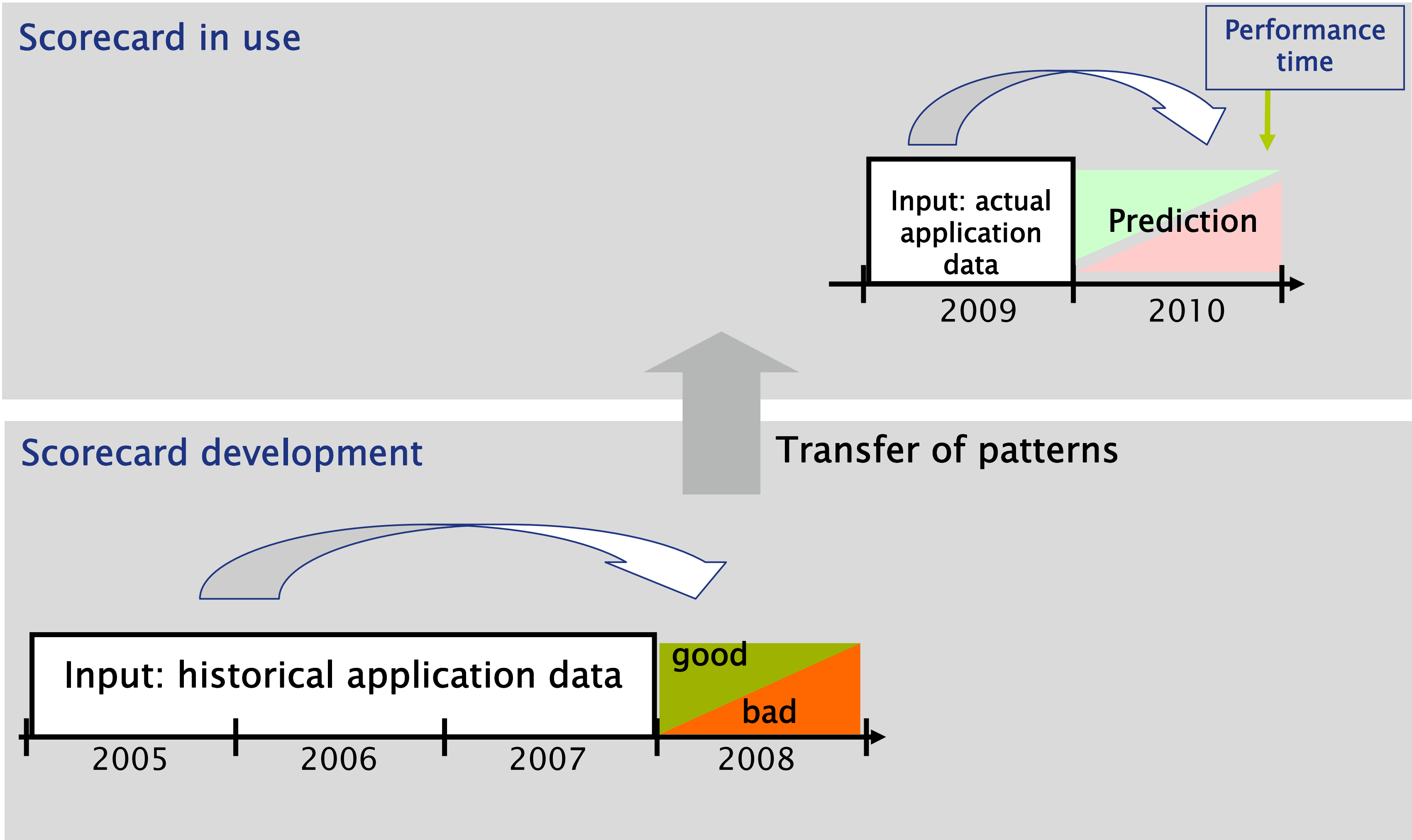
**Generally**: The more data the better!

⇕

**But**: Longer data history often means less representative data for the future

due to changes in target groups, policy rules, collection processes,...

# Choice of Data: Data History (II)



# Choice of Data: Data History (III)

- **Solution**: Use rather long data history and weight data according to future application structure or just weight recent applications stronger

- **Example**: Application scorecard for car loans

  The difference between the historical and the future portfolio can be characterised by the age of the vehicle

Weighting:

| Age of vehicle | Weight |
|---|---|
| Car <0.5 years | 1,66366 |
| 0.5 years<= car <5 years | 0,97852 |
| Car > 5 years | 0,90451 |
| Other vehicle | 1,0 |

# Choice of Data: Potential Predictors

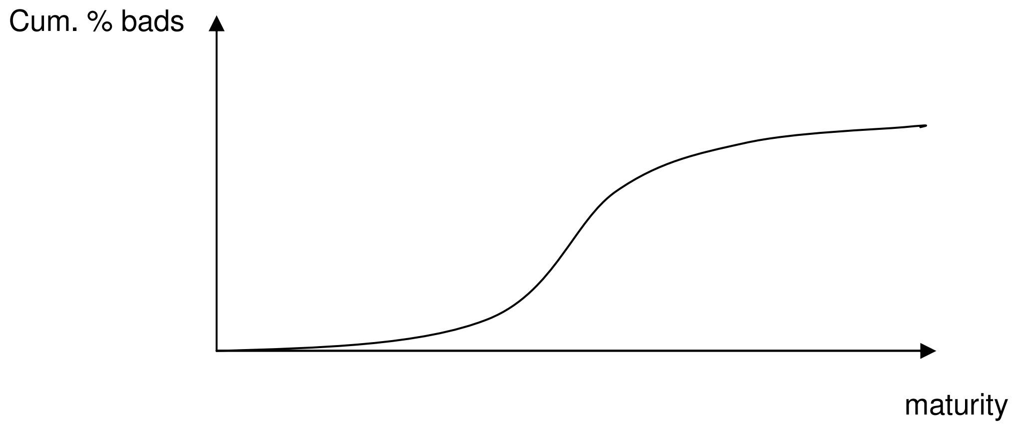
The development data should cover a wide range of potential predictors from different areas.

- **Client specific**: age, marital status, profession, time in job, type of accomodation, time since last move, income, bank account, credit bureau data,...
- **Loan/product specific**: Loan amount, maturity, purpose of the loan, type of security,...
- **Combined predictors**: instalments/income, sum of all loans/income, amount paid in cash/loan amount,...

# Choice of Data: Performance Definition

- Criteria: e.g. 60 days past due, last dunning, cancellation, Basel II criteria
- Time horizon: within 12 (18, 24) months or life time of the loan

Typically , there is a certain saturation of ‚bad' accounts after 12-24 months, depending on the type of loan.



# Data preparation

Checking data quality

- Completeness
- Correctness/consistency

→ Exclusions

Preparing data for model development

- Standardise formats (e.g. car brands)
- Build combined predictors (e.g. ratios, interaction variables)
- Build binary performance variable

# Typical Steps of Scorecard Development

- Choice of method
- Choice of data
- Performance definition
- Data preparation
- **Univariate analyses**
- Multivariate analyses
- Rejection inference
- Final model

# Univariate Analysis: Classing Predictors

**Few classes**: - Model more stable

- Less risk of overfitting (random effects on small classes are smoothened)

**Many classes:** Better discriminatory power of the predictor and the model

# Univariate Analysis: Classing strategies

Categorial predictors

- Start with all categories
- Combine (small) classes based on expert knowledge and risk

Continuous predictors

- Start with fine classes, e.g. 5% or 10% quantiles, or standard classes e.g. ranges of 5 years for age
- Combine small classes based on expert knowledge and risk
- Split big classes
- Eliminate inconsistencies by re-classing

**Check change of discriminatory power after each step!**

# Mulivariate Analysis: Choice of predictors (I)

**Few predictors**:- Lower costs for implementation and maintenance

- Less risk of overfitting
- Less information to ask the client

⇕

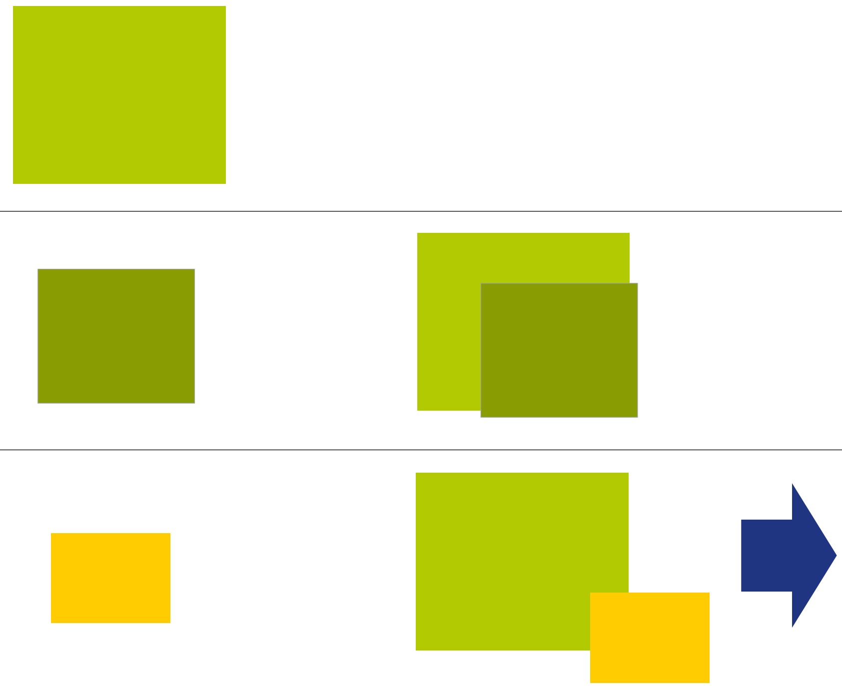
**Many predictors:** Better discriminatory power of the model

**Not every predictor with good univariate discriminatory power is useful in a multivariate model!**

# Mulivariate Analysis: Choice of predictors (II)

## Added value to multivariate models

| | Univariate | Multivariate |
|---|---|---|
| Age | | |
| Time in job | | |
| Purpose of loan | | |



Lower univariate discriminatory power
but higher added value

# Mulivariate Analysis: Predictor Selection Strategy

- Build groups of correlated predictors
- Initial model:
  - Select the strongest predictor from each group
  - Cover all aspects of the application
- Stepwise logistic regression (e.g. using SAS)
- Calculate discriminatory power and check Gini-curve of the resulting model
- Check consistency of scores for the different categories of one predictor
- Exchange predictors with others from the same correlation group and check effects on significance of predictors, discriminatory power and scores per category

# Rejection inference (I)

- Rejected applications are excluded from the development data because the performance information is missing

➡ Bias!

- Rejection inference provides a way to include the typical characteristics and patterns of rejected applications in the model development
- Two steps:
  – First develop an interim-model to ‚predict‘ the performance of the rejected clients
  – Then develop the final model based on all data incl. rejected clients

# Rejection inference (II)

- Build interim scorecard based on accepted applications
- Score all rejected applications with the scorecard
- Duplicate rejected applications and mark one record as ‚good‘ and one as ‚bad‘
- Weight the ‚bad‘ record with the bad rate associated with the calculated score (vice versa for the ‚good‘ record)
- Redo classing and predictor selection on the complete data set and develop the final model

# Predictor Selection for Final Model

Aspects to consider:

- (Future) availability
- Cost of data (credit bureau data can be very expensive!)
- Consistency of risk structure
- Correlation with other predictors in the model
- Additional discriminatory power
- Stability (e.g. ratios are more stable than absolute amounts)
- Legal aspects (laws against discrimination)

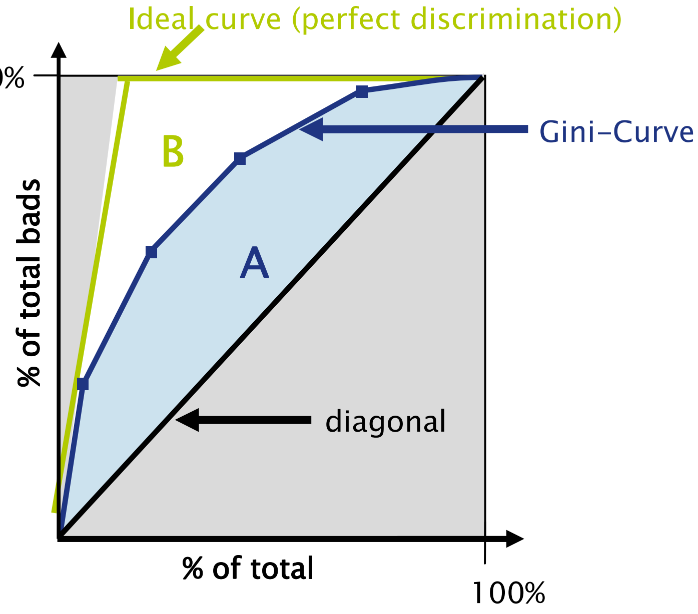
Sabine Detering
Risikocontrolling
Völklinger Str. 4
40219 Düsseldorf

Phone: +49 211 4955-2232
Email: sabine.detering@readybank.de

# Backup: Discriminatory Power

- **Gini coefficient** (Accuracy Ratio, Power)

$$\text{Gini} = \frac{A}{A+B}$$




- **Information value**:

$$\text{Weight of Evidence} = \ln\left(\frac{\%\text{ of total goods}}{\%\text{ of total bads}}\right)$$

$$\text{Information Value} = \sum_{\text{all classes}} \left(\%\text{ of total goods} - \%\text{ of total bads}\right) * \text{WoE}$$