General Approximators for Credit Scoring: Practical Considerations

Thomas Hill, Ph.D., StatSoft Inc.
Vladimir Rastunkov, Ph.D., StatSoft Inc.
Knut Opdal, StatSoft Norway AS

2013 Credit Scoring and Credit Control XIII Conference

StatSoft®

Making the World More Productive®

Headquarters: StatSoft, Inc. • 2300 E 14 St • Tulsa, OK 74104 • USA • +1(918) 749-1119 • Fax: (918) 749-2217 • info@statsoft.com • www.statsoft.com

Australia: StatSoft Pacific Pty Ltd.
Brazil: StatSoft South America Ltda.
Bulgaria: StatSoft Bulgaria Ltd.
Chile: StatSoft South America Ltda.
China: StatSoft China
Czech Rep.: StatSoft Czech Rep. s.r.o.
France: StatSoft France
Germany: StatSoft GmbH
Hungary: StatSoft Hungary Ltd.
India: StatSoft India Pvt. Ltd.
Israel: StatSoft Israel Ltd.
Italy: StatSoft Italia srl
Japan: StatSoft Japan Inc.
Korea: StatSoft South Korea
Netherlands: StatSoft Benelux
Norway: StatSoft Norway AS
Poland: StatSoft Polska Sp. z o.o.
Portugal: StatSoft Iberica Lda
Russia: StatSoft Russia
S. Africa: StatSoft S. Africa (Pty) Ltd.
Spain: StatSoft Iberica Lda
Sweden: StatSoft Scandinavia AB
Taiwan: StatSoft Taiwan
UAE/Egypt: StatSoft Middle East
United Kingdom: StatSoft Ltd.

Agenda

- Statistical Modeling vs. General Predictive Modeling and Pattern Recognition
- Pattern Recognition Algorithms
  - The data are the model
  - Interpretability of results
  - What-if or scenario analysis, and reason scores
  - Caveats
  - Practical considerations
- Summary

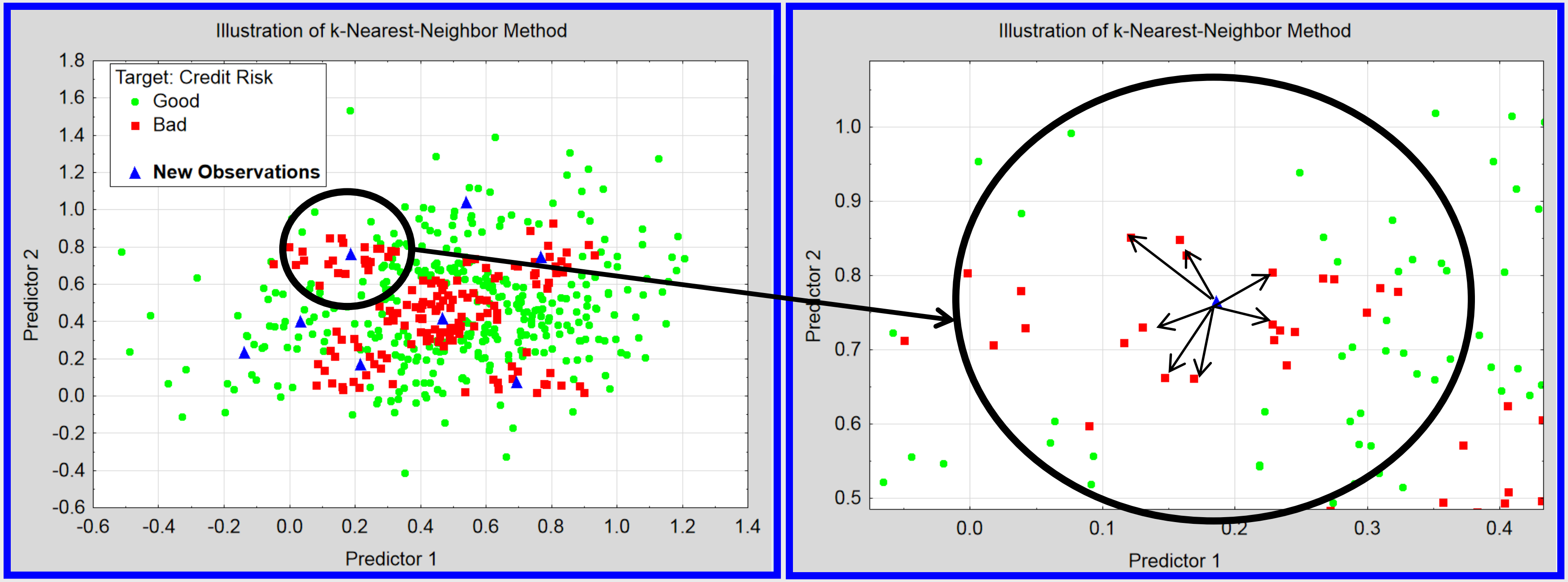
Overview:
Data Mining and Statistical Modeling

Knowledge Discovery vs. Statistical Analysis

- **Statistical Analysis**
  - Focuses on “hypothesis testing” and “parameter estimation”
  - Fits “parsimonious statistical models” with the goal to “explain” complex relationships with fewer parameters
  - Examples: Regression, nonparametric statistics, factor analysis, traditional quality control
- **Data Mining**
  - **The data are your model!**
  - Focuses on knowledge discovery, detection of patterns, clusters, and so on; we only have data and no (or few) expectations and hypotheses
  - Fits simple models or complex models (such as neural nets) to enable valid prediction
  - Examples: K-nearest-neighbor methods, recursive partitioning (trees), neural nets, stochastic gradient boosting of tree classifiers, random forests, support vector machines
  - General approximators

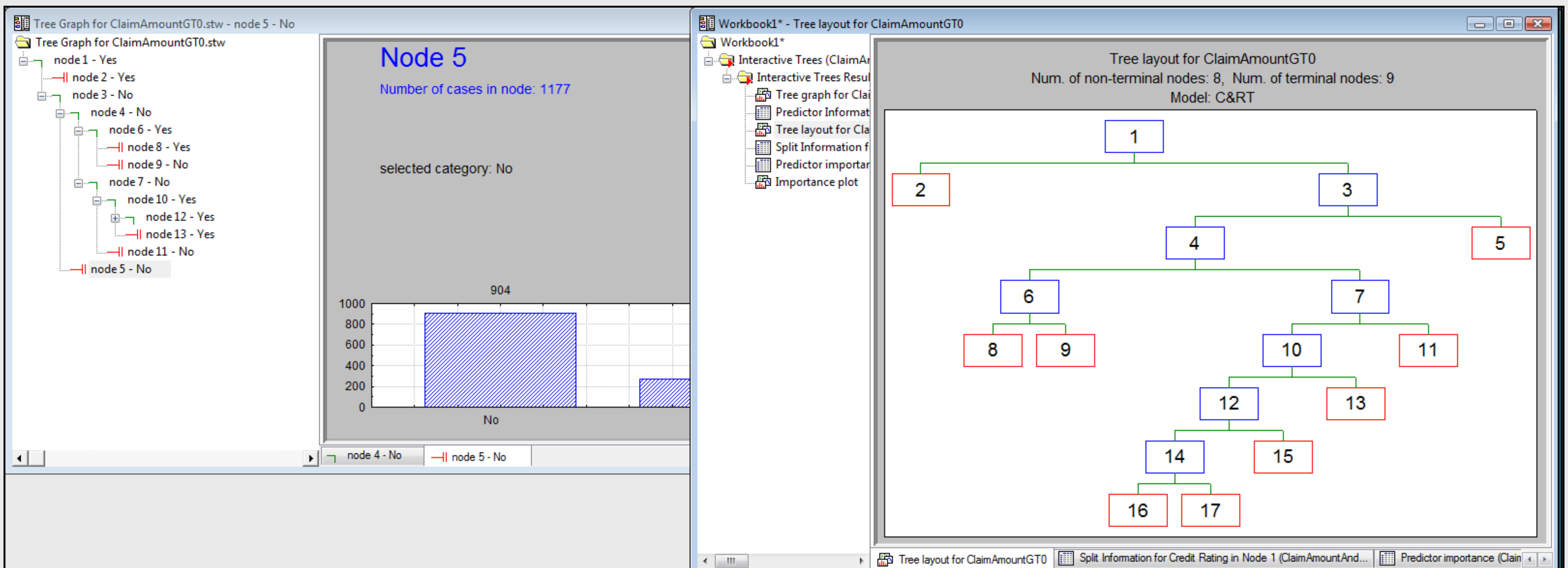
The Data Are Your Model: Predicting "Missing Data"

- E.g., consider prediction as a "missing data" problem
  - Consider imputing a prediction for future (not yet observed) credit default using *k-Nearest-Neighbor* methods
  - Missing data (not yet observed) are replaced with the observed (non-missing) values of similar or "neighboring" observations
  - i.e., "*The data are your model*"



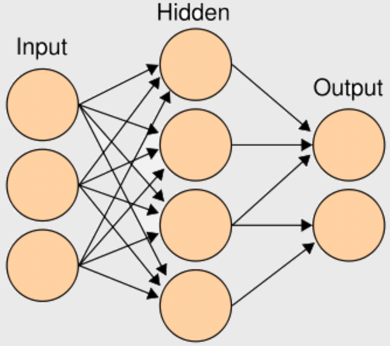
Recursive Partitioning or Decision Trees

- Recursive partitioning, or decision trees (C&RT, CHAID)
- Yields prediction models that are easily deployed
- Model predictions will not be based on extrapolation, but rather by “assigning” new cases to the “closest” partition (repeated pattern) derived from historical data
- The question “what predictors are important” is difficult to answer



General Approximators for Predictive Modeling: Neural Nets, boosting, etc.

- Neural networks, support vector machines
  - Effectively highly nonlinear models/equations to predict continuous or categorical outcomes
  - Will give models with continuous response functions
  - Generally considerered difficult to interpret "black box"



- Boosting, Bagging, Ensembles, Metalearning
  - *Boosting* is the repetitive application of a general approximator to consecutive residuals of preceding steps; thus the algorithm can focus in consecutive steps on the cases not well represented in the preceding step
  - *Bagging* is the (weighted) averaging of predictions from the application of a general approximator to different subsets of the data (variables, observations)
  - *Ensembles* describes the (weighted) averaging of predictions from the application of different learning algorithms to the same data.
  - Often, these methods can improve the quality/accuracy of predictions

Reason Scores via What-If or Scenario Analysis

- The algorithms of data mining and general predictive modeling are complex, yielding complex predictions
- But it is possible to derive reason scores that *explain* a specific prediction via what-if or scenario analysis
  - Effectively computing numeric derivatives at the point of the observed response
  - This is equivalent to the interpretation of coefficients in linear logistic regression models (for risk)
- For example, consider the case of predicting risk of default[1]
- Given any prediction model of risk, by changing ("moving") one-predictor-at-a-time, while holding others constant at the observed values, the effects of each predictor on expected risk can be computed

[1]http://www.statsoft.com/portals/0/products/data-mining/credit_scoring.pdf

Reason Scores via What-If or Scenario Analysis

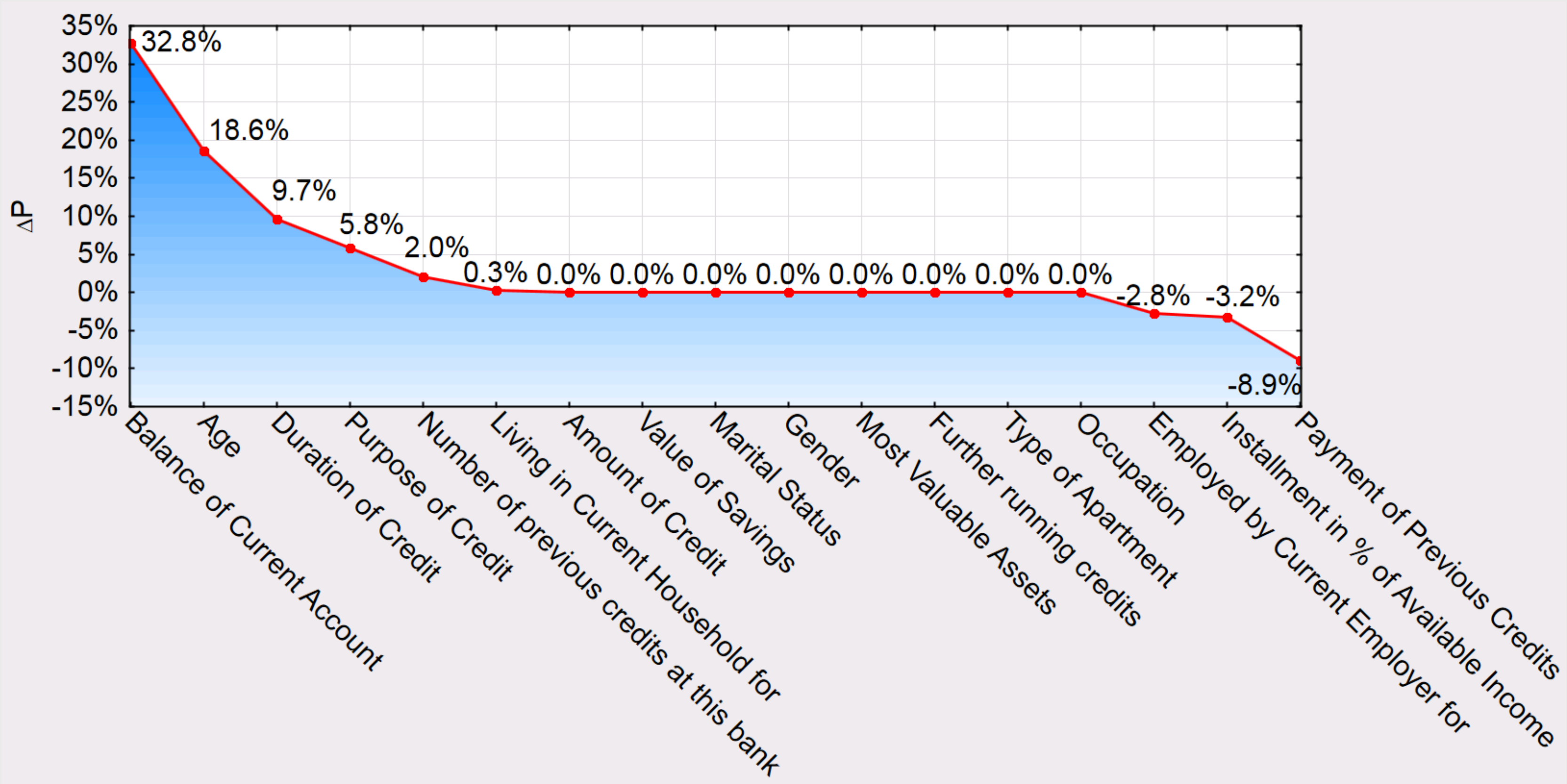
- Given any prediction model of risk, by changing (“moving”) one-predictor-at-a-time, while holding others constant at the observed values, the effects of each predictor on expected risk can be computed

No	Parameter	Value	No	Parameter	Value
1	Balance of Current Account	no running account (>$300)	10	Gender	male (male)
2	Duration of Credit	36 (18)	11	Living in Current Household for	< 1 year (>8 years)
3	Payment of Previous Credits	no problems with current credits (no previous credits)	12	Most Valuable Assets	life insurance (life insurance)
4	Purpose of Credit	retraining (furniture)	13	Age	22 (31)
5	Amount of Credit	3003 (3148.6)	14	Further running credits	no further running credits (no further running credits)
6	Value of Savings	no savings (no savings)	15	Type of Apartment	rented (rented)
7	Employed by Current Employer for	5-8 years (1-5 years)	16	Number of previous credits at this bank	2- 4 (one)
8	Installment in % of Available Income	25-35 (< 15)	17	Occupation	skilled employee (skilled employee)
9	Marital Status	single (single)			

Note. Value in parenthesis represents the baseline (median for continuous, mode for categorical factors)

Reason Scores via What-If or Scenario Analysis

- Given any prediction model of risk, by changing ("moving") one-predictor-at-a-time, while holding others constant at the observed values, the effects of each predictor on expected risk can be computed



Reason Scores via What-If or Scenario Analysis

- Top 3 recommendations from effects of each predictor “top 3”

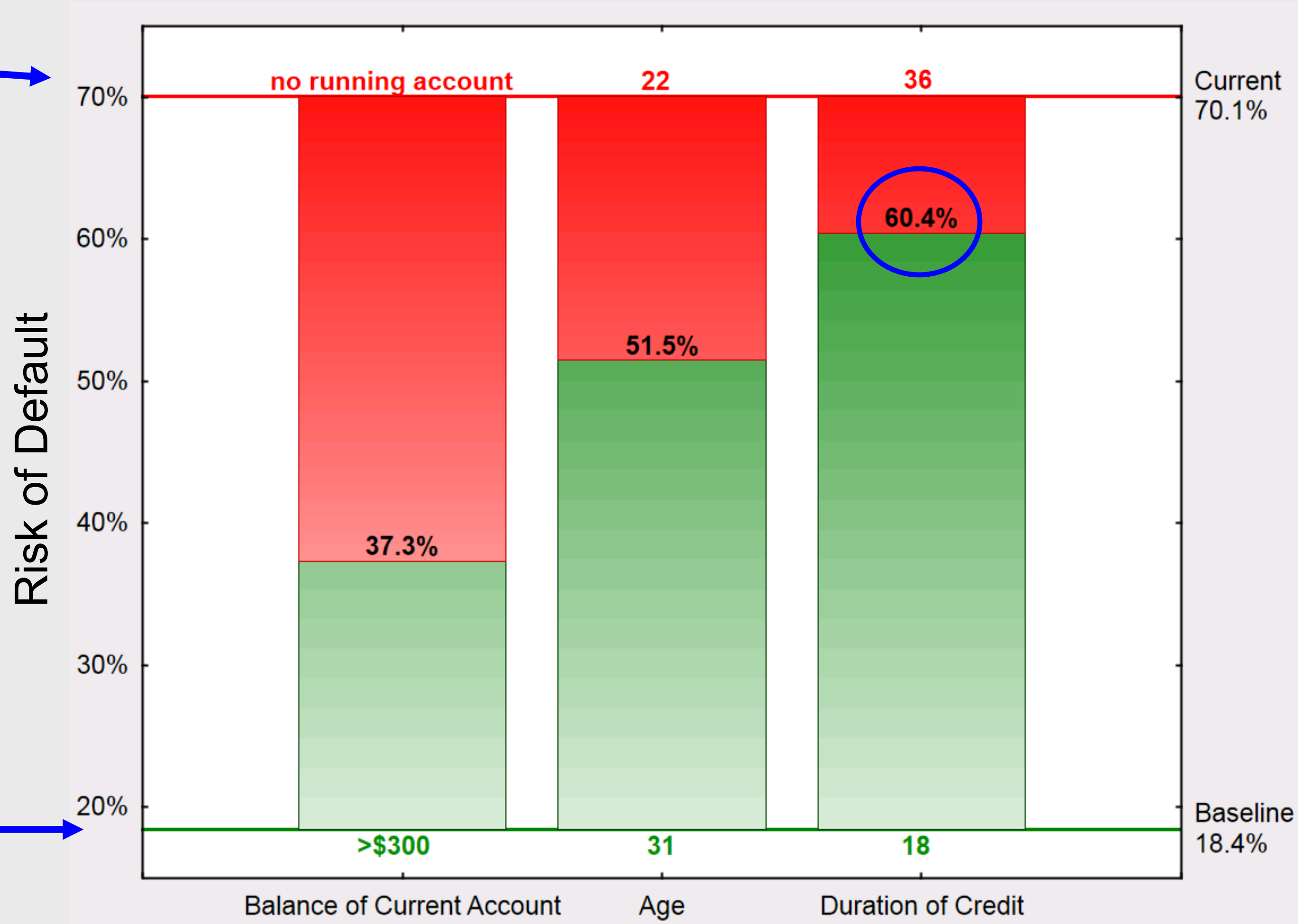
Current position

IF $\begin{matrix}\mathit{Duration\ of} \\ \mathit{Credit}\end{matrix} = 18$

IF $Age = 31$

IF $\begin{matrix}\mathit{Balance\ of} \\ \mathit{Current\ Account} \\ " > \$300"\end{matrix}$

Target (baseline)

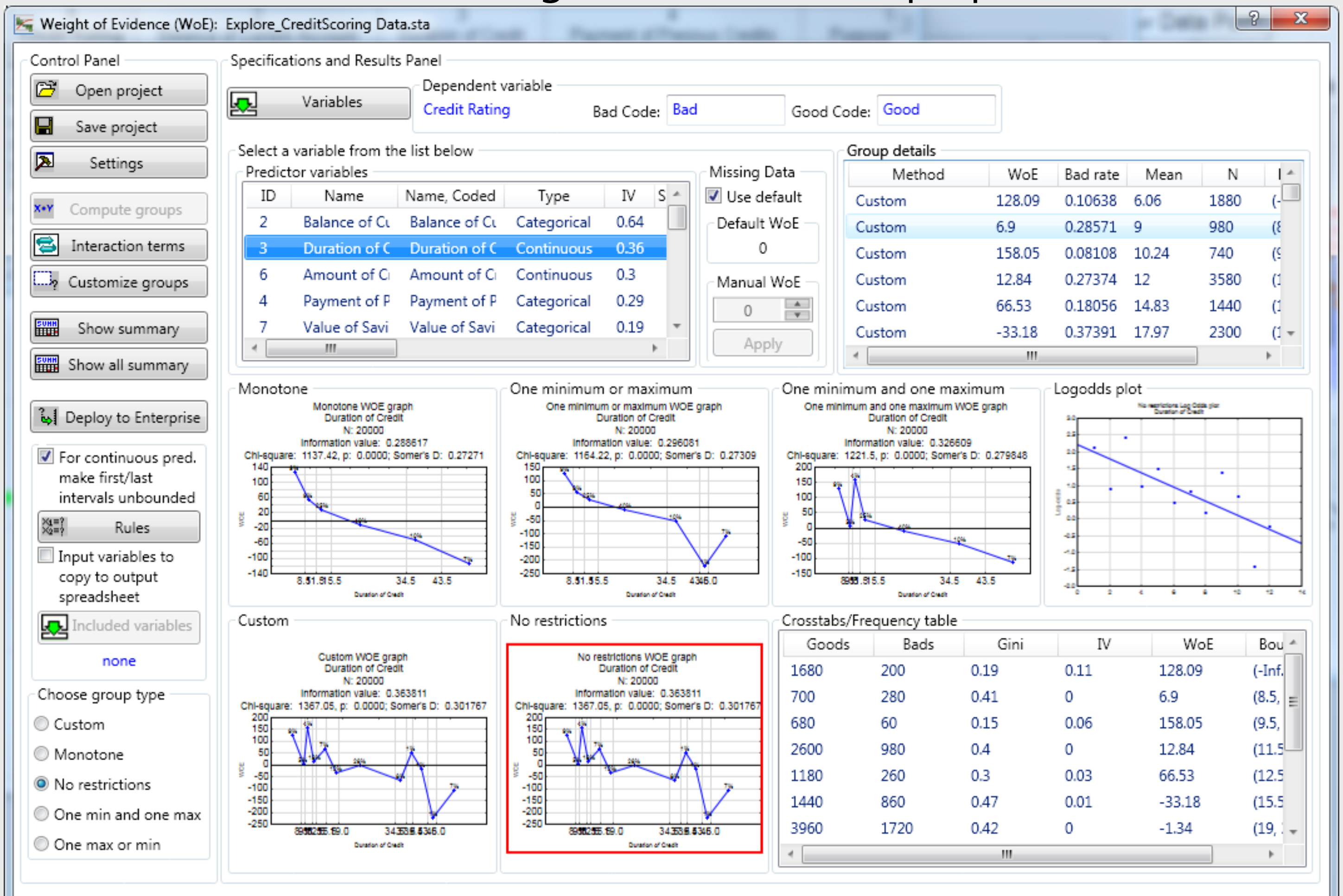


Caveats

- While one-predictor-at-a-time partial derivative at the point of the new observation can provide insights into the reasons for a particular prediction, it may also lead to falsely simplistic interpretations of the how a model “functions”
  - Interaction effects: Only the combination of specific value ranges for two or more predictors affect credit default probability
  - Nonlinear effects: The relationship between the values in a predictor to default probability are not monotone, but complex
    - There, simple statements cannot be derived such as: *“The values of predictor A were lower, then default risk would decrease”*
  - The partial derivatives evaluated at the point of observed new response can be entirely different than the partial derivatives evaluated at the mean or modal response in the training data
  - What-if or scenario-analysis can also be performed by setting specific inputs to missing data
    - Sensitivity analysis

Caveats and Solutions

- Proper binning of predictors can ensure simpler relationships to risk
  - As is common when building in standard data preparation for scorecards



Practical Considerations

- The methods discussed here are easily implemented as part of an automated and/or real-time credit scoring system
- All that is required is a computer system that can efficiently implement a prediction model for predicting credit default risk (fraud probability, etc.)
- And perform the necessary what-if analyses based on metadata stored along with each predictor regarding how much each will have to be changed during the what-if analyses to derive reason scores.
- If the predictive model is built based on appropriately binned and Weight-of-Evidence recoded predictor variables it is easy to implement an algorithm to compute reason scores for any model.

Summary

- This presentation discussed briefly the difference between traditional statistical modeling and general predictive modeling methods based on general pattern recognition “learning” algorithms
- The latter approach is now wide used in various domains where accurate predictions of future outcomes based on historical data have significant business value
  - Marketing and sales
  - Fraud detection
  - Product quality from manufacturing data
  - …
- General approximators, that can represent accurately *any* repeated pattern in the historical data, and leverage those patterns for more accurate predictions.
- The reasons for a specific prediction can be derived by applying what-if scenario analysis to derive partial derivatives of risk with respect to each predictor at the point of the observed values (or the average/modal response)
- Thus, with only a little more complexity at the point of scoring new data points more accurate prediction models can be used to extend more credit without increasing the credit default rate.

Thank You