ENTERPRISE MINER:
4 – CREDIT SCORECARD DEVELOPMENT

DR IAIN BROWN, ANALYTICS & INNOVATION PRACTICE, SAS UK

15 NOVEMBER, 2016

AGENDA

- **SAS – 15th November 2016 at 10:00am**
- **Enterprise Miner: Credit Scorecard Development**
- *The session looks at:*
  - *Variable Classing and Selection*
  - *Scorecard Modelling*
  - *Reject Inference*
  - *Scorecard Assessment*

ROLE OF SAS ENTERPRISE MINER

THE ANALYTICS LIFECYCLE

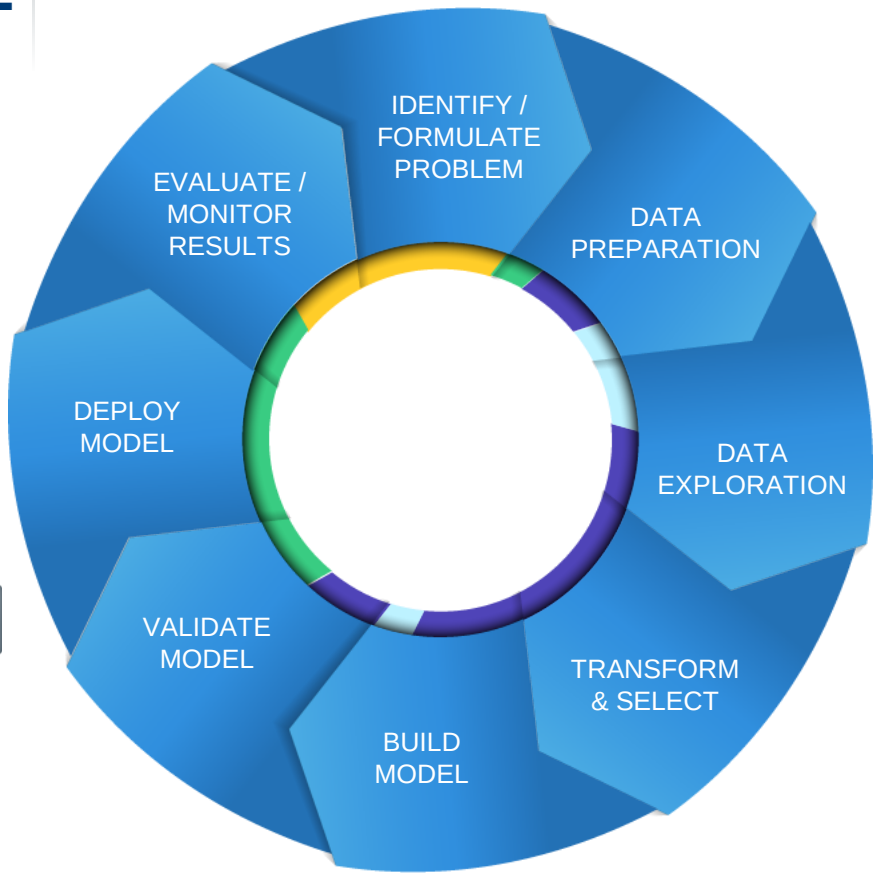
PREDICTIVE ANALYTICS AND DATA MINING

BUSINESS MANAGER

Domain Expert
Makes Decisions
Evaluates Processes and ROI

BUSINESS ANALYST

Data Exploration
Data Visualization
Report Creation



IT SYSTEMS / MANAGEMENT

Model Validation
Model Deployment
Model Monitoring
Data Preparation

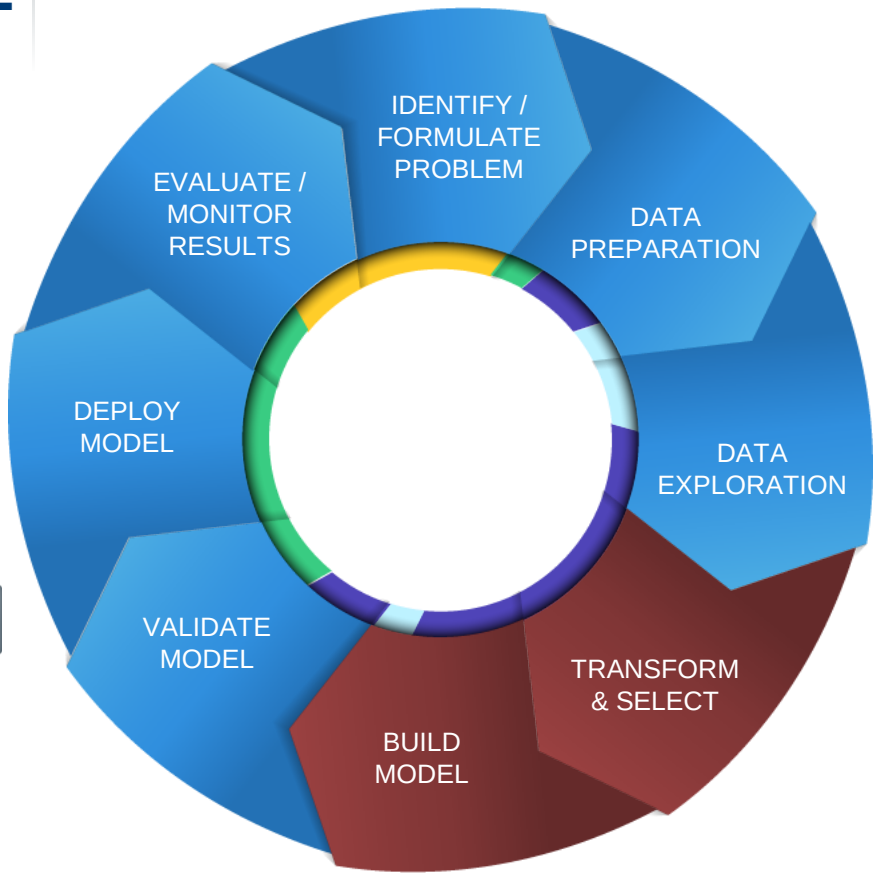
DATA MINER / STATISTICIAN

Exploratory Analysis
Descriptive Segmentation
Predictive Modeling

THE POWER TO KNOW.

THE ANALYTICS LIFECYCLE

ANALYTICAL MODEL DEVELOPMENT



BUSINESS MANAGER

Domain Expert
Makes Decisions
Evaluates Processes and ROI

BUSINESS ANALYST

Data Exploration
Data Visualization
Report Creation

IT SYSTEMS / MANAGEMENT

Model Validation
Model Deployment
Model Monitoring
Data Preparation

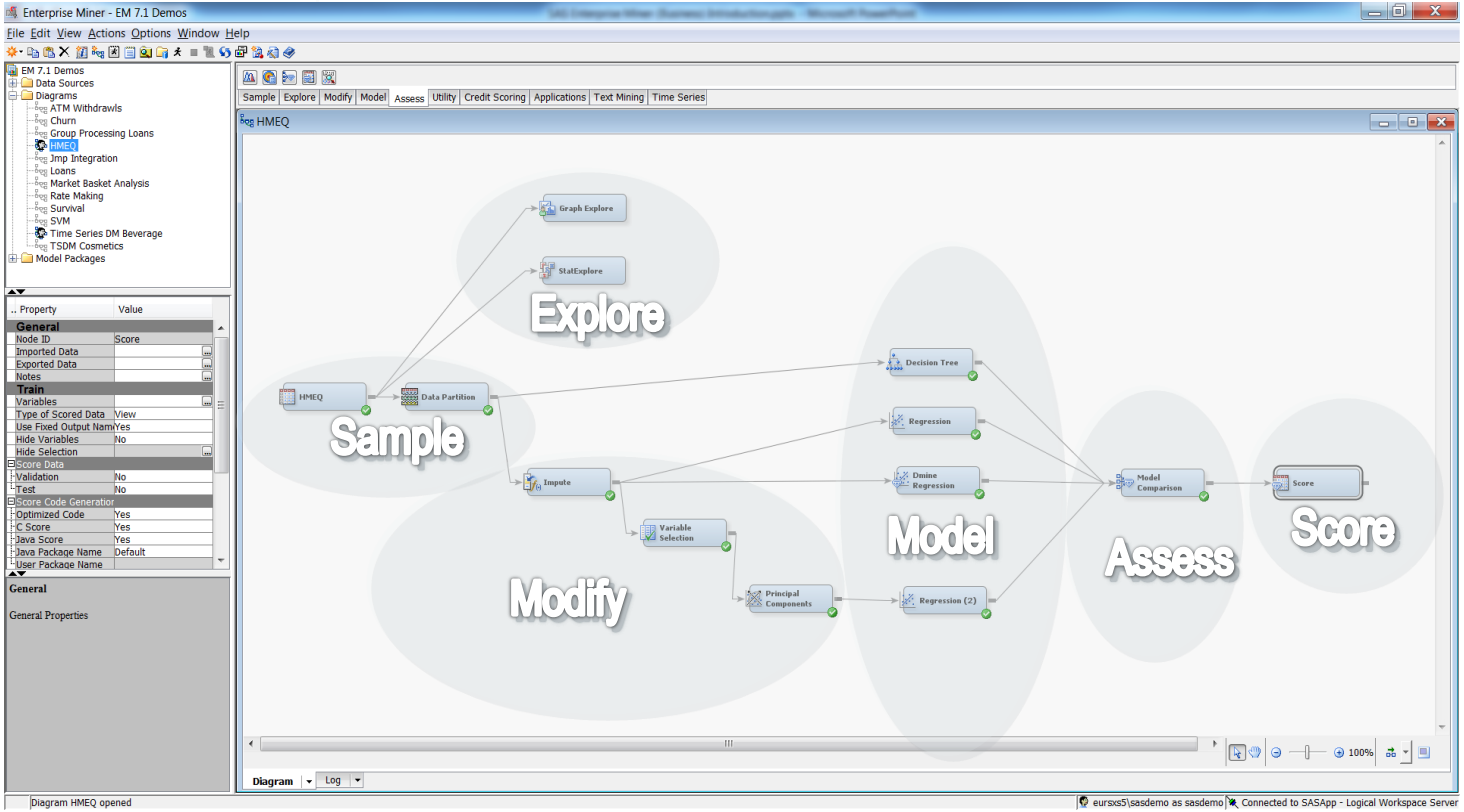
DATA MINER / STATISTICIAN

Exploratory Analysis
Descriptive Segmentation
Predictive Modeling

THE POWER TO KNOW.

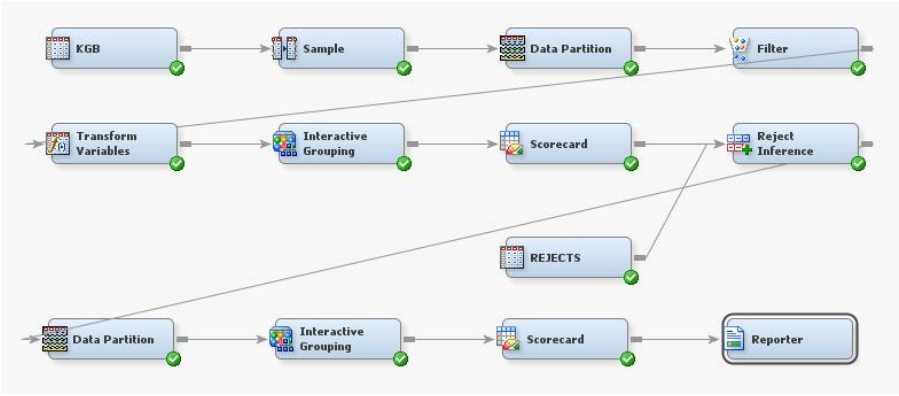
SAS® ENTERPRISE MINER™

SEMMA IN ACTION – REPEATABLE PROCESS



ADD-IN: SAS® CREDIT SCORING FOR ENTERPRISE MINER™

- Build Scorecards to assign risk scores to customers:
  - Application Scoring
  - Behavior Scoring
  - Probability of Default Scoring (rating)
  - Collection scoring
- Lot's of control over the scorecard construction. Not a black box.
- Broad set of tools to support the complete data mining process
- Easy-to-use GUI to develop more –and better – scorecards
- Easy to distribute and scalable system
- An open, extensible design for ultimate flexibility
  - GUI + SAS Code Node + Extension Nodes
  - Batch processing
- Extensive model scoring alternatives
  - SAS, Java, PMML, Scoring Accelerator for Teradata



CREDIT SCORING

CREDIT SCORING | OVERVIEW

- Application scoring: Likelihood that an applicant will not repay a loan and will therefore fall into default
  - Applicant characteristics (for example, age, income, employment status, time at address, …)
  - Credit bureau information
  - Application information of other applicants
  - Repayment behaviour of other applicants
- Develop models (scorecards) estimating the probability of default of a customer
- Typically, assign points to each piece of information, add all points, and compare with a threshold (cutoff)

CREDIT SCORING | OVERVIEW

- Behavioural scoring: Through the cycle risk of default based on an existing customer's behaviour
- Update risk assessment taking into account recent behaviour
- Example of behaviour:
  - Average/Max/Min/trend in checking account balance, bureau score, …
  - Delinquency history (payment arrears, …)
  - Job changes, home address changes, …
- Dynamic

CREDIT SCORECARD DEVELOPMENT PROCESS

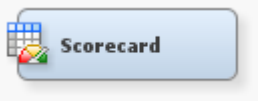
1. Variable Classing and Selection
   - Calculation of Weight of Evidence statistics
   - Calculation of Information Value or Gini statistics
   - Automatic and Interactive
     - Enhances productivity
     - Incorporate business knowledge
2. Scorecard construction
   - Fitting a logistic regression model
   - Scaling and calculating score points
3. Assessing Scorecard quality
   - KS, GINI, ROC and Trade-off charts
4. Reject Inference
   - Model for scoring through-the-door population

SPECIFIC CREDIT SCORECARD CAPABILITIES

EM Credit Scoring nodes:

- **Automatic and interactive variable grouping**
  - Computes Weights of Evidence
  - GINI and Information Values for variable selection

- **Scorecard construction**
  - Logistic regression based using WOE or dummy variables as inputs
  - Parameterized score points scaling
  - Assessment statistics and charts



- **Reject inference**
  - Through the door impact analysis

EXAMPLE CREDIT SCORECARD

Let cutoff = 500 points

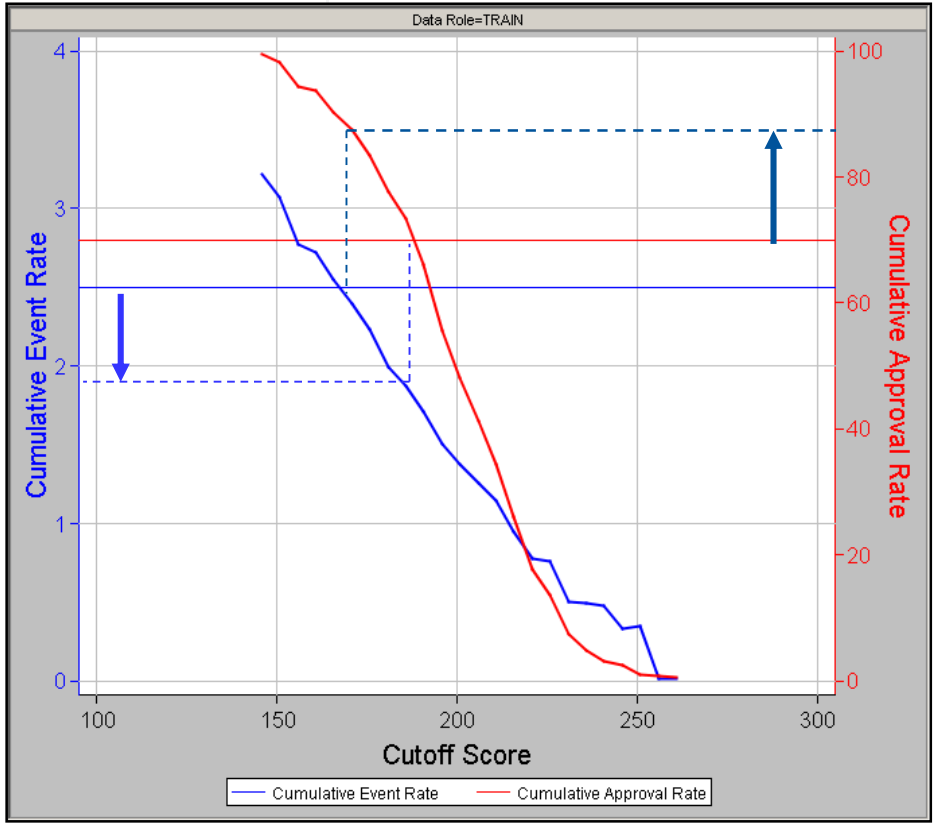
A new customer applies for credit.

AGE	32	120 points
HOUSE	OWN	225 points
INCOME	$45K	200 points
Total		545 points

ACCEPT FOR CREDIT

Characteristic Name	Attribute	Scorecard Points
AGE	Up to 25	100
AGE	26-33	120
AGE	34 - 45	185
AGE	45+	225
HOUSE	OWN	225
HOUSE	RENT	110
INCOME	Up to 10k	120
INCOME	10k-25k	140
INCOME	26k-35k	160
INCOME	36k-50k	200
INCOME	50k+	240

CREDIT SCORING IMPROVEMENTS



At constant approval rate:

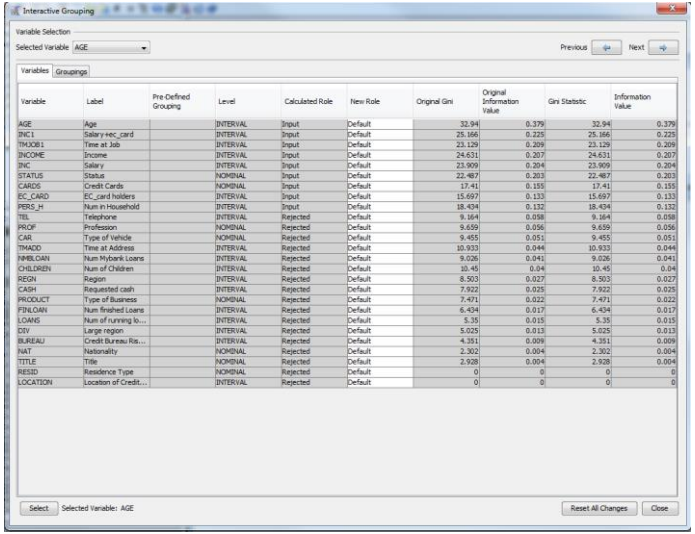
Reduce bad rate
from 2.5% to 1.9%

At constant bad rate:

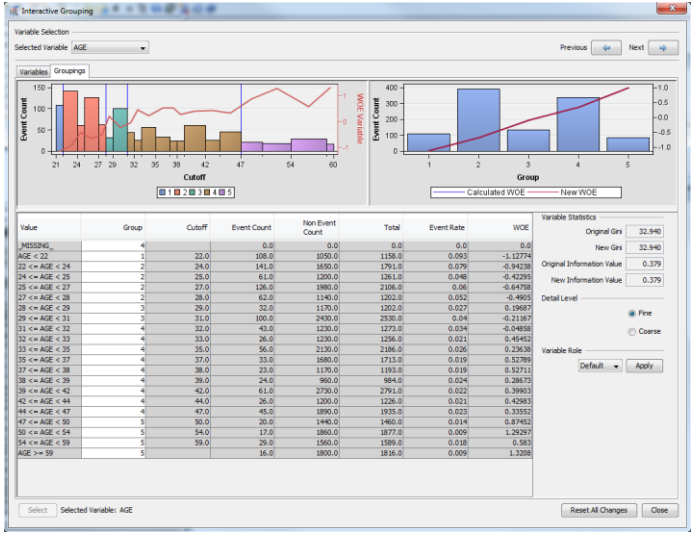
Increase approval rate
from 70% to 87%

CREDIT SCORING | INTERACTIVE GROUPING NODE EXAMPLE RESULTS

- Automatic and interactive variable grouping
  - Computes Weights of Evidence



Auto-updating IV and Gini



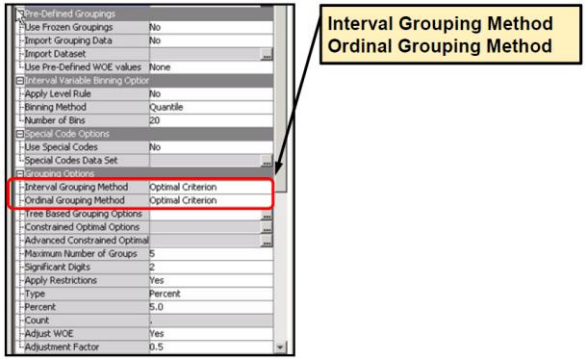
Fine/Coarse Detail

Metrics for Characteristic Analysis

- Measures commonly used in a characteristic analysis
  - Weights of Evidence (WOE) – measures the strength of the attribute of a characteristic in separating good and bad accounts
  - Information Value (IV) – measures the overall predictive power of the characteristic; that is, its ability to separate good and bad loans. (weighted sum of WOEs)
  - Gini Statistic – alternative approach to IV to assess the overall predictive power of a characteristic.

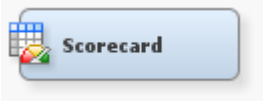
Grouping options

- The Interval and Ordinal Grouping Method fields determine how the grouping algorithm groups the pre-binned interval and ordinal variables:
  - **Optimal Criterion** – groups the variables based on the criterion property
  - **Quantile** – enables the algorithm to generate groups with approximately the same frequency in each group.
  - **Monotonic Event Rate** – generates groups that result in a monotonic distribution of event rates across all levels
  - **Constrained optimal** – generates groups based on pre-defined constraints



CREDIT SCORING | SCORECARD = LOGISTIC REGRESSION + SCALING

- In order to determine the relative contribution of each characteristic logistic regression is used.
- One regression co-efficient per characteristic is produced if Weight Of Evidence coding is used
- Dummy variable (1-of-n coding) would instead lead to one co-efficient per attribute.



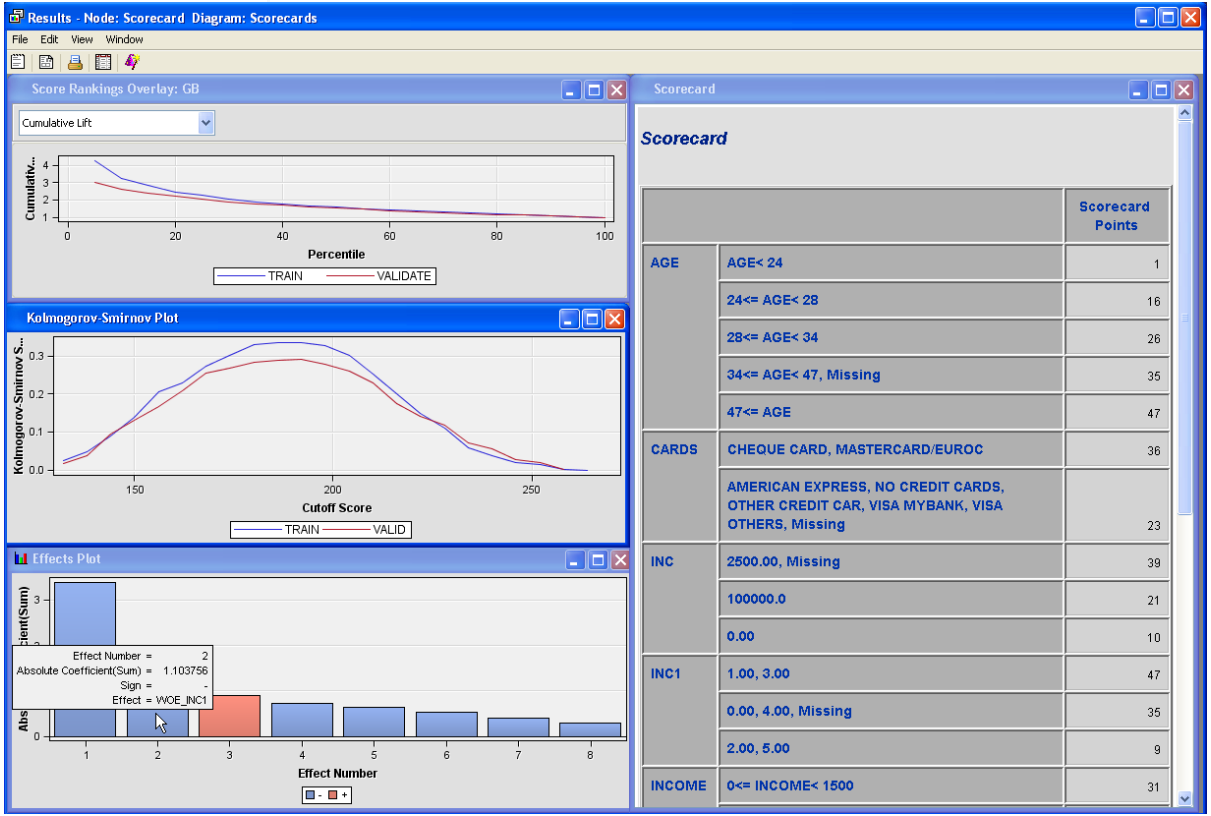
SCORE POINTS

- An attribute's points reflect:
  - Risk of an attribute relative to other attributes of the same characteristic; determined by an attributes WOE, Weight of Evidence, value
  - Relative contribution of each characteristic to the overall score; determined by a characteristic's co-efficient in a logistic regression model

SCALING

- In order to obtain more ‘friendly’ scores, the points per attribute are linearly scaled with a factor and an offset.
- The score is proportional to the good/bad odds and not the bad/good odds that are modeled in the logistic regression, thus a negative sign is introduced.
- Smaller scores thus correspond to higher risk.

CREDIT SCORING | SCORECARD NODE SAMPLE RESULTS

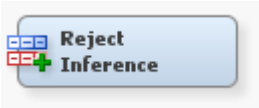


Scorecard

		Scorecard Points
AGE	AGE< 24	1
	24<= AGE< 28	16
	28<= AGE< 34	26
	34<= AGE< 47, Missing	35
	47<= AGE	47
CARDS	CHEQUE CARD, MASTERCARD/EUROC	36
	AMERICAN EXPRESS, NO CREDIT CARDS, OTHER CREDIT CAR, VISA MYBANK, VISA OTHERS, Missing	23
INC	2500.00, Missing	39
	100000.0	21
	0.00	10
INC1	1.00, 3.00	47
	0.00, 4.00, Missing	35
	2.00, 5.00	9
INCOME	0<= INCOME< 1500	31

REJECT INFERENCE

- Good / bad Information is only available for past accepts, but not for past rejects.
- In order to arrive at models for the through-the-door population, it is common practice to perform ‘Reject Inference’: apply the scorecard to the rejects and classify rejects as inferred good / inferred bad.
- An ‘augmented’ input data set then is created by adding the inferred good / bad to the actual good / bad. The scorecard is re-adjusted using this data set.



REJECT INFERENCE – TECHNIQUES

- The reject inference node enables three types of augmentation technique to be applied:
  - Hard Cutoff
  - Parcelling
  - Fuzzy

REJECT INFERENCE – TECHNIQUES

Hard Cutoff Augmentation

1. Build a scorecard model using the known good/bad population (that is, accepted applicants).
2. Score the rejected applicants with this model to obtain each rejected applicant's probability of default and their score on the scorecard model.
3. Create weighted cases for the rejected applicants.
4. Set a cutoff score level above which an applicant is deemed ***good***. All applicants below this level are deemed ***bad***.
5. Add the inferred goods and bads back in with the KGB's and rebuild the scorecard.

REJECT INFERENCE – TECHNIQUES

Parcelling Augmentation

1. Build a scorecard model using the known good/bad population.
2. Score the rejected applicants with this model to obtain each rejected applicant's probability of default.
3. Create weighted cases for the rejected applicants.
4. The inferred good/bad status of the rejected applicants will be assigned randomly and proportional to the number of goods and bads in the accepted population ***within each score range***.
5. If desired, apply the event rate increase factor to P(bad) to increase the proportion of bads among the rejects (rule of thumb 2-8 times that of the accepted applicants)
6. Add the inferred goods back in with the known goods and bads and rebuild the scorecard.

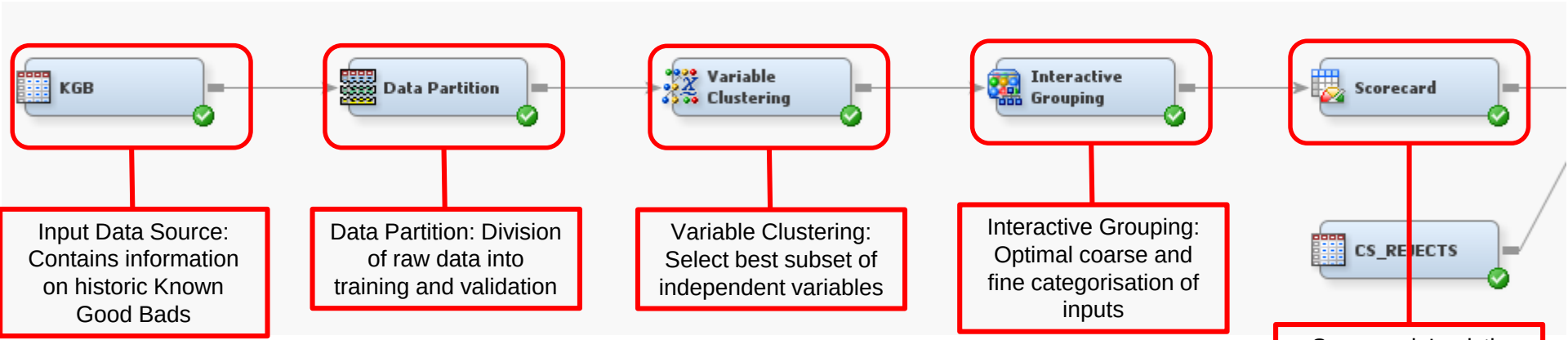
Score	# Bad	# Good	% Bad	% Good	Reject	Rej - Bad	Rej - Good
0-169	290	971	23.0%	77.0%	1646	379	1,267
170-179	530	2,414	18.0%	82.0%	1732	312	1,420
180-189	365	2,242	14.0%	86.0%	3719	521	3,198
190-199	131	1,179	10.0%	90.0%	7334	733	6,601
200-209	211	2,427	8.0%	92.0%	1176	94	1,082
210-219	213	4,047	5.0%	95.0%	3518	176	3,342
220-229	122	2,928	4.0%	96.0%	7211	288	6,923
230-239	139	6,811	2.0%	98.0%	3871	77	3,794
240-249	88	10,912	0.8%	99.2%	4773	38	4,735
250+	94	18,706	0.5%	99.5%	8982	45	8,937

REJECT INFERENCE – TECHNIQUES

Fuzzy Augmentation

1. Build a scorecard model using the known good/bad population and score the rejected applicants with this model to obtain the probability of the rejected applicant being a good (P(good)) and the probability of being a bad (P(bad)).
2. Do not assign a reject to a good/bad class. Instead, create two weighted cases for each rejected applicant using P(good) and P(bad).
3. Multiply P(good) and P(bad) by the user-specified rejection rate to form frequency variables.
4. Results are in two observations for the rejected applicants. One observation has a frequency variables (rejection weight *P(good)) and a target variable value of 0. The other observation has a frequency variable (rejection weight*P(bad)) and a target variable value of 1.

CREDIT SCORING | EM – APPLICATION SCORECARD DEVELOPMENT



Input Data Source: Contains information on historic Known Good Bads

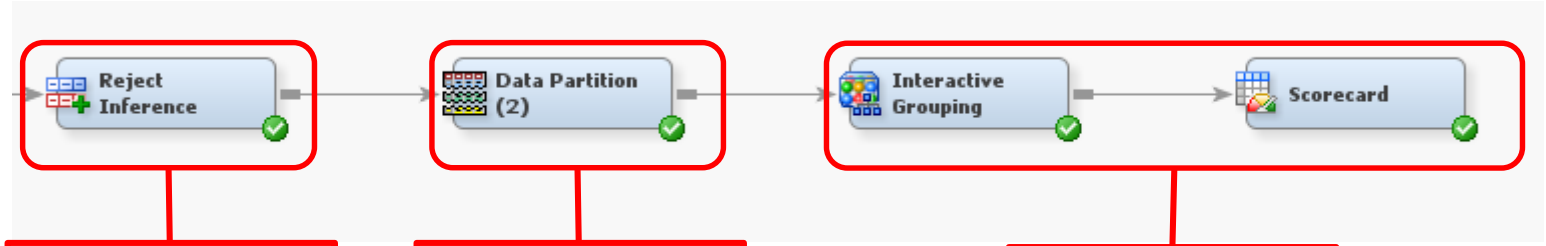
Data Partition: Division of raw data into training and validation

Variable Clustering: Select best subset of independent variables

Interactive Grouping: Optimal coarse and fine categorisation of inputs

Scorecard: Logistic Regression model to predict a customer's probability of default

CREDIT SCORING | EM – APPLICATION SCORECARD DEVELOPMENT



Reject Inference: Inferred rejected population outcome

Data Partition: Division of augmented sample into training and validation

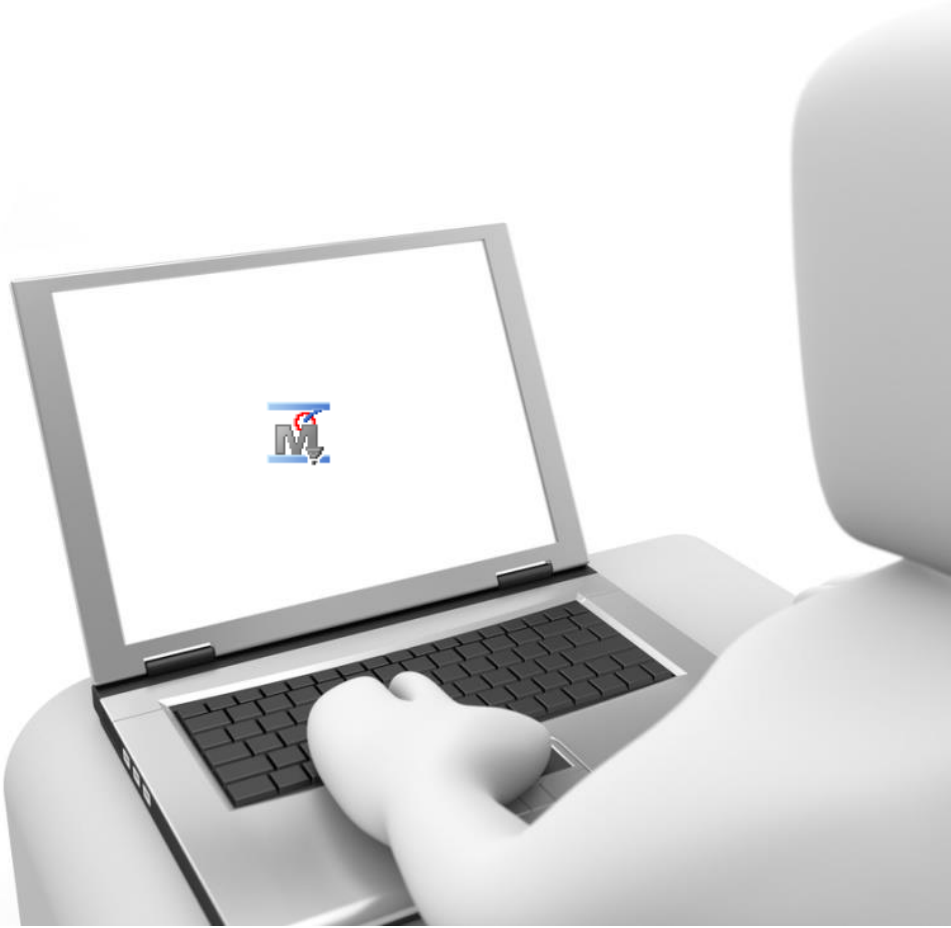
Final augmented Good Bad model

SCORECARD DEPLOYMENT

- Model Deployment
  - Application of model in business processes for delivery of extracted knowledge in the right time to the right people
  - Realises model value by providing decision support
  - Integration requires scoring code in appropriate language
- Score code
  - includes all data transformation of the scorecard process
  - produced in SAS, C, Java and in-database (Teradata)
  - Enables batch and real-time scoring in SAS, databases and on the web
- Metadata
  - Integration into SAS Business Analytics Framework

CREDIT SCORING | DEMONSTRATION

- EM – Application Scorecard



MODEL VALIDATION AND REPORTING

MODEL MONITORING

- Three types of model quality criteria are monitored
  - Model stability
  - Model performance
  - Model calibration
- Multiple statistical indicators suggested under Basel Committee Working Paper
- Additionally indicators for monitoring LGD and CCF models

MODEL STABILITY

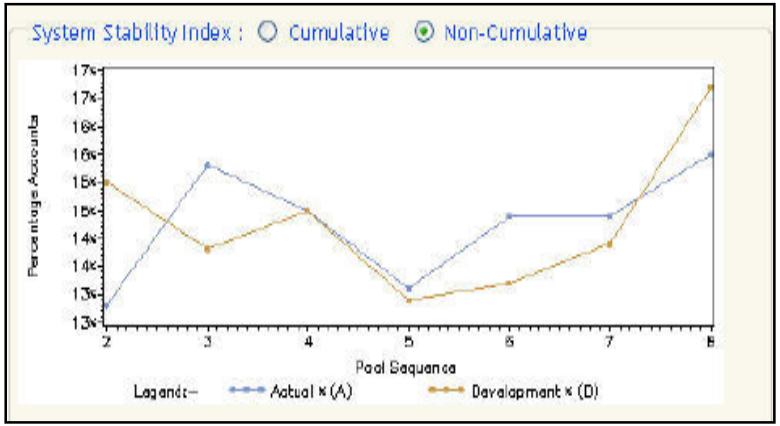
- Important for detecting population shifts, for example for pre-deployment sanity check
  - Is the distribution of scores still similar to when you developed the model?
  - Have you now more high (low) scoring customers then previously?

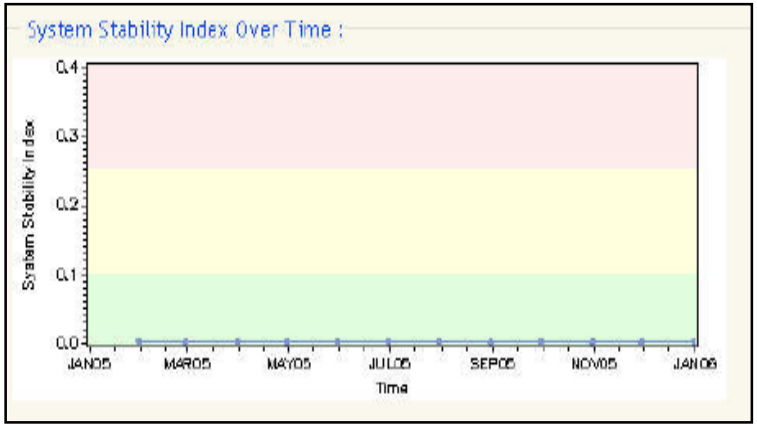
STABILITY – SYSTEM STABILITY INDEX

- A measure used to monitor the score distribution:

SSI = Sum ((A-D)*Log (A/D))

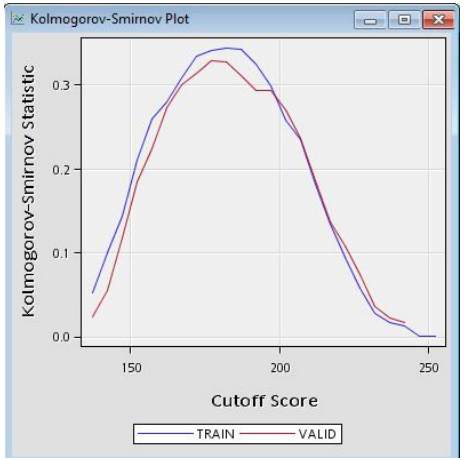
where A and D are the No. of Accounts in the Actual and Development sample.

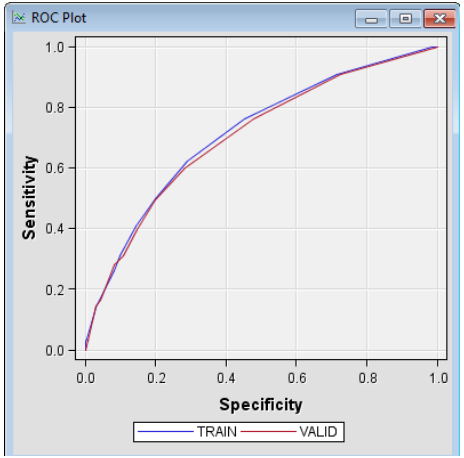




MODEL PERFORMANCE

- Important for ensuring high quality pooling and approval decisions
  - Do the bad customers have low scores?
  - Do the good customers have high scores?
  - To what degree do the score distributions of good and bad customers overlap?
  - How well does the model separate the good from the bad customers?
  - How many of all bad customers can you find within the low scoring customers?



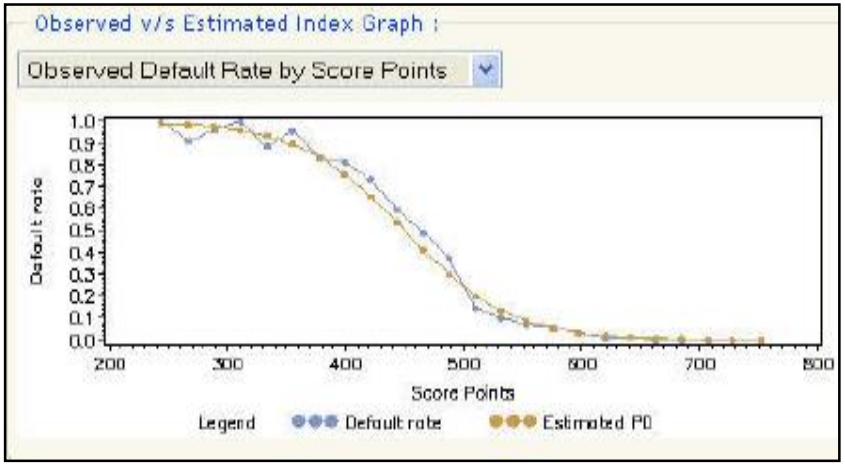


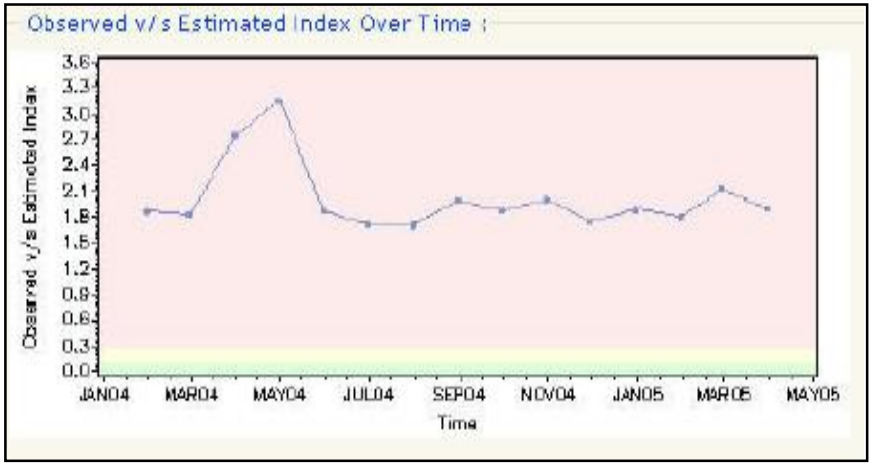
MODEL CALIBRATION

- Validation of the calibration is more complex than the validation of the discriminatory power of internal rating systems
- To transform a PD from a score, direct or indirect methods can be used
- Important for ensuring correct risk assessment, for example Risk Weighted Assets calculation
  - How similar are predicted default rates to actual default rates?
  - Are the differences significant?

CALIBRATION – OBSERVED VS. ESTIMATED INDEX

- Measures the “closeness” of observed vs. estimated default rates





SAS® MODEL MANAGER

REGULATORY BASEL 2 REPORTING

Monitor Model Performance

REPORTING FOR FS REGULATORS BASED ON BASEL 2 REQUIREMENTS

- Probability –of-Default (PD) and Loss-Given-Default (LGD) model performance monitoring
- Calculation of 30 model performance statistics
  - Model Stability
  - Model Performance
  - Model Calibration
- Customizable traffic lighting
- Trend time series charts

Model Validation Report for Probability of Default (PD)
Project: HMEQ PD Scorecard Version: 2012

Model Validation Summary

Test Level	Test Name	Development Score	Current Score	Current and Development Scores Compared with Benchmark
0: Stability	System Stability Index	<.0001	0,5237	
1: Performance	Accuracy	0,8601	0,7499	
	Accuracy Ratio (Gini)	0,7547	0,3762	
	Area Under the Curve (AUC)	0,8773	0,6881	
	Bayesian Error Rate	0,1339	0,2462	
	D Statistic	1,8543	0,8102	
	Error Rate	0,1399	0,2501	
	Information Statistic	2,5454	0,6133	
	Kendall's Tau-b (p-value)	<.0001	<.0001	
	Kullback-Leibler Statistic	1,3272	0,3505	
	Kolmogorov-Smirnov Statistic	0,6205	0,3215	
	(1-PH) Statistic	0,9185	0,6251	
	Pietra Index	0,2194	0,1137	
	Precision	0,6459	0,5640	

Model Validation Report for Loss Given Default (LGD)
Project: LGD Models Version: 2012

Model Validation Summary

Test Level	Test Name	Development Score	Current Score	Current and Development Scores Compared with Benchmark
0: Stability	System Stability Index	<.0001	0,0271	
1: Performance	Accuracy Ratio (Gini)	0,2885	0,4254	
2: Calibration	Correlation Analysis	0,8585	0,8152	
	Mean Absolute Deviation (MAD)	0,0595	0,0756	
	Mean Absolute Percent Error (MAPE)	66,4713	148,9093	
	Mean Squared Error (MSE)	0,0094	0,0133	

THE POWER TO KNOW.

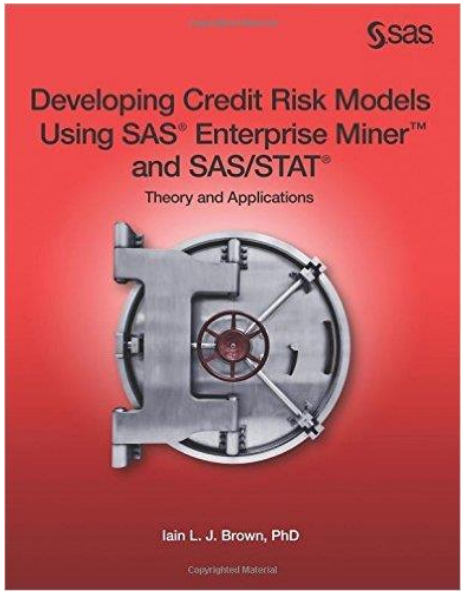
SUMMARY

SUMMARY

HOW TO LEVERAGE SAS ENTERPRISE MINER FOR CREDIT RISK MODELLING

- Make accurate and timely risk calculations to improve
  - Approval processes
  - Risk management
  - Basel compliancy
  - Collection management

- Provides
  - In–house model development
    - A-IRB models
  - Flexible model deployment
  - Out-of-the-box reporting



QUESTIONS AND ANSWERS

IAIN.BROWN@SAS.COM

THE POWER TO KNOW.