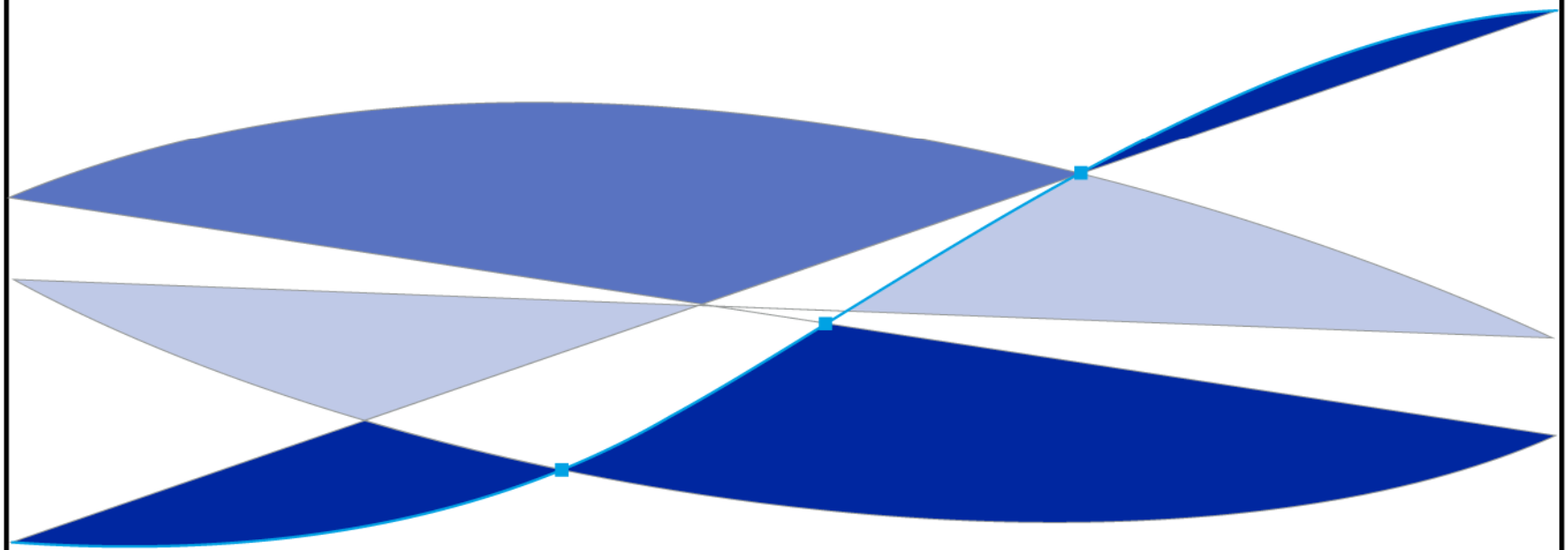
Moody's | K·M·V

Solving the Data Challenge for Developing and Validating IRB Systems



Jack Gregory

Managing Director, Prof. Services, Moody's KMV

jack.gregory@mkmv.com

Agenda

1. Introduction
2. Data Challenges Faced by Organizations
3. Best Practices in Data Management
4. Case Studies
5. Conclusion
6. Q&A

1 Introduction

The Experience of Moody's KMV in Solving Data Challenges

History:

- Over 20 years of preparing data for credit risk modeling
- Over 10 years of procuring global credit default and loss data of public and private firms for Moody's KMV databases used to develop Moody's KMV Credit Edge and RiskCalc
- Over 3 years of providing design, validation and calibration for internal client models to support Basel initiatives

Key Expertise:

- Development of rules and checks to create optimal data sample
- Local knowledge of accounting and credit principles in many markets
- Proprietary application for assessing data quality and data processing (CRD Analyst Toolkit)
- Standardization of loan performance metrics across wide variety of participant systems

Moody's KMV's Wide Range of Data Production Systems

MKMV'S CURRENT PRODUCTION PROCESSES		
Monthly Processes (All Output is Monthly)		
System Name	**Primary Output**	**Down Stream Product**
Public	Monthly EDFs, Drivers, Industry Breakdown	CreditMonitor, Portfolio Manager, Deal Analyzer
Private	Private EDFs and Drivers	Private Firm Model
CDO	Model collateral for individual CDOs	CDOEdge
LossCalc	LGD model parameters	LossCalc
Daily Processes (All Output is Daily)		
System Name	**Primary Output**	**Down Stream Product**
Public	Daily EDFs and Groups Data	CreditEdge, CreditEdge Plus
EIS	EDF Implied Spreads and Fair Value Spreads	CreditEdge Plus & CreditMark
TPM/CMK	Market Sharpe Ratio, LGDs	CreditMark
Note: CRD Production is run autonomously by the CRD team in NYC		

MKMV'S PRODUCTION PROCESSES IN THE NEAR FUTURE		
Monthly Processes (All Output is Monthly)		
System Name	**Primary Output**	**Down Stream Product**
RiskCalc	ddFactor	RiskCalc v3.1

For Private Middle-Market Credits, the Credit Research Database is a Source of Data for Model Development & Testing

CRD Private Firm Statistics (as of June 2005)

	Number of Obligors	Defaulted Obligors	Number of Fin Stmts
North America	106,052	10,496	536,515
Europe	1,711,990	141,178	9,043,417
Asia	263,083	15,609	1,219,300
Australia	31,576	2,715	102,778
Africa	19,633	415	52,613
Global CRD	**2,132,334**	**170,413**	**10,954,623**

- Data from over 20 countries covering over 80% of the world's GDP
  - Most major industrialized nations (e.g. US, Canada, Germany, France, UK, Italy, Japan, Australia, Spain, Nordic Region, South Korea)
  - Targeted emerging markets (e.g. Mexico, South Africa)
- Over 40 financial institutions have contributed data, with several supplementary data vendor relationships

Pre-Basel II Data Practices

- Credit data spread across "islands" of systems (automated and paper-based)
- Data for defaults and facility data stored on paper and non-automated
- Limited history archiving of detailed credit data
- Inputs to ratings very rarely stored…sometimes outputs
- Data often extracted as "one-off" projects, typically from enterprise risk systems with limited value for validation and calibration efforts

Benefits of Standardizing Credit Administration and Internal Ratings Process in a Controlled Data Environment

Controlled data environment → **Standardized** risk measurement → **Enhanced** risk management

Basel II and the Growing Importance of Managing Data Quality

"Institutions using the IRB approach for regulatory capital purposes will need advanced data management practices to produce credible and reliable risk estimates."

(US Draft Supervisory Guidance on IRB Systems for Corporate Credit, 2003)

Data Requirements for IRB System

The data must be of sufficient depth, scope and reliability to:

- Validate IRB system processes
- Validate parameters (PD, LGD)
- Refine the IRB system
- Develop internal parameter estimates
- Apply improvements historically
- Calculate capital ratios
- Produce internal and public reports
- Support risk management

Credit Data Management – What We Have Learned So Far?

- Production data from credit systems, including Moody's KMV, not sufficient alone to meet data requirements
- Structured, formal, systematic and documented procedures, processes and verification for data are necessary
- Data management requires collaboration and effort from Risk Management, Credit Operations, Information Technology and Vendors
- Aligning data along time dimension is challenging and industry codes important and challenging
- There is never enough structured credit data (yet)

2

Data Challenges Faced by Organizations

Issues Addressed in This Section

Data challenges

#1: Availability

#2: Quality

#3: Consistency

#4: Low or non-default industries

#5: Qualitative data

#1: Data Availability – Significance for Basel Compliance

Minimum Data Observation Period

PD Estimates: At least 5 years of data

LGD Estimates: At least 1 complete economic cycle and no shorter than 7 years

EAD Estimates: At least 1 complete economic cycle and no shorter than 7 years

Time Horizon

Time horizon for PD estimation is 1 year; but banks are expected to use a longer time horizon in assigning ratings

Alignment of Data from Different Sources by Time

Linkages between data from financial statement/loan accounting systems

Controlled data environment required

What happens if banks have limited data?

General principle: The less data a bank has, the more conservative it must be in its estimation

Affects ability to develop better models and test models for performance across different dimensions

#1: Data Availability – Significance for Accurate Modeling

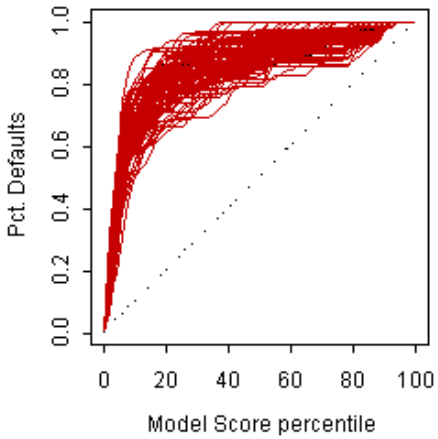
The number of defaults in the dataset drives the accuracy of your test*

Research:

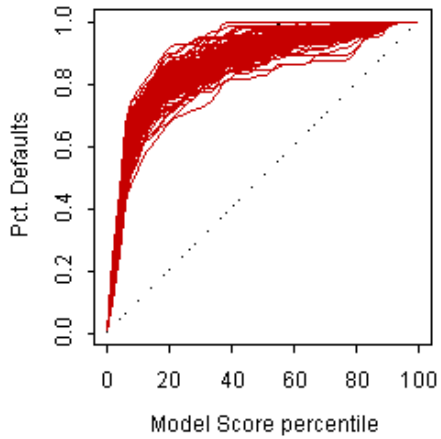
- A sample of 20,000 non-defaulters with a varying number of defaulters was analyzed 100 times using a bootstrapping technique
- As more defaults are added, the accuracy of the test is greatly improved
- We can see, however, that even with 400 defaults, we can still do better
- Increasing the number of non-defaulters, on the other hand, has less of an impact

* Stein, Roger M., 2002, Benchmarking Default Prediction Models: Pitfalls and Remedies in Model Validation, (Moody's KMV, New York).

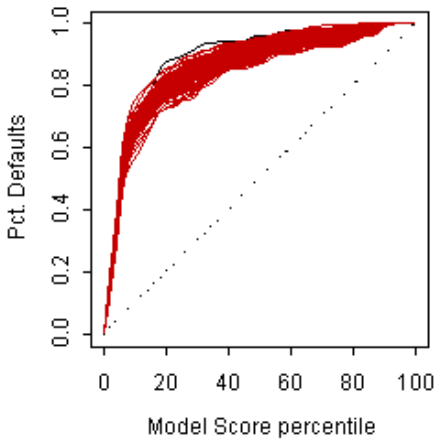
100 data sets: Ngood=20,000, Ndef=50



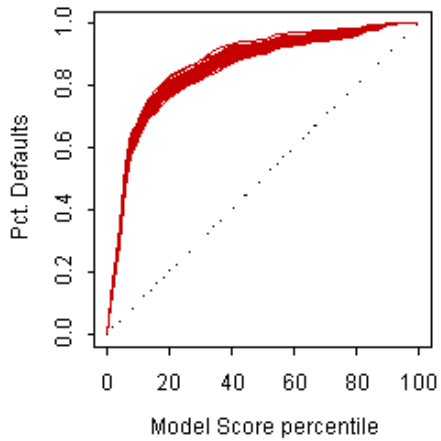
100 data sets:Ngood=20,000, Ndef=100



100 data sets:Ngood=20,000, Ndef=200

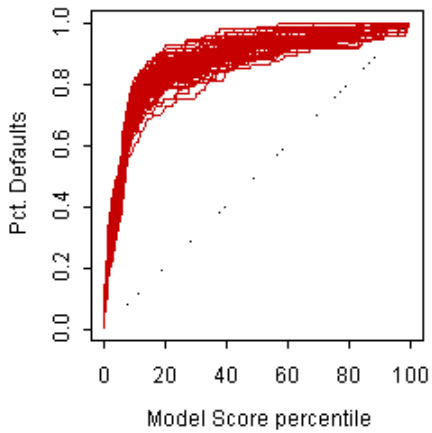


100 data sets:Ngood=20,000, Ndef=400

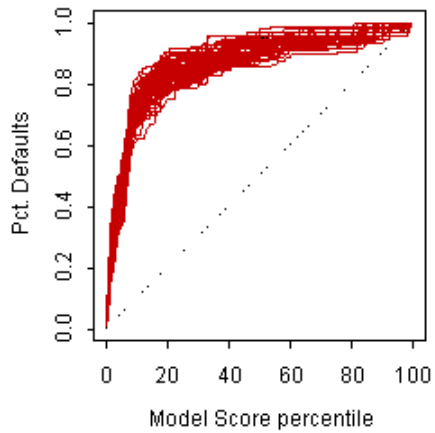


Stein, Roger M., 2002, Benchmarking Default Prediction Models: Pitfalls and Remedies in Model Validation, (Moody's KMV, New York).

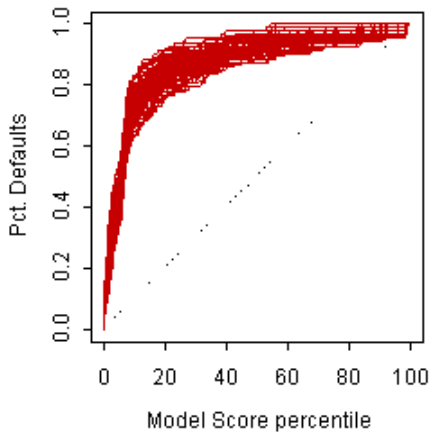
100 data sets: Ngood=10,000, Ndef=100



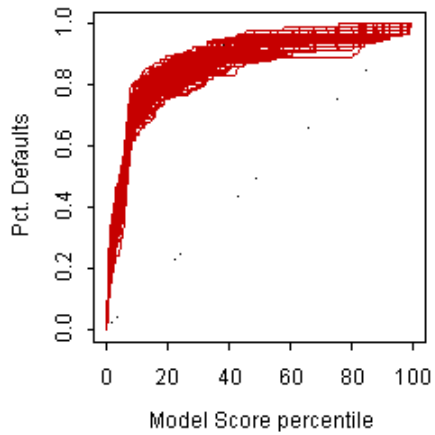
100 data sets: Ngood=15,000, Ndef=100



100 data sets: Ngood=20,000, Ndef=100

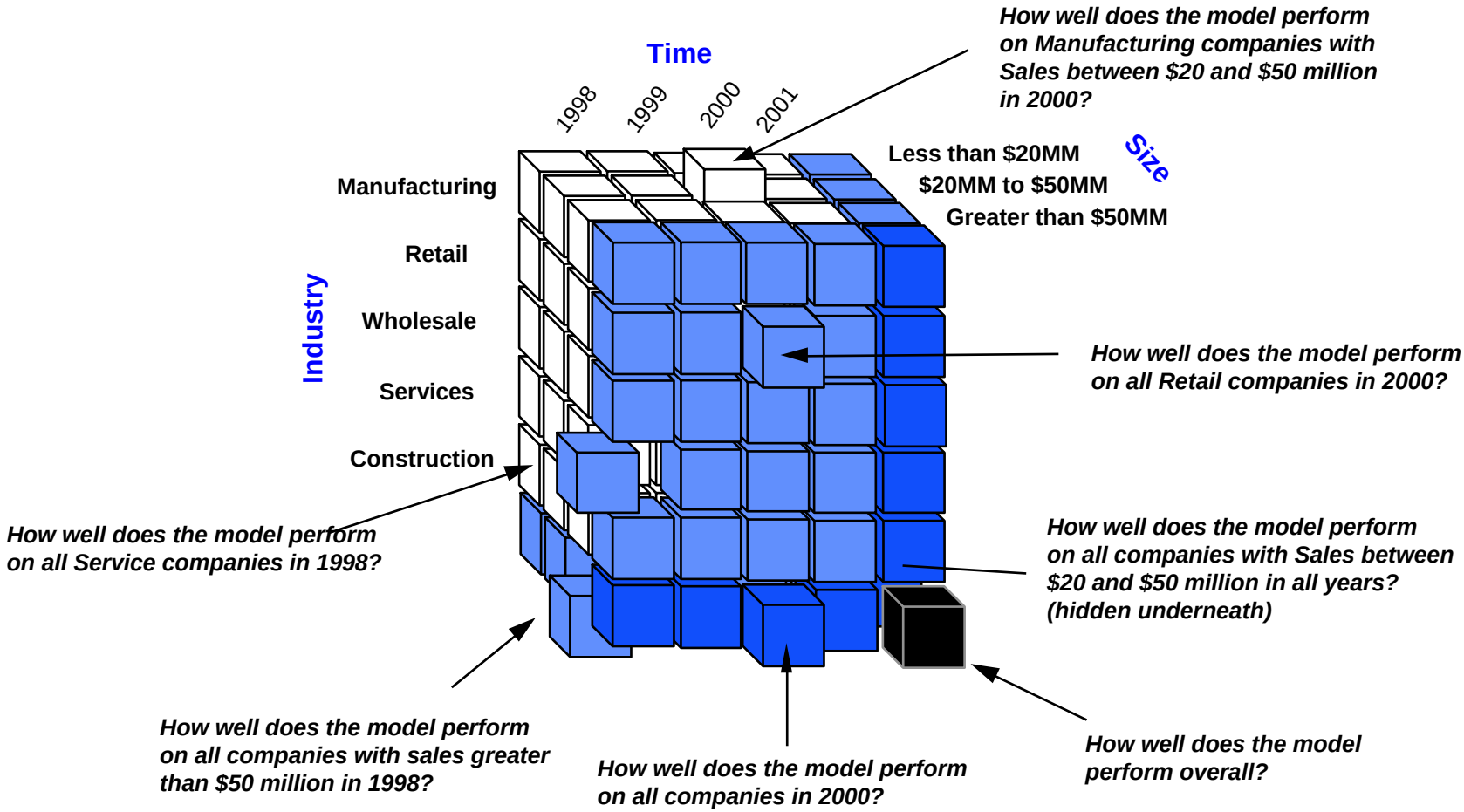


100 data sets: Ngood=50,000, Ndef=100



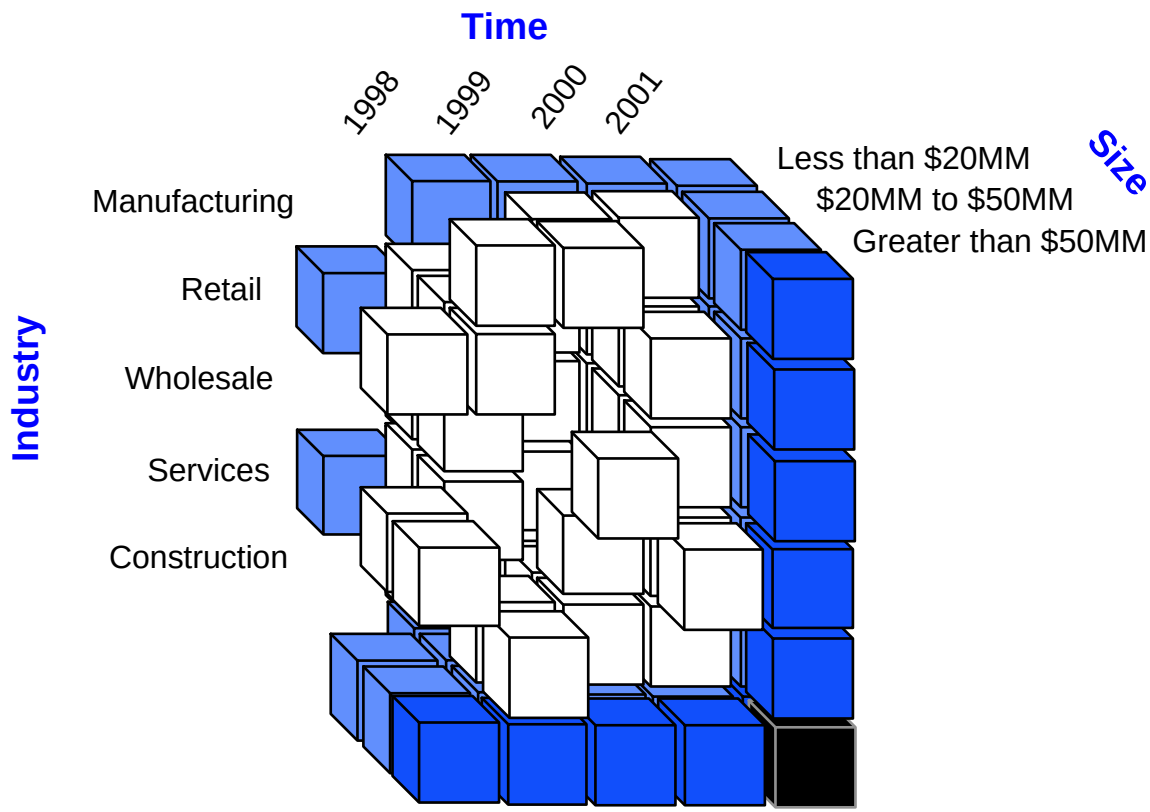
Stein, Roger M., 2002, Benchmarking Default Prediction Models: Pitfalls and Remedies in Model Validation, (Moody's KMV, New York).

#1: Data Availability – Analyzing How a Model Performs Across Industry, Size and Time Periods



From Dhar, B. and Stein R. (1997), *Seven Methods for Transforming Corporate Data into Business Intelligence*, Prentice Hall, NJ.

Most Banks Only Have Enough Data to Fill Part of the Picture They Need to Answer These Questions…



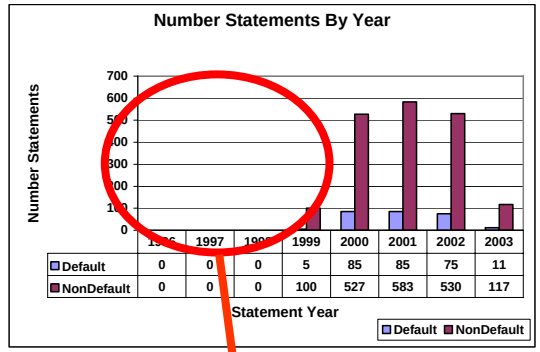
From Dhar, B. and Stein R. (1997), *Seven Methods for Transforming Corporate Data into Business Intelligence*, Prentice Hall, NJ.

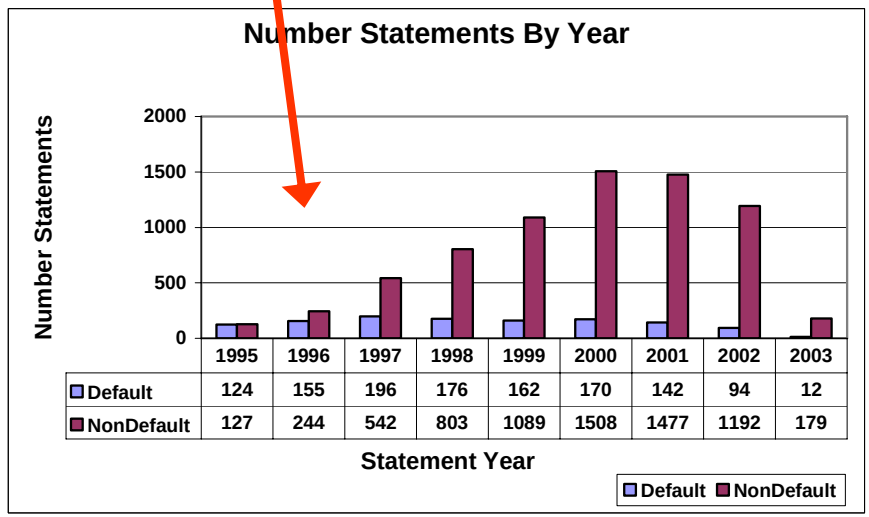
Augmentation of Dataset Using External Data

Under Basel II guidelines, a bank can use external data (public data/pooled industry data) as long as it demonstrates that its PD/LGD estimates are representative of its long-run experience

Augmenting historical default or loss experience using data extracted from the Moody's KMV CRD (private cos.) and CE database (public cos.)

Augmentation structured according to client's existing/projected portfolio distributions by size, industry, private/public, country and risk profile





#2: Data Quality – Granularity of Data

SAMPLE

Company Descriptive Information
- Proper industry classification

Balance Sheet

Income Statement/Profit & Loss

Loan Accounting System
- Defaulted loan information

Financial Statement Information
- Income Statement / Profit & Loss
  - Net Sales
  - Cost of Goods Sold
  - Gross Profit

Loan Portfolio Information
- Loan Performance
  - Origination Date, Amount
  - Maturity Date
  - Interest Rate & Spread
  - Collateral type
- Default Information
  - Delinquency
  - Bank internal grades corresponding to default
  - Non Accrual
  - Write-off / Charge-off
  - Bankruptcy
  - Debt Restructuring

#2: Data Quality – Criteria for Assessment

Criteria for Assessing Data Quality:

- Diversity of data by industry, geography and size
- Accuracy of portfolio representation
- Accuracy of data elements
- Timeliness of historical reporting
- Financial auditing and reporting quality
- Availability of data, especially number and proportion of defaults captured

#3: Data Consistency

Application of definitions and filters

- Are key issues (default definition, default horizon, ratings philosophy) defined clearly and consistently across the dataset?
- Are data filters (proper SIC codes, "no duplicates" rule, similar data points) applied consistently to the dataset?

Changing organizational policies

- Where organizational policies shift, how should a bank maintain data integrity and avoid data gaps?
- If historical data or rating histories are unavailable for loans acquired through M&As or portfolio purchases, how should a bank fill the gaps in its IRB system?

#3: Data Consistency – Standardization of Loan Data Example

Tremendous variability within and across portfolios

- Loan type codes 5,700+
- Collateral type codes 750+
- Loan purpose codes 300+
- Loan index codes 900+
- Miscellaneous codes 800+ (call codes, flags, GL types)

Single codes contain multiple types of data

Bank	Source System	Code	Description
ABC	FGH	4568	Bank Prime Residential Real Estate: Construction
ABC	BAN	4568	Working Capital: Variable Rate

Wide variation in actual information database and limited consistency over time

- e.g. Basic pricing data such as interest margin over Loan Index is unavailable or miscalculated, such as when Interest Margins are uniformly indicated at 0% over Prime indexed loans, but actual Interest Rates provided vary, and are clearly different from current Prime Rate

#4: Industries with Low or No Defaults

- Typically smaller portfolios with limited structure financial analysis and standardized credit data collection
- Limited default experience and capture, and qualitative data often less structured
- Internal ratings often free-form and based on agency rating proxy or simulated model
- Cost and effort to develop internal models for these segments may not be practical
- Considering evaluating commercial models and data sources

#5: Qualitative Factors – Challenges

- Lack/insufficient usable history of qualitative data
- Qualitative factors contain subjective elements that are often open to interpretation
- Application of qualitative factors in historical ratings often difficult to determine and inconsistent
- Qualitative data does support composite or complex data
- Qualitative data lacks support for pooled data (cross-organization)

3

Best Practices in Data Management

Issues Addressed in this Section

Best practices in data management in terms of:

- Data collection
- Data cleansing
- Data maintenance

Discussion of some of Moody's KMV's data management practices:

- Basic content production system
- EDF production QC
- CRD data quality rules

Data Collection

Sources - What are the sources from which data is collected? Are they verified for accuracy, timeliness and completeness?

Data Capture - Is the bank capturing all significant quantitative and qualitative factors used to build and test the models?

Reference data requirements:

US Draft Supervisory Guidance on IRB Approach for Corporate Credit (2003) provides a listing of the type of data elements required under an IRB data framework

CRD data requirements

(examples in section on granularity of data)

Data Cleansing

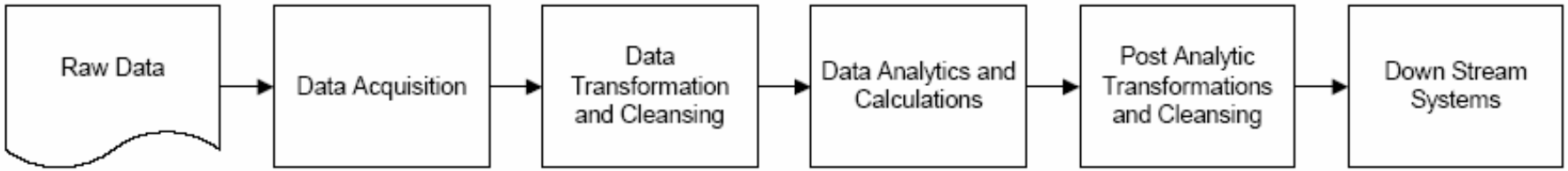
Rules and filters – What business rules and data quality filters has a bank adopted to prepare data for modeling purposes?

Examples:

- Statements that do not balance
- Duplicate identification
- Statements with too many missing values
- Incorrect or impossible values (triggers for checks: significant changes in data input between statements, extreme outliers)
- Proper SIC codes
- Confirmed borrowers
- Financial statement-related rules such as negative dividends, negative total liabilities, etc.

Data Maintenance

For most parts, Moody's KMV content production systems follow a basic model:



- Acquire data from external/internal sources
- Transform data for use in common components
  - Data mapping
  - Data cleansing
- Run data through analytics/calculation engine
- Transform output into a format usable by downstream systems
- Transfer data to target systems

CRD Data Handling Process

CREDIT INSTITUTION

DATA SUBMISSION

LOAN SYSTEM & FINANCIAL STMTS

CD, FTP

TXT, ACCESS, EXCEL, MFA

CRD ANALYST MAPS DATA

STANDARD FORMATS

SANITY CHECKS

EXAMPLE: TOTAL ASSETS < $0

DISTRIBUTION ANALYSIS

DISTRIBUTIONS OF INTERNAL GRADES, STMT TYPES, PORTFOLIO SIZE, ETC.

CONFIRMED BORROWER LINKS: NAME MATCH OR PROVIDED

MFA ID# 267308 KAMEN DISTRIBUTORS

LOAN ID# 599053475 KAMEN DIST LLC

2ND ANALYST REVIEW OF MAPPING & DISTRIBUTIONS

CLARIFICATION WITH CREDIT INSTITUTION

FIELD DEFINITIONS, EXPECTED PORTFOLIO CHARACTERISTICS, DEFAULT FLAGS, ETC.

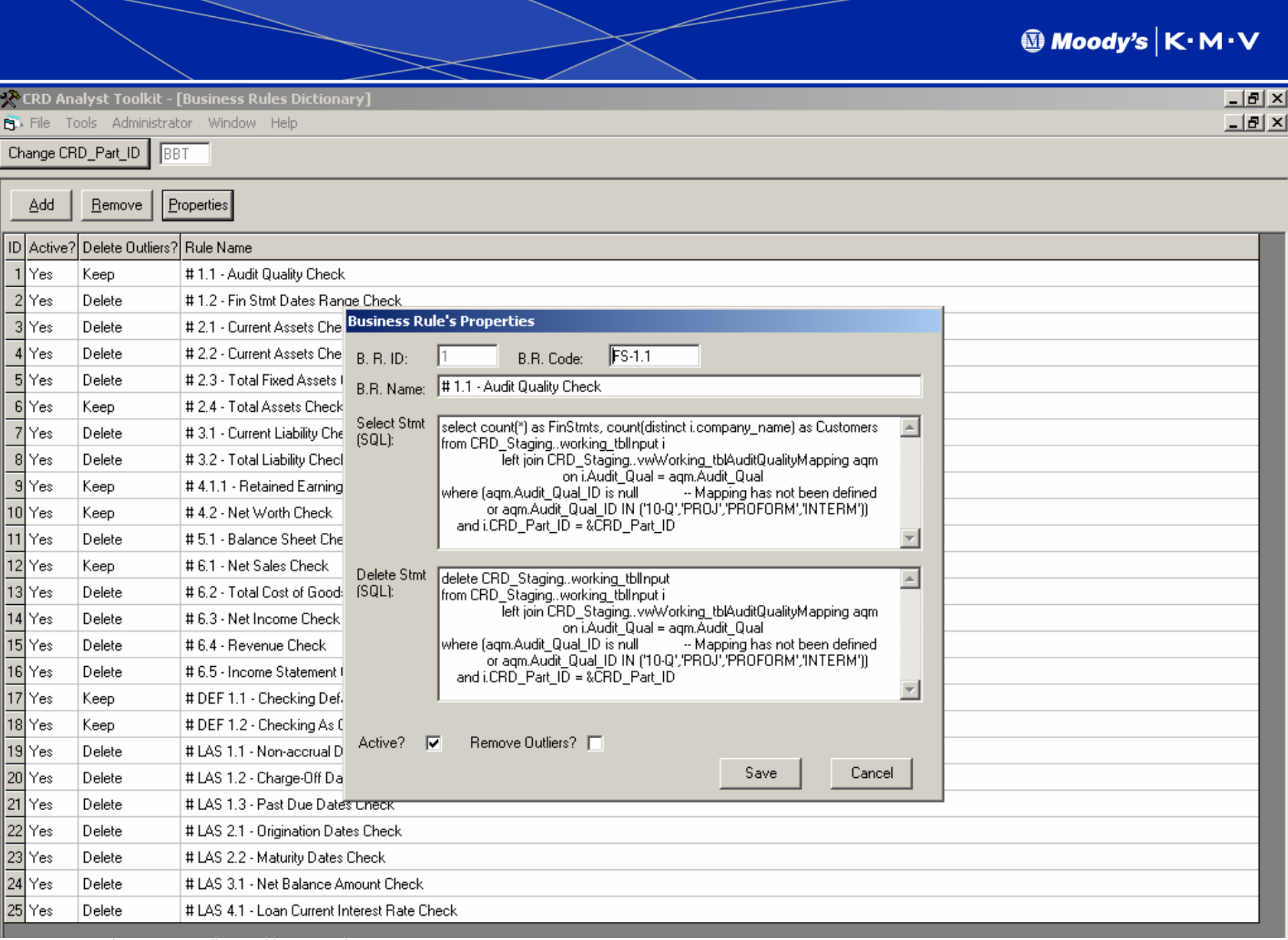
Moody's KMV Data Quality Rules (CRD Analyst Toolkit)

Moody's KMV-developed application standardizes individual participant data to CRD structure

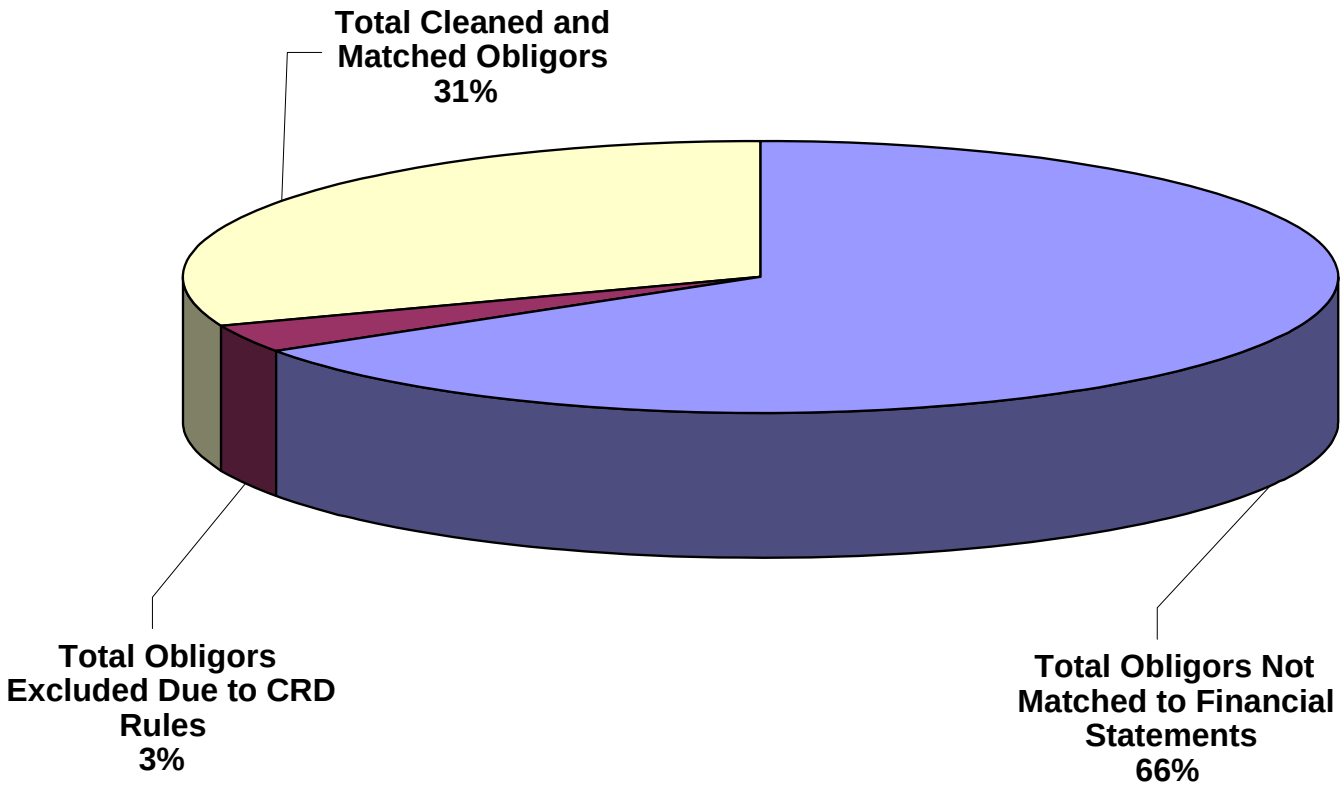
- Facilitates customer matching, validation, default identification and loading of participant data

Encompasses over 200 data quality rules

- QA rules developed over 7-year period
- Financial statement, loan accounting system checks designed to eliminate "noisy" data and create the cleanest comprehensive database for modeling and research
- For example, one rule is "Total Assets must equal Total Liabilities plus Net Worth (Equity)"



69% of CRD Participants' Portfolios are Unusable for CRD Purposes



4

Case Studies

Case Study: Large Global Bank
Data Management in a Validation-Calibration Project

Client Situation:

- Wanted to validate and calibrate their entire loan portfolio across a region, from small middle-market firms to global Fortune 100 companies
- Had the expertise to conduct such an analysis, but did not have manpower, computer hardware or internal organization to do it efficiently

Client Challenges:

- Lack of uniformity in observations for each country's dataset; substantially less defaults in some datasets than others
- Compare the behavior of the new version of the rating system to the previous version
- Validating the entire commercial and corporate portfolio for all major industries
- Calibrate each portfolio separately

Case Study: Large Global Bank

Value Added by Moody's KMV:

- Undertook data waterfall analysis to filter out inconsistent or invalid data, partly using MRA's Update Portfolio function
- Augmented the data with global defaults for validation purposes
- Validated and calibrated all commercial and corporate portfolios for all major industries
- Benchmarked results against RiskCalc, Credit Monitor and CreditEdge

End Result:

- Client had a system that was validated and calibrated in all portfolios, using more and cleaner data
- Extensive documentation of the whole process
- The project was completed much faster than if they had done it themselves

Case Study: Large Domestic Bank
Specific Data Diagnostics and Management Work

Client Situation:

- Wanted Moody's KMV to provide from scratch a total credit risk management solution (data analysis, scorecards/models building, credit process, implementation, validation, calibration)
- Discussion here focuses on the data aspects

Client Challenges:

- Insufficient historical loan performance data to build internal rating models but wishes to have industry-specific models
- Market in which client was operating was highly competitive and rapidly changing, rendering historical data relatively less meaningful in estimating future PD

* Stein, Roger M., The Relationship between Default Prediction and Lending Profits, Journal of Banking & Finance 29 (2005) 1213-1236

Case Study: Large Domestic Bank

Value Added by Moody's KMV:

- Investigated characteristics of client's loan portfolio, customer base and collateral pool
- Analyzed the type of data that should be incorporated in the database and the proper process for gathering and cleansing data
- Ran data checking algorithms (which allow semi-automation of data cleaning) and diagnosed issues (e.g., questionable audit flags, odd concentrations of low debt values)

End Result:

- Client established "clean, modeling-ready" dataset for the rest of the stages of the project
- Client received advice and recommendations as to the implementation of data infrastructure and management practices

* Stein, Roger M., The Relationship between Default Prediction and Lending Profits, Journal of Banking & Finance 29 (2005) 1213-1236

Concluding Remarks

- Active and structured management of credit data is essential for supporting validation and calibration of models
  - Start early and establish measured and systematic process for creating, enhancing and managing data within your organization
  - Always move from a measured process and progress in stepped process
  - Turnover in data management functions personnel can impact operation of data management
- Pooling of data will remain a viable and valuable option for most organizations
- Moody's KMV will continue to work and collaborate with you to bring ideas and solutions to support data management