Comptroller of the Currency
Administrator of National Banks
US Department of the Treasury

ENSURING A SAFE AND SOUND
NATIONAL BANKING SYSTEM
FOR ALL AMERICANS

Validation of Credit Rating and Scoring Models

A Workshop for Managers and Practitioners

February 2-3, 2006
The Omni Shoreham Hotel
Washington, DC

V2

Table of Contents

Welcome

Welcome to ***Validation of Credit Rating and Scoring Models***, a workshop sponsored by the Office of the Comptroller of the Currency (OCC). That you for attending, we hope you find the conference useful and informative.

For some time, banks have relied upon automated statistical models for assessing borrower creditworthiness during the underwriting process. Increasingly, banks are using rating and scoring models for measuring and managing portfolio credit risk. Even those banks that not participated in this recent trend will have to develop models if they choose to participate in the advanced internal ratings-based (AIRB) approach to regulatory capital.

The application to portfolio credit risk and regulatory capital elevates the level of model risk in rating and scoring models. This implies a commensurate increase in the emphasis that must be placed upon model-validation as a control.

The banking industry and OCC have identified sound validation principles over the years, and we have elaborated on those principles in OCC Bulletin 2000-16. This workshop expands upon 2000-16 in two ways. First, we discuss the application of those general validation principles to the specific case of scoring and rating models. Second, we describe the degree of validation rigor that we expect to become the industry standard for portfolio credit-risk models.

The Office of The Comptroller of the Currency

The OCC was established in 1863 as a bureau of the U.S. Department of the Treasury to charter, regulate, and supervise all national banks. It also supervises the federal branches and agencies of foreign banks. Headquartered in Washington, D.C., the OCC has four district offices plus an office in London to supervise the international activities of national banks.

The OCC is headed by the Comptroller, who is appointed by the President, with Senate confirmation, for a five-year term. The Comptroller also serves as a director of the Federal Deposit Insurance Corporation and as a director of the Neighborhood Reinvestment Corporation.

The OCC's nationwide staff conducts on-site reviews of national banks and provides sustained supervision of bank operations. OCC staff analyze a bbank's loan and investment portfolios, funds management, capital, liquidity, sensitivity to market risk, and compliance with applicable laws and regulations. They review banks' internal and external audit and risk-management information systems to help evaluate bank management's ability to identify and control risk.

The agency issues rules, legal interpretations, and corporate decisions concerning banking, bank investments, bank community development activities, and other aspects of bank operations. In regulating national banks, the OCC has the power to

- Examine the banks.
- Approve or deny applications for new charts, branches, capital, or other changes in corporate or banking structure.
- Take supervisory actions against banks that do not comply with laws and regulations or that otherwise engage in unsound banking practices. The agency can remove officers and directors, negotiate agreements to change banking practices, and issue cease and desist orders as well as civil money penalties.
- Issue rules and regulations governing bank investments, lending, and other activities.

The OCC does not receive any appropriations from Congress. Instead, its operations are funded primarily by assessments on national banks. National banks pay for their examinations, and pay for the OCC's processing of their corporate applications. The OCC also receives revenue from its investment income, primarily from U.S. Treasury securities.

An Important Note

The workshop organizers have worked hard to put together a set of presentations that describe sound model-risk control processes in the context of credit rating, scoring, and portfolio credit-risk measurement and management. Nevertheless, the fact that a particular procedure is described during a presentation does not necessarily imply that the OCC requires or expects banks to implement the procedure; nor does it imply that a particular set of procedures will necessarily ensure a sound validation process. Statements made by presenters are their own, and do not necessarily represent the views of the Office of the Comptroller of the Currency or the U.S. Department of the Treasury.

Program Overview

Validation of Credit Rating and Scoring Models

The Omni Shoreham Hotel
Washington, DC 20008

Thursday, February 2, 2006

10:00am-1:00pm On-Site Registration
The Regency Ballroom Prefunction Area

1:00 Workshop Opens
All presentations will be be held in the Regency Ballroom

1:00-1:30 Opening Remarks: **Model Validation as a Process**

1:30-2:45 Session I: **Building and Validating Credit Rating and Scoring Models**

2:45-3:00 Break
All Breaks will be held in the Ambassador Ballroom

3:00-4:15 Session II: **Evaluating Discriminatory Power and Forecast Performance**

4:15-4:30 Break

4:30-5:15 Session III: **Examples of Model Design and Quantification**

5:15-6:00 Audience Discussion

6:30-8:00 Reception
The Empire Ballroom

Friday, February 3, 2006

7:00am Breakfast

8:00-8:45 Session IV: **Process Verification and Data Maintenance**

8:45-9:00 Break

9:00-10:15 Session V: **Monitoring and Benchmarking**

10:15-10:30 Break

10:30-12:00pm Session VI: **Industry Panel and Audience Discussion**

12:15-1:30 Lunch and Speech by Comptroller Dugan
The Empire Ballroom

1:45-3:00 Session VI: **Validation as a Control Function Under Basel II**

3:00pm Workshop Ends

Session Detail

Opening Remarks: Model Validation as a Process

Jeffrey Brown, Senior Deputy Comptroller
OCC International and Economic Affairs

Jeffrey A. Brown is the Senior Deputy Comptroller for International and Economic Analysis in the Office of the Comptroller of the Currency. At the OCC, Brown also served as the director of the Risk Analysis Division. He has published research on the causes of bank failure, the condition of real estate markets, the risks of bank lending to real estate activities, bank behavior in reserving for loan and lease losses, and bank cost efficiencies and credit rating. As a policy advisor at the OCC, Brown has worked on a broad range of issues dealing with risk management, risk measurement, and risk modeling practice. He has participated in the group developing the U.S. interagency supervisory implementation procedures for the internal ratings-based approach to the new Basel Capital Adequacy Framework. Brown holds a PhD. in economics from Brown University.

Session I:
Building and Validating Credit Rating and Scoring Models

Dennis Glennon, Deputy Director for Credit Risk Modeling
OCC Risk Analysis Division

Credit rating and scoring models are used for multiple purposes: underwriting, account management, pricing, and economic capital allocation. As such, the models should be designed to reflect the specific purpose for which the models will be used. Credit scoring models used for underwriting and account management purposes are generally developed to identify relative risk across the portfolio; however, models used for pricing and capital allocation should be constructed to measure absolute risk.

There are industry-accepted methods for constructing and validating credit rating and scoring models. These methods are designed primarily to assess the discriminatory power of the models. In this session, we discuss several of the more commonly used model development and diagnostic tools. We demonstrate the importance of applying a broad range of diagnostic and validation tests to assess the reliability of model performance.

Session II: Evaluating Discriminatory Power and Forecast Performance

C. Erik Larson, Lead Expert for Enterprise Risk
OCC Risk Analysis Division

Increasingly, credit models are being relied upon to produce accurate forecasts of individual loan performance. This presentation considers estimation technique and model assessment when accuracy and goodness-of-fit are modeling objectives.

Session III: Examples of Model Design and Quantification

Mitch Stengel, Lead Expert for Basel II
OCC Modeling and Analysis

Earlier presentations have laid the conceptual groundwork for rating and scoring model validation. This presentation applies that conceptual framework to a concrete example in the form of a bank's risk rating model for its corporate portfolio. While the names and specific details have been changed, this case study is based on current practice at a number of large national banks.

The case study goes through the principal stages of risk rating model design and construction, from data sample selection through the analysis of a hold-out / out-of-time sample. It illustrates and emphasizes the importance of validation processes (developmental evidence) at every step of the way. Finally, it shows that, in spite of differences in detail and terminology, the model building process and concomitant validation activities are fundamentally the same for retail and wholesale portfolios.

Session IV: Process Verification and Data Maintenance

Michael Carhill, Director
OCC Risk Analysis Division

OCC Bulletin 2000-16 points out that models have three components: inputs, processing, and outputs. Most of the intellectual discussion about model validation turns around the validation of outputs, and this conference focuses on that area. However, most of the resources required for a sound validation process will be expended in the validation of inputs and processing. A good validation process will ensure that a model is functioning as intended. In a loose sense, it is practicable to achieve perfection in the model inputs and processing, while it is impossible for model outputs to be perfect. This suggests why some believe that a good model-validation process is successful when it can guarantee that the models are "perfectly wrong."

Session V: Monitoring and Benchmarking

M. Nazmul Hasan, Lead Expert for Credit Risk Modeling
OCC Risk Analysis Division

This presentation provides a conceptual framework for ongoing monitoring and benchmarking. Although monitoring and benchmarking may appear to be two distinct and independent validation processes, effective monitoring cannot be conducted without appropriate benchmarking. This may include front-end analysis of the score distribution, back-end analysis of performance measures, and analysis of risk characteristics (risk drivers). To conduct such analyses the common benchmarks considered are development sample and alternative models. Banks should continue to explore and develop new statistical methodologies and quantitative techniques for this element of validation.

Industry Panel

Chaired by Nicholas M. Kiefer
Cornell University and OCC Risk Analysis Division

Luncheon Speaker

John C. Dugan, Comptroller of the Currency

John C. Dugan was sworn in as the 29th Comptroller of the Currency in August 2005. Prior to his appointment as Comptroller, Mr. Dugan was a partner at the law firm of Covington & Burling, where he chaired the firm's Financial Institutions Group. He specialized in banking and financial institution regulation. He also served as outside counsel to the ABA Securities Association.

He served at the Department of Treasury from 1989 to 1993 and was appointed assistant secretary for domestic finance in 1992. While at Treasury, Mr. Dugan had extensive responsibility for policy initiatives involving banks and financial institutions, including the savings and loan cleanup, Glass-Steagall and banking reform, and regulation of government-sponsored enterprises. In 1991, he oversaw a comprehensive study of the banking industry that formed the basis for the financial modernization legislation proposed by the administration of the first President Bush.

From 1985 to 1989, Mr. Dugan was minority counsel and minority general counsel for the U.S. Senate Committee on Banking, Housing, and Urban Affairs. There he advised the committee as it debated the Competitive Equality Banking Act of 1987, the Proxmire Financial Modernization Act of 1988, and the Financial Institutions Reform, Recovery, and Enforcement Act of 1989.

Among his professional and volunteer activities before becoming Comptroller, he served as a director of Minbanc, a charitable organization whose mission is to enhance professional and educational opportunities for minorities in the banking industry. He is also a member of the American Bar Association's committee on banking law, the Federal Bar Association's section of financial institutions and the economy, and the District of Columbia Bar Association's section of corporations, finance, and securities laws.

A graduate of the University of Michigan in 1977 with an A.B. in English literature, Mr. Dugan also earned his J.D. from Harvard Law School in 1981. Born in Washington, D.C. in 1955, Mr. Dugan lives in Chevy Chase, Md., with his wife, Beth, and his two children, Claire and Jack.

Session VI: Validation as a Control Function Under Basel II

Mark Levonian, Deputy Comptroller
OCC Modeling and Analysis

The Basel II framework includes explicit expectations for robust validation of banks' credit-risk systems under the internal ratings-based (IRB) approach to capital. Banks using IRB must validate their internal processes for differentiating risk as well as for quantifying that risk. Validation should be viewed as an integral part of the broader control processes around IRB systems, and evaluated within the context of those other controls. Generally accepted validation principles and generic types of validation tools can and should be applied. However, the specifics of validation may be different in this new setting, and banks are likely to need new tools and data.

Comptroller of the Currency
Administrator of National Banks

US Department of the Treasury

ENSURING A SAFE AND SOUND
NATIONAL BANKING SYSTEM
FOR ALL AMERICANS

Opening Remarks: Model Validation as a Process

Jeffrey Brown

Senior Deputy Comptroller
OCC International and Economic Affairs
Washington, DC 20219

2006 Validation of Credit
Scoring Models Workshop

Model Risk

Financial firms and their regulators are comfortable thinking about financial risks—credit risk, market risk, and interest rate risk. While it would be convenient if we could physically observe and measure those risks, we cannot. Therefore, among the tools that we use to think about those risks are models to identify and quantify them.

It is intrinsic to the notion of models that they feature risk. In the realm of finance modeling, the object of interest is future outcomes—either expected or potential—and those modeled outcomes can be wrong. Humans construct models, so models are exposed to all of the sources of error of any other human construct—errors in logic, errors in execution, and errors in use. Those errors are what we call model risk.

However, determining whether models are wrong is difficult. It is often said—but still true—that all models are wrong by design, because they are simplifications of reality. Models reflect knowledge, beliefs or assumptions about important casual relationships. Those causal statements are, by design, highly abstract. They are not designed to capture every detail of reality, including the idiosyncratic factors that contribute to observed outcomes. Furthermore, when models are used to generate predictions—and risk models are included in that class—the 'predicted' outcomes reflect future outcomes of exogenous variables. Even a very good statement of the causal relationship will deliver wrong predictions under most outcomes for the exogenous variables. So the modeled outcomes are conditional on a very precise artificial construct and specific set of conditions. Under the most favorable circumstances, tests of such models are probability statements, but under most circumstances model adequacy is determined by expert judgment. While this is a disturbing notion for the non-modeler, it is one that must be confronted because it determines strategic choices in model building, model administration and model use.

Model-Risk Analysis

The appropriate response to model risk is to risk manage the use of models, just as we manage the risks of all other aspects of running the enterprise. The objective of the model builder is to devise the best model for the business use; the objective of model-risk management is to determine whether that has been accomplished.

Model users must acknowledge at the outset that models are imperfect and put in place a process for controlling the risk that they are not good enough to use. Model users need to employ a model validation process that is designed to provide the best available evidence that a model is good. Such a process entails the evaluation of model development, verification that it is operating as planned, and monitoring for evidence that contradicts the model. The model validation process is ongoing, critically dependent on expertise, and costly.

At the OCC, we have responded to the growing importance of modeling in banking by examining for model validation processes. We embodied our expectations in a banking bulletin, (OCC Bulletin 2000-16, "Model Validation"), which is generally applicable in all modeling contexts. This bulletin describes a framework that values the principles of independence and assigned responsibilities in checking models, recognizes the importance of documentation and ongoing testing, and makes it clear that bank management is responsible for recognizing model risk and devoting adequate resources to addressing it.

Model-Risk Analysis in a Credit Risk Context

Credit rating and scoring models present a distinct set of challenges to model validation. The primary event of interest—default—is rare. When defaults do occur, they tend to happen in batches, implying long spells during which defaults are more rare. All of this means that the comparison of model predictions to outcomes—back testing—is not statistically powerful. Adding to the level of difficulty of the validation challenge, there is shortage of generally accepted standard models against which to compare.

Recognizing the challenges of model validation for credit rating and scoring models, it becomes increasingly important that the users of those models employ a complete process that offsets the limitations of any individual test.

- The first element of that process is to demonstrate that the model is well developed. Models should be logical on an a priori basis. Models need to be supported with empirical evidence that they can identify credit risks in a data set that is well designed for model development purposes. And modelers should be sensitive to the risk of trying to describe a development data set perfectly when some of the outcomes in the developmental data may be random.
- Given a well-developed model, the second element of the process is ongoing verification that the model is working as expected. Ongoing verification includes activities designed to confirm that the model is implemented as designed and activities designed to get an early read on whether the models is likely to be working. Process verification includes checking equations and the computer code that deploys the model. Equally important, process verification must include mechanisms to assure the quality of the data inputs. And process verification includes the evaluation of reports to confirm that they are understandable and well used. Another key aspect of ongoing verification is the comparison of model predictions to

predictions from other useful sources—benchmarking—to confirm the likely correctness of the predictions.

- The third element of model validation is outcomes analysis. In this phase, where practicable, model predictions are compared to actual outcomes. While theoretically compelling, model users must understand that statistically powerful outcomes tests may be rare, and must not count on this evidence alone.

Model-Risk Analysis in the Basel II Context

The advent of regulatory capital requirements that will be a function of internal bank credit risk assessments raises the stakes for model validation. While Basel II is a regulatory capital framework, bank management will be responsible for model validation; bank validation processes will be the first line of defense against bad credit risk models. Just as we do in other risk management contexts, bank supervisors will examine the bank validation processes. It must be recognized, however, that the added importance of being part of the capital framework means that some validation deficiencies that supervisors might otherwise have deemed immaterial will be brought to management's attention.

Comptroller of the Currency
Administrator of National Banks
US Department of the Treasury

ENSURING A SAFE AND SOUND NATIONAL BANKING SYSTEM FOR ALL AMERICANS

Session I
Building and Validating Credit Rating and Scoring Models

Dennis Glennon

Deputy Director – Credit Risk Modeling
Risk Analysis Division
Washington, DC 20219
Dennis.Glennon@occ.treas.gov

2006 Validation of Credit Scoring Models Workshop

Outline

- Introduction
- Developmental Evidence: Building statistical-based rating/scoring models
- Performance Evaluation: Verifying the model works

Introduction

- Is there a supervisory concern?
  - Sound modeling practices
- How do we approach the supervision of model risk?
  - Focus on two fundamental properties of a valid modeling process:
    - logically consistent model/sample design
    - valid statistical methods

Introduction

- Sound modeling practices
  - There are generally accepted, or industry-accepted, methods of building and validating models.
  - These methods incorporate procedures developed in the statistics, econometrics, information-theory, and operations research literature.
  - Although these methods are valid, they may not be appropriate in all applications.
    - A model selected for its ability to discriminate between high and low risk may perform poorly at predicting the likelihood of default.

Developmental Evidence

- Model development is a process.
- Simply stated:
  - define the purpose – discrimination vs prediction;
  - <u>select a sample</u> that reflects/represents the targeted population – a reference data set;
  - <u>select a modeling technique</u> consistent with the purpose;
  - <u>identify risk factors</u> that reflect the lender's knowledge and historical experience;
  - <u>fit the model</u> and check for model mis-specification or overfitting of the data;
  - <u>develop methods of verifying</u> that the model works – outcome-based methods.

Developmental Evidence

- Sample-design issues
  - Missing data
    - not available
    - censored data (i.e., reject inference)
    - truncated data (i.e., prepayment/attrition)
  - Omitted variables (implicitly held constant)
    - product terms (e.g., price, payment options)
    - economic conditions (e.g., interest rates, employment, business/industry conditions)
  - Pooling time-sensitive data

Developmental Evidence

- Modeling techniques
  - Expert systems
  - Regression
    - logit, probit, least squares, neural network
  - Decision-tree methods
    - CHAID, CART
  - Linear programming

Developmental Evidence

- *Step 1:* Univariate analysis used to reduce the set of potential risk factors to a subset of feasible risk factors
  - correlation
  - weight of evidence
- *Step 2:* Multivariate analysis used to capture the combined effect of multiple factors on expected performance
  - regression approach

Developmental Evidence

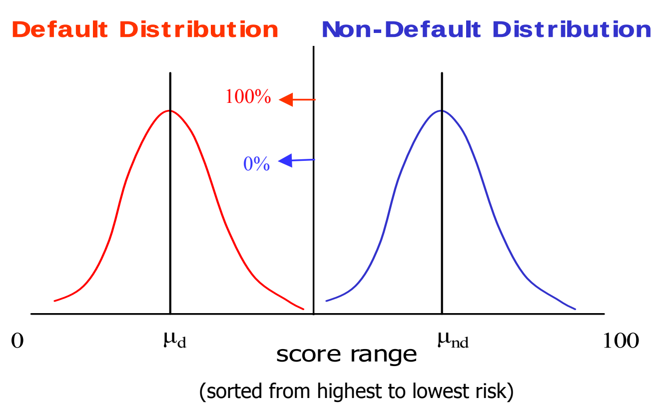
- Selecting the "best" model
  - Business sense
  - Diagnostic test
  - Out-of-sample analysis
    - In-time sample (i.e., hold-out)
    - Out-of-time sample
  - Cross-validation/benchmark analysis

Performance Evaluation

- Common performance measures
  - Kolmogorov-Smirnov (K-S)
  - Gains charts/cumulative accuracy profiles (CAP)
  - Divergence
  - Log-odds

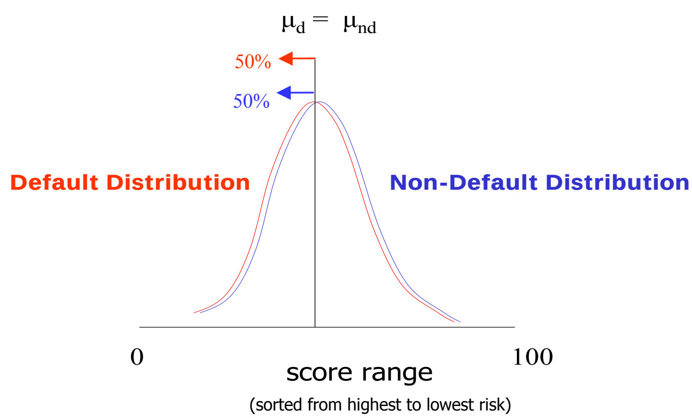
Performance Evaluation

- Model Performance Measures - K-S
  - *Upper Bound:* If the scores partition the population into two separate groups in which one group contains all the defaulted accounts and the other all the non-defaulted accounts, then the K-S is 100.



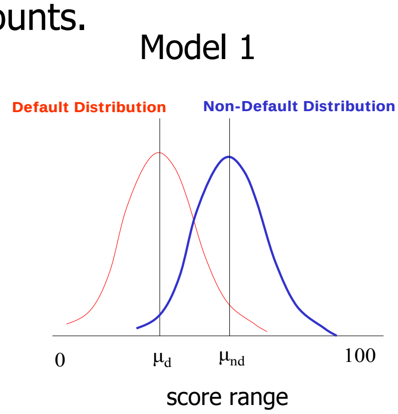
Performance Evaluation

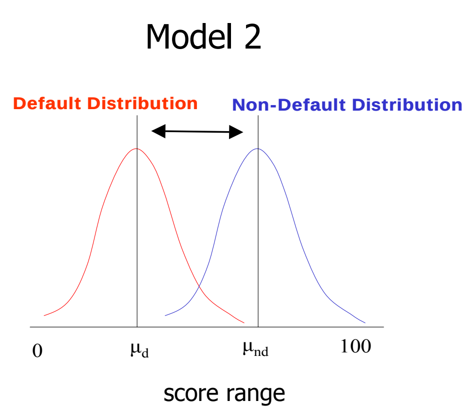
- Model Performance Measures - K-S
  - *Lower Bound:* If the model can not differentiate between non-defaulted and defaulted accounts, then it is as if the model selects individuals randomly from the population. There would be no difference in the location of the distributions. The K-S would be 0.



Performance Evaluation

- Model Performance Measures - K-S
  - These results suggests that the K-S value will fall between 0 and 100, and that the higher the value the better the model is at separating the non-defaulted from defaulted accounts.





Performance Evaluation

Model performance: Kolmogorov-Smirnov (K-S)

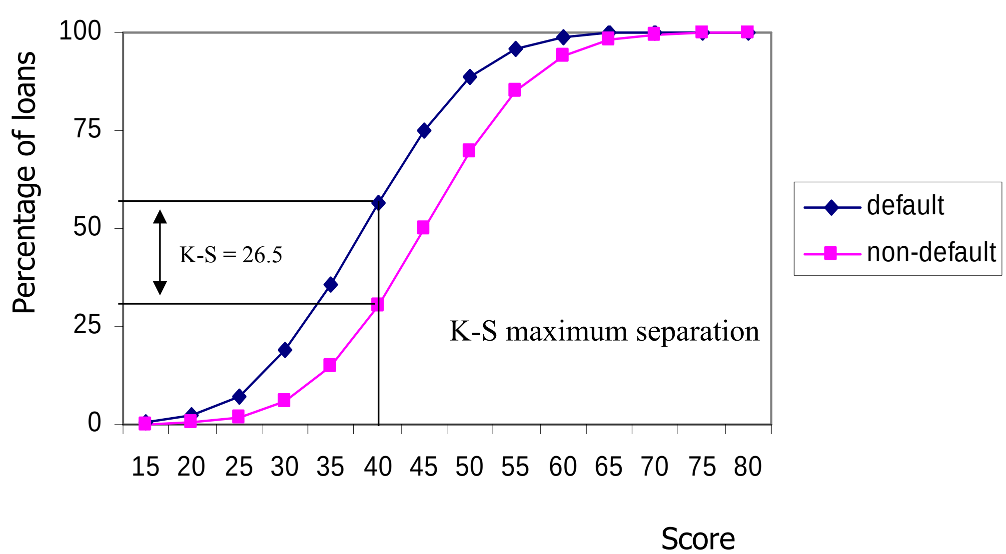
Obs (i)	Score Range lower	Score Range upper	Distributions Default (#)	Distributions Non-Def (#)	Cumulative Distributions Default (%)	Cumulative Distributions Non-Def (%)	K-S
1	0	15	82	458	0.34	0.05	0.29
2	15	20	428	3205	2.12	0.37	1.75
3	20	25	1235	13886	7.24	1.75	5.49
4	25	30	2778	41657	18.77	5.92	12.85
5	30	35	4074	91645	35.69	15.09	20.60
6	35	40	5092	152741	56.82	30.36	**26.46**
7	40	45	4365	196381	74.94	50.00	24.94
8	45	50	3274	196381	88.53	69.64	18.89
9	50	55	1698	152741	95.58	84.91	10.67
10	55	60	764	91645	98.75	94.08	4.67
11	60	65	232	41657	99.71	98.24	1.47
12	65	70	58	13886	99.95	99.63	0.32
13	70	75	9	3205	99.99	99.95	0.04
14	75	80	1	458	100	100	0.00
15	80	100	1	31	100	100	0

(82+428)/24092

Total Bad = 24092

Performance Evaluation

- Cumulative non-default and default distributions



Performance Evaluation

- Separation as a modeling objective
  - *Comment:* The K-S statistic is *not* a measure derived from the difference between the actual and predicted values of the dependent variable; as such, it is *not* an R^2-type measure of model accuracy.
  - *Comment:* For that reason, in practice, the K-S test is used to evaluate the model as a segmentation or classification tool. As a result, this test does not necessarily identify the model that is best at predicting the probability of default.

Performance Evaluation: K-S Test

- *Hypothesis Test:* the difference between two distributions
  - Test statistic (K_α)

$$\mathrm{K}_{\alpha} = 100\,\{\,\mathrm{D}\left[\frac{(\#\,non\_defaults + \#\,defaults)}{(\#\,non\_defaults)(\#\,defaults)}\right]^{1/2}\}$$

where α = significance level (e.g., .95)
D = critical value (table value)

Performance Evaluation: Hypothesis Test

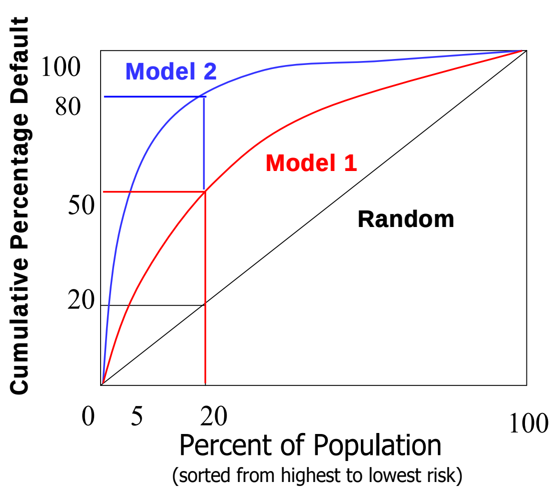
- Difference between two distributions
  - *Example 1 (from above):* Default Rate = 2.35%
    - # Defaults = 24,091
    - # Non-defaults = 999,977
    - $K_{\alpha=.95}$ = 0.80% < KS = 26.5%

- *Example 2:* Default Rate = 4.41%
  - # Defaults = 441
  - # Non-defaults = 9,559
  - $K_{\alpha=.95}$ = 5.94%

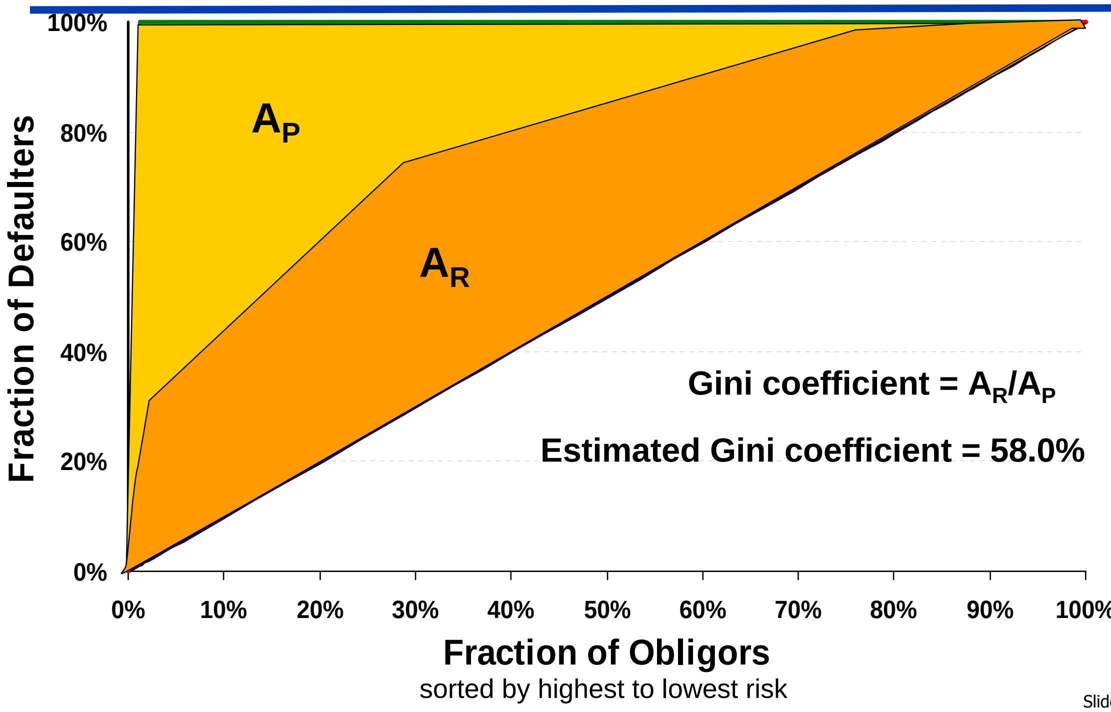
- *Example 3:* Default Rate = 50.0%
  - # Defaults = 1,500
  - # Non-defaults = 1,500
  - $K_{\alpha=.95}$ = 4.45%

Performance Evaluation: Gains Chart

- Model Performance Measure



Performance Evaluation: Gini Coefficient



OCC VCRSM Workshop, February 2006

Performance Evaluation: Gini Coefficient

- There is no magic number
  - Higher is better, but there is no ratio that says a scoring or rating system is "good" or "bad"
- All errors are *not* created equal
  - Gini coefficient treats "false negatives" and "false positives" as equally bad
- Be careful about making comparisons
  - Dispersion of credits across score ranges or grades
  - Number of defaulters in sample
  - Portfolio composition

Performance Evaluation: Diagnostics

- Goodness-of-fit measures
  - R^2-type measure of goodness-of-fit are generally not used.
- Robustness test: out-of-sample analysis
  - *In-time sample*: Observations randomly selected from the development or reference data.
  - *Out-of-time sample*: Observations randomly selected from a population with observation and performance periods different from those of the reference sample.

Performance Evaluation

- In-time and out-of-time analyses
  - The models are evaluated in terms of their ability to maintain:
    - stable parameter estimates across the different validation samples, and
    - a given level of separation between the good and bad distributions (i.e., stable K-S statistics).

Performance Evaluation

- Are these tools really useful?
- *Illustrative Example:* Developing a Credit Scoring Model for Risk Segmentation Purposes
  - *Sample:* A simulated random sample of 10,000 observations.
  - *Performance:* Derived from the following data generating process.

Illustrative Example: Data

- *Data Generating Process*
  - $y = 0$ if $Y < 0$
    $y = 1$ otherwise
  - $Y = \mathbf{-52.5} + \mathbf{1.0}\ r_1 + \mathbf{1.0}\ r_2 + \mathbf{1.0}\ r_3 + \mathbf{2.0}\ r_4 + \mathbf{2.0}\ r_5 + \mathbf{2.0}\ r_6 - e$
  - where
    - $e \sim \text{logistic}(0,\pi^2/3)$
    - $r_1 - r_6$ are uncorrelated continuous random variables
  - pr(default) = pr(y=1)
  - mean of y = 0.0866

Illustrative Example: Univariate

Variables	Estimated Parameters β	P-Values Pr > ChiSq	Divergence Index
r1	0.1442	0.0001	0.3746
r2	0.1415	0.0001	0.4875
r3	0.1339	0.0001	0.4488
r1	0.3729	0.0001	3.746
r2	0.2917	0.0001	1.383
r3	0.2911	0.0001	1.345
w1	0.0356	0.3172	0.0048
w2	-0.0795	0.0256	0.0192
w3	-0.0197	0.5792	0.0039
w4	0.0523	0.1019	0.7826
w5	0.2237	0.0001	0.2839
w6	0.0412	0.2465	0.0033

The estimated parameters β are derived from the univariate (logit) regression models:
$y = b_0 + b_1 x$ where x = {r1, r2, r3, ..., w6}; and y = 1 if default, 0 otherwise.
The Divergence Index is: $D = \Sigma^{10}_{g=1} (p_g - q_g) \ln(p_g/q_g)$, where p_g (q_g) is the percentage of non-default (default) accounts in the g^{th} decile.

Illustrative Example: Logit Model

Variables	Development Parameters	Exact: Estimated Parameters	Over-specified: Estimated Parameters	Over-specified: Estimated Parameters	Mis-specified: Estimated Parameters	
Intercept	-52.5	-53.1642	-53.2803	-53.2157	-11.7397	<.0001
r1	1.0	1.0127	1.0141	1.0128	0.3476	<.0001
r2	1.0	0.9777	0.9781	0.9779	0.3255	<.0001
r3	1.0	0.9745	0.9775	0.9746	0.3201	<.0001
r4	2.0	2.0327	2.0312	2.0330	0.6712	<.0001
r5	2.0	1.9921	1.9715	1.9673	1.3224	<.0001
r6	2.0	2.0185	2.0304	2.0254		
w1	0		-0.0575			
w2	0		-0.0769			
w3	0		0.0166			
w4	0		0.4387	0.4401	-9.7901	<.0001
w5	0		0.0312			
w6	0		0.0307			

Illustrative Example

Parameter Stability

Variables	Exact Specification Development	Exact Specification Validation	Over-Identified Development	Over-Identified Validation	Mis-Specified Development	Mis-Specified Validation
Intercept	-53.1642	-52.8708	-53.2157	-52.8258	-11.7397	-10.3896
r1	1.0127	1.0195	1.0128	1.0197	0.3476	0.3161
r2	0.9777	1.0070	0.9779	1.0072	0.3255	0.3108
r3	0.9745	1.0214	0.9746	1.0217	0.3201	0.3188
r4	2.0327	1.9487	2.0330	1.9494	0.6712	0.6006
r5	1.9921	2.0400	1.9673	2.0669	1.3224	1.0034
r6	2.0185	2.0384	2.0254	2.0315		
w4			0.4401	-0.1527	-9.7901	-1.8131

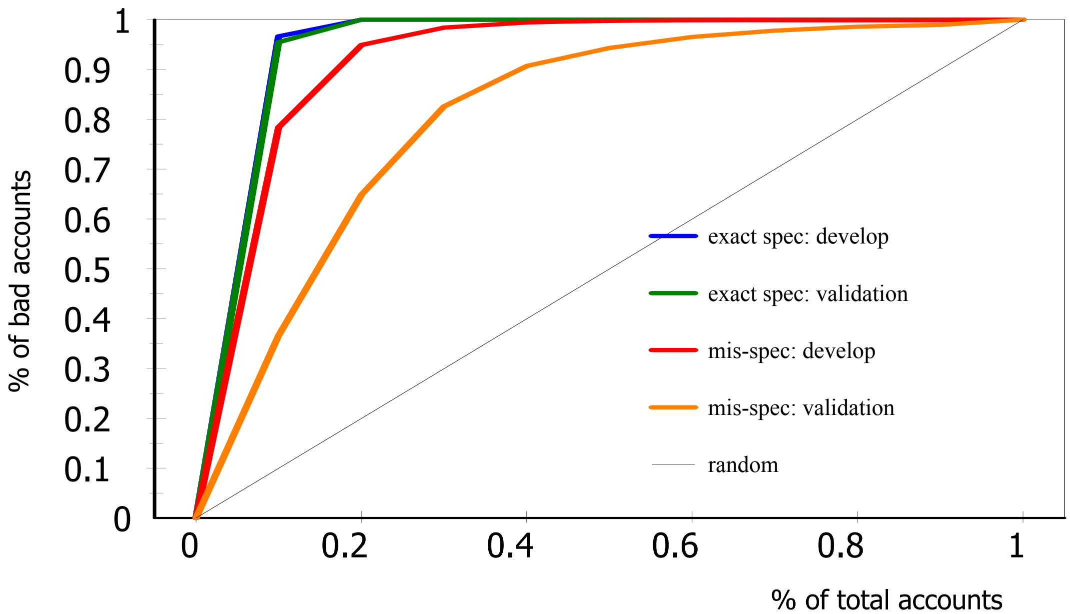
black – statistically significant at the 1% level
blue – statistically significant at the 10% level
red – statistically insignificant at the 10% level

Illustrative Example

- K-S Test
  - The K-S stat is a measure of the degree of separation between the non-default and default distributions.

Model	Sample Development	Validation	Total
Exact	94.9	93.6	94.1
Over-Identified	94.7	93.2	94.0
Mis-Specified	82.0	57.6	66.2

Illustrative Example: Gains Charts



Performance Evaluation: Other Issues

- Developing benchmarks for performance monitoring and early-read/early-warning analysis.
  - Benchmark values and distributions constructed at time of model development are used to differentiate between
    - temporary shifts due to "random" shocks
    - permanent drift due to structural changes

Conclusion

- Model development is a process.
- Models should be developed using sound modeling practices.
- Model verification is an integral part of the model development process.

Session II

Evaluating Discriminatory Power and Forecast Performance

C. Erik Larson

Lead Expert for Enterprise Risk
Risk Analysis Division
Washington, DC 20219
Erik.Larson@occ.treas.gov

2006 Validation of Credit Scoring Models Workshop

Outline

- Modeling Objectives
- Traditional Credit Risk Model Design
- Discriminatory Power
- Forecast Performance
  - Business Decisions
  - Accuracy and Precision
  - Evaluating Rating or Score Level Prediction
  - Evaluating Global Fit

Modeling Objectives

- Should be linked to business needs and use
- Can influence:
  - the logical design of the model
  - the sampling design
  - the statistical techniques employed in estimation
  - the benchmarking and performance tracking techniques
  - the interpretation of validation results
- Should generally be determined early in the modeling process

Modeling Objectives

Discrimination and Prediction

- The qualitative or ordinal **discrimination** between two or more types of credit
- *Examples:*
  - *Risk ranking of delinquent borrowers to allocate followup-efforts*
  - *Segmentation of applications for different review*

- The **forecasting** of cardinal risk levels for individual credits
- *Examples:*
  - *Default probability estimation*
  - *Loss forecasting*

Traditional Credit Risk Model Design

- Default, delinquency and segmentation models have traditionally been developed to meet a classification objective.
- The dependent variable of interest takes a limited set of values, {0,1}, corresponding to membership in a class.
- *Examples:*
  - *Good vs. Bad*
  - *Non-Delinquent vs. Delinquent*
  - *Non-Default vs. Default*
  - *Low Risk vs. High Risk*

Traditional Credit Risk Model Design

- Rating and scoring models develop predictions of class membership as a function of borrower characteristics, X_i.
- Typical Model
  - The score: $z_i = Z(X_i, \hat{\beta})$
  - Implementation:
    Choose a score cutoff z*.
    If borrower's score is less than cutoff, predict bad;
    if score is greater than or equal to cutoff, predict good.
- Let F(z|Good) and F(z|Bad), respectively, represent the cumulative distribution functions of "good" borrowers and "bad" borrowers generated by the score.
- Question: How should z* be chosen?

Discriminatory Power

Choosing a Score Threshold: Types of Errors

- Model predictions of class membership are compared to realized outcomes

		Realized Outcome	
		Good	**Bad**
Predicted Outcome	**Good** $z_i > z^*$	No Error	Type I Error
	Bad $z_i <= z^*$	Type II Error	No Error

- This is closely related to retail scorecard "swap-set" analysis

Discriminatory Power
One Score Threshold: The K-S Statistic

- One way to choose z* is by picking a value that minimizes expected costs from making Type I and Type II errors:

$$c_b \text{ Prob[Type I Error]} + c_g \text{ Prob[Type II Error]}$$

$$c_b \text{ (1-F(z*|Bad) Prob[Bad]} + c_g \text{ F(z*|Good) Prob[Good]}$$

- Here "c_b" is the cost from making a loan that turned out bad, and "c_g" is the opportunity cost of failing to make a loan that would have turned out good.

- Note that if c_gProb[Good] = c_bProb[Bad], then this problem reduces to maximizing

$$F(z^*|\text{Bad}) - F(z^*|\text{Good})$$

- *This is equivalent to setting the score threshold at the value for which the K-S statistic is maximized (see Thomas, et. al. [2002])*

A problem with this argument....

- We seldom observe approve/decline decisions made by setting a cutoff equal to the score value which maximizes K-S.

- In fact, we usually see many thresholds used in decisioning. What should we conclude?

- ***The use of K-S to evaluate a model's discriminatory power might not provide insight into the model's performance in the range required by the business decision*** *(Hand [2004])*

- ***Other metrics might be needed.***

Forecast Evaluation
Accuracy and Precision

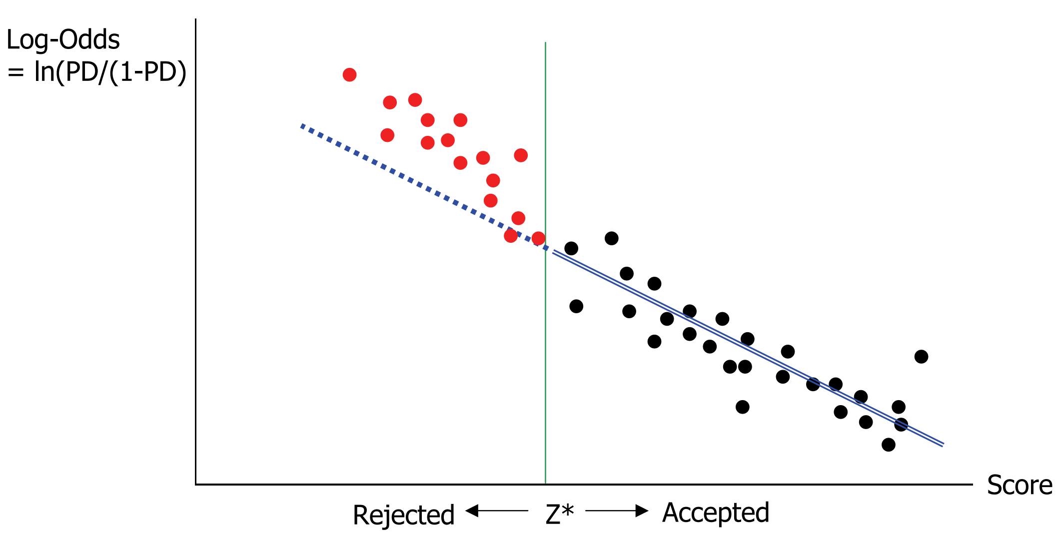
- The concepts of **<u>accuracy</u>** and **<u>precision</u>** can be employed when evaluating rating and scoring model performance at a number of different thresholds.

- A forecast is considered *accurate* if it is “right” on average, i.e. if the predicted outcome on average coincides with the actual outcome. This concept of accuracy is closely related to the *unbiasedness* of a statistical estimator.

- *Precision* is usually defined as the inverse of the standard error (or variance) of an estimator. Less precision is reflected by a larger standard error.

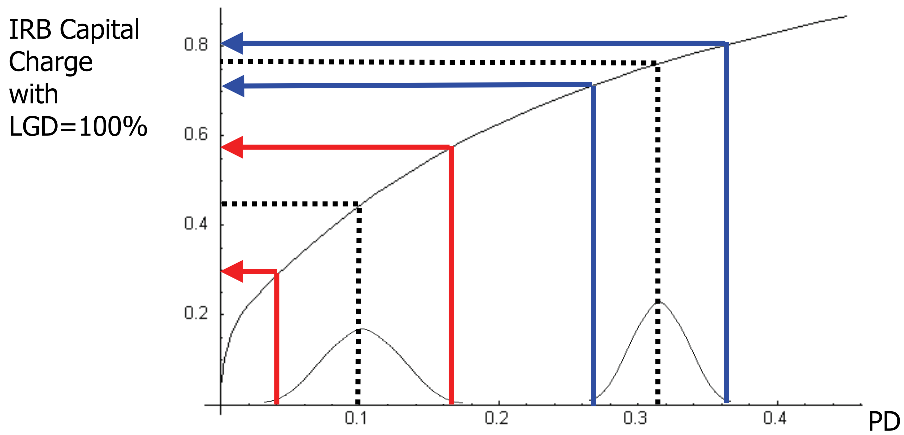
Business Decisions Influenced by Forecast Accuracy

- Reject-inference (prediction of performance for rarely-booked credits)



Business Decisions Influenced by Forecast Precision

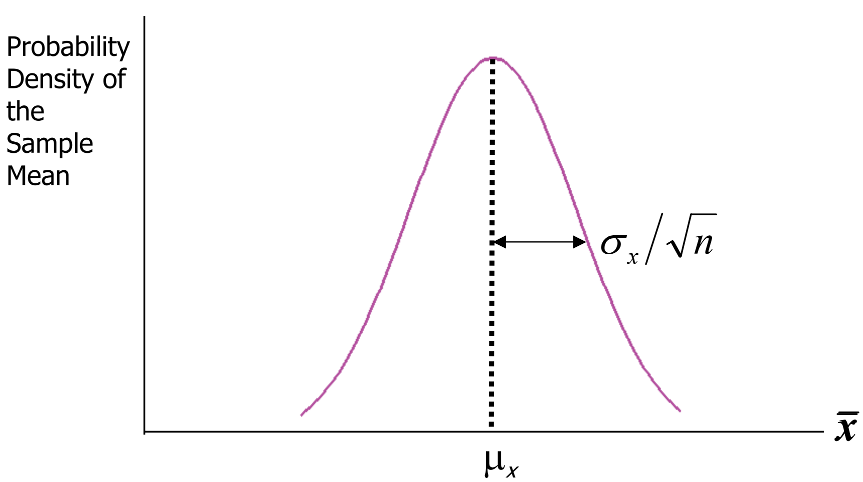
- The variability in capital that could be induced by variability in PD.



Source: Basel II formula for corporates, sovereigns, and banks(BCBS [2005]).

Statistical Evaluation of Accuracy and Precision

- The **Central Limit Theorem** tell us that that when based upon a sufficiently large sample, the sample mean of an estimator, $(\bar{x})$, will be distributed normally around the true population mean (μ_x), with a standard deviation equal to the population standard deviation (σ_x) divided by the square root of the sample size (n).



Evaluating Rating or Score Level Forecasts
The Interval Test

- To examine the accuracy and precision of a PD or LGD forecast for an individual rating grade, we can use the Central Limit Theorem to construct a test of the null hypothesis that the true mean is equal to the predicted value for the grade. We then compare the observed value of PD or LGD with this interval.

- We construct a 95% confidence interval as

Parameter Estimate +/- 1.96*Parameter Standard Error

- When focusing on PD, the standard error can be computed as SquareRoot(PD*(1-PD)/N), where N equals the number of observations in a rating bucket. The interval is computed as ranging from

$$PD - 1.96 \times \sqrt{\frac{PD \times (1-PD)}{N}} \quad \text{to} \quad PD + 1.96 \times \sqrt{\frac{PD \times (1-PD)}{N}}$$

Example
Rated loan portfolio for RMH Bank

RMorrisHamilton

Grade Default Report, December 31, 2002

Period covered by report: 1997 – 2002

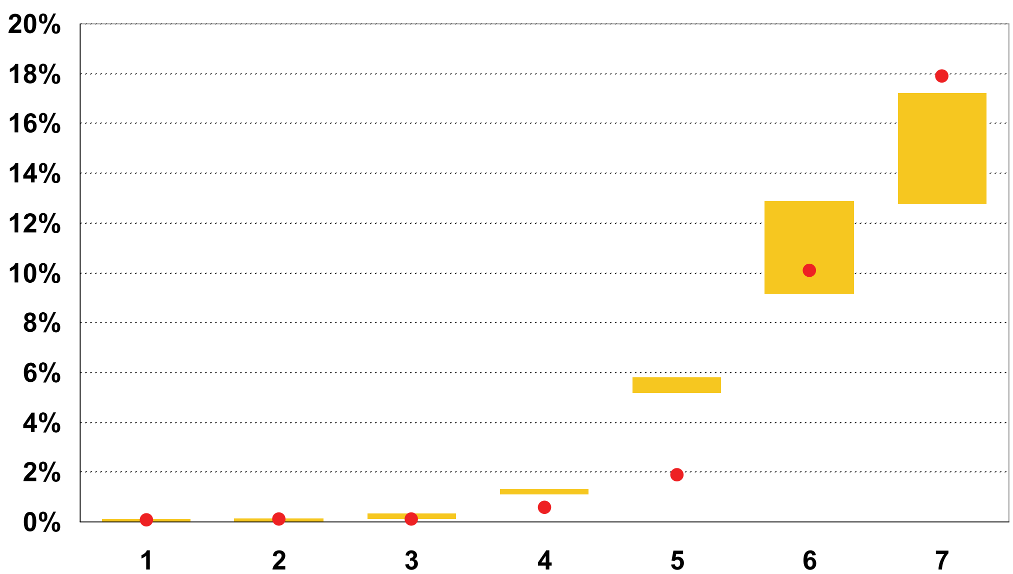
1 Internal Grade	2 Estimated PD (12-31-02)	3 Number of Obligors	4 Portfolio Share	5 Number of Defaults	6 Actual Default Rate
1	0.03%	3660	4.5%	3	0.08%
2	0.05%	5800	7.2%	5	0.09%
3	0.25%	9500	11.8%	10	0.11%
4	1.20%	38200	47.5%	217	0.57%
5	5.50%	21240	26.4%	396	1.86%
6	11.00%	1100	1.4%	111	10.09%
7	15.00%	990	1.2%	177	17.88%
Total		80490	100%	919	1.14%

Example
Interval Tests for RMH Bank's PD Estimates

Rating Grade	Expected Default Rate (PD)	N	Standard Error	Confidence Interval Lower	Confidence Interval Upper	Actual Default Rate
1	0.0003	3660	0.000286	0.000	0.001	0.0008
2	0.0005	5800	0.000294	0.000	0.001	0.0009
3	0.0025	9500	0.000512	0.001	0.004	0.0011
4	0.0120	38200	0.000557	0.011	0.013	0.0057
5	0.0550	21240	0.001564	0.052	0.058	0.0186
6	0.1100	1100	0.009434	0.092	0.128	0.1009
7	0.1500	990	0.011348	0.128	0.172	0.1788

$$PD - 1.96 \times \sqrt{\frac{PD \times (1 - PD)}{N}} \quad \text{to} \quad PD + 1.96 \times \sqrt{\frac{PD \times (1 - PD)}{N}}$$

Example
Interval tests for RMH Bank's PD Estimates



(Bars denote 95% confidence interval around grade PD;
dots are actual realized default rates for each grade.)

Evaluating Forecast Performance Globally

The Chi-Square Test

- The Chi-Square Goodness-of-Fit statistic (Pearson [1900]) can be used to test the null hypothesis that the observed data follow a specified distribution.
- If there are k grades and c=2 states (default and non-default) then we are testing a null hypothesis about k binomial random variables. If the outcomes for each grade are independent, then the joint test will be distributed as a Chi-Square random variable with k degrees of freedom.
- The observed (O) and expected (E) frequencies of default and non-default are compared for each grade, and the statistic is computed as:

$$\chi^2 = \sum_{i=1}^{kc} (O_i - E_i)^2 / E_i$$

Example

The Chi-Square Test for RMH Bank's PD Estimates

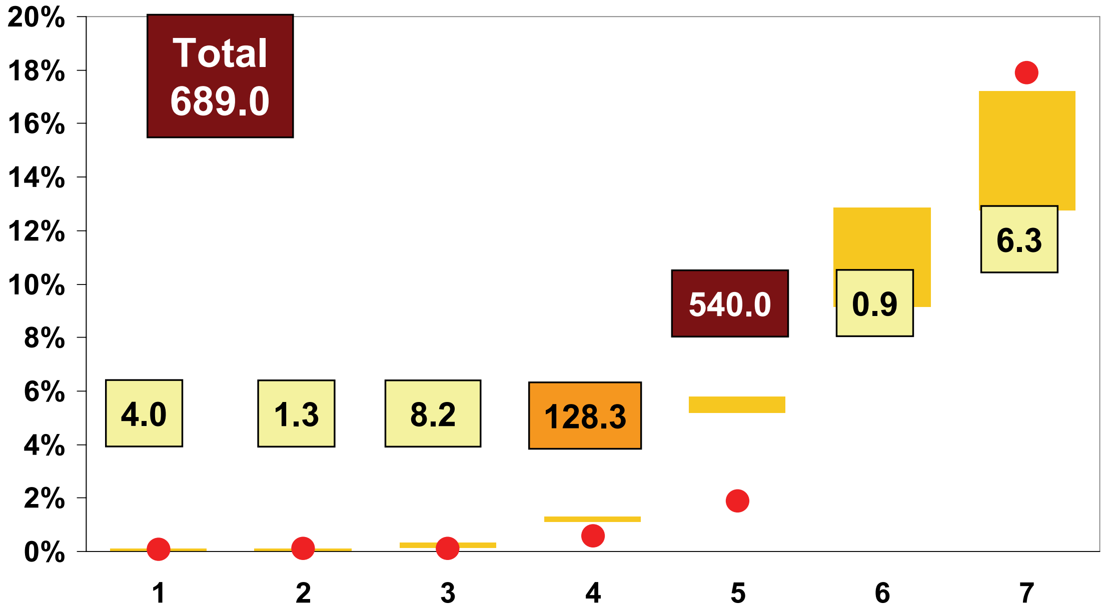
Rating Grade	PD	N	Observed Default	Observed Non-Default	Expected Default	Expected Non-Default	ChiSq Contrib Default	ChiSq Contrib Non-Default
1	0.0003	3660	3	3657	1	3659	4.00	0.00
2	0.0005	5800	5	5795	3	5797	1.33	0.00
3	0.0025	9500	10	9490	24	9476	8.17	0.02
4	0.012	38200	217	37983	458	37742	126.81	1.54
5	0.055	21240	396	20844	1168	20072	510.26	29.69
6	0.11	1100	111	989	121	979	0.83	0.10
7	0.15	990	177	813	149	842	5.26	1.00

Chi-Square Statistic=	689.02
Degrees of Freedom =	7
Prob(Chi-Sq>Critical Val)=	0.00

$$\chi^2 = \sum_{i=1}^{kc} (O_i - E_i)^2 / E_i$$

Example

The Chi-Square Test for RMH Bank's PD Estimates



(Bars denote 95% confidence interval around grade PD; dots are actual realized default rates for each grade.)

Be careful with these tests!

- Default rates are very low for most grades. With such low default rates, need a very large number of loans to achieve desirable levels of statistical confidence.
- The tests assume that defaults in each grade are independent, and they almost certainly are not.
- The tests assume that the "true" default rate is constant, and it almost certainly is not.
- ***The practical implication is that the true 95% confidence bands for the PD estimates are probably wider than derived.***

Testing Global Accuracy
Other Related Tests

The Chi-Squared test's sensitivity to how observations are distributed across the k grades has led to the development of some alternative tests:

- The **Hosmer-Lemeshow Test** (Hosmer and Lemeshow [2000])

  A Chi-Square test where the data is regrouped into deciles rather than k grades

- The **Modified HL Test** (Phibbs, et. al. [1991])

  A Chi-Square test where deciles are defined in terms of the expected number of outcomes, rather than the number of observations in the grades

Measuring Accuracy and Precision
Mean-Squared Error

- Errors are made whenever decisions about an unknown quantity, such as PD, are based upon sample information.
- As we have seen, these errors will generally have two components:
  - some error may be due to **bias or inaccuracy**;
  - some error is due to **random variance or imprecision** arising from use of a sample
- A statistical measure that reflects both the accuracy and precision of an estimator is the Mean Squared Error of the Estimate (MSE):

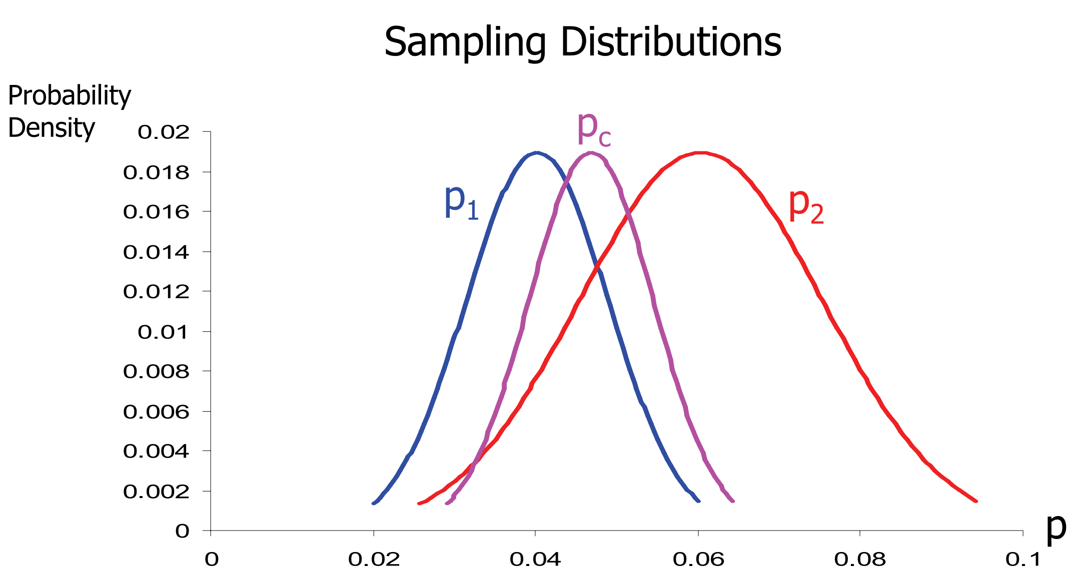
MSE = Variance of the Estimate + Squared Bias of the Estimate

Example

Using MSE to Evaluate PD Rating System Granularity (Kiefer and Larson [2004])

- Consider two different groups of obligors, with respective true (but unobservable) default rates given by θ_1=.04 and and θ_2=.06. We assume that defaults are uncorrelated.
- **We are interested in the question of whether these two groups should or should not be combined for the purposes of estimating default.**
- If we have n_1=500 and n_2=250 obligors from each group, we can we compute the sample default rates p_1 and p_2 to use as estimators of θ_1 and θ_2.
- Alternately, we could pool the sample data and estimate a single combined-group default rate, which we will call p_c.

Example: Using MSE



Example: Using MSE

Estimator	Expected Value	Bias		Variance	MSE =Variance+ Bias^2
		Θ1 =.04	**Θ2** =.06		
p1 (n1=500)	Θ1	0	n.a.	Var(p1) =θ1(1-θ1)/n1 =.0000768	.0000768
p2 (n2=250)	Θ2	n.a.	0	Var(p2) =θ2(1-θ2)/n2 =.0002256	.0002256
Portfolio with two rating buckets	Θ1 and Θ2	0	0	Var(p1)+Var(p2) =.0003024	**.0003024**
pc (Portfolio with one rating bucket)	(n1θ1+ n2θ2) /(n1+n2)	n2(θ2-θ1) /(n1+n2) =.0067	-n1(θ2-θ1) /(n1+n2) =- .0133	Var(pc) =(n1θ1(1-θ1)+ n2θ2(1- θ2)) /(n1+n2)^2 = .0000592	**.0003406**

Since MSE from using the single combined estimate, p_c is greater than the overall MSE from estimating p_1 and p_2 separately, the granularity is warranted from a perspective of minimizing errors in default rate estimation.

Conclusions

- Models can be built to different objectives
- Accuracy and precision are often required by the business use of a model
- Models should be evaluated in how they meet both design and use objectives
- Discriminatory power and forecast performance should both be assessed at the time of development and on a continuing basis subsequent to implementation.

References

- Basel Committee on Banking Supervision (BCBS), *International Convergence of Capital Measurement and Capital Standards: A Revised Framework,* updated November 2005.
- Hand, D.J. [2004] “Good Practice in Retail Scorecard Assessment”, *OCC Credit Rating and Scoring Models Conference*, Washington, DC.
- Hosmer, D.W. and Lemeshow, S. [2000] *Applied Logistic Regression, Second Edition*. New York: John Wiley & Sons, Inc.
- Kiefer, N.M and Larson, C.E. [2004] “Evaluating Design Choices in Economic Capital Modeling: A Loss Function Approach” in *Economic Capital: A Practitioner Guide* (A. Dev, Editor). London: Risk Books.
- Pearson, K. [1900] "On the Criterion that a given System of Deviations from the Probable in the case of a Correlated System of Variables is such that it can be reasonably supposed to have arisen from Random Sampling." *Karl Pearson's Early Statistical Papers*. London: Cambridge University Press, 1956. pg. 339-357.
- Phibbs, C.S., Romano, P.S., Luft, H.S., Brown, B.W., Katz, P.P. [1991] “Improving the fit of logistic models for mortality and other rare events”. *119th Annual Meeting of the American Public Health Association*, Atlanta, GA.
- Thomas, L.C., Edelman, D.B., and Crook, J.N. [2002] *Credit Scoring and it's Applications.* Philadelphia: SIAM

Comptroller of the Currency
Administrator of National Banks

US Department of the Treasury

ENSURING A SAFE AND SOUND NATIONAL BANKING SYSTEM FOR ALL AMERICANS

Session III
Examples of Model Design and Quantification

Mitchell Stengel

Lead Expert – Basel II
Modeling & Analysis
Washington, DC 20219
Mitch.Stengel@occ.treas.gov

2006 Validation of Credit Scoring Models Workshop

Introduction

Case study: corporate rating model

- Intended to assign risk ratings to individual obligors in U.S. corporate portfolio. That is, a classification model, not a (PD) predictive model.
- Through development process – until ready to cut the ribbon.

Introduction

Overview of typical rating / scoring model design and construction process – applies to both wholesale and retail

- Decision: what is the business purpose of the model?
- Data: sample design
- Model specification
  - Choice of variables and formats.
  - Choice of statistical techniques.
  - Qualitative, discretionary, or override factors.
  - Final rating estimates.
  - In-time / out-of-time sample testing

Introduction

Validation processes are appropriate at every stage

- Three stages of model construction above correspond directly to the "developmental evidence" validation processes discussed in earlier presentation:
  - Detailed statement of business purpose.
  - Sample design: selection of dataset that represents target population.
  - Selection of valid and appropriate modeling techniques: expert judgment, statistical methodology, or combination.

Data

KEAL BanCorp., NA

- Sample selected from Compustat
  - 4,861 firms; 72,915 company-years
- 475 defaults from Compustat, bank internal database, and external data sources (such as bankruptcy.com)

Data

Data cleaning and scrubbing

- Deletions from dataset (most important only)
  - Non-commercial / non-industrial firms (by SIC code).
  - Cases of multiple defaults in 3 year period (only one retained).
  - Cases where could not find CUSIP.
  - Cases of major fraud litigation.
  - Cases of firms that declared bankruptcy to avoid large lawsuit pay-outs.
  - Cases of default of parent and sub (only parent retained).

OCC VCRSM Workshop, February 2006

Data Issues

- Low defaults: numbers and rates*
- Missing data: can fill in sometimes
- Use of external data sources: mapping
- Internal data: sample design, selection of variables, bank information systems
- Combined cross-section / time-series

* Basel Committee on Banking Supervision, Newsletter No. 6 (September 2005), "Validation of low-default portfolios in the Basel II Framework."

Choice of Variables

- Starting point: variables used for this particular type of model in past by bank or others
- Typically financial ratios
  - Large number to choose from.
  - Often alternative definitions.
- Begin with univariate analysis
  - Correlations of individual variables with defaults.

OCC VCRSM Workshop, February 2006

Choice of Variables

LOT of trial and error

- Criteria for selection
  - Sets of variables with best discriminatory power.
  - Parsimony: minimize multicollinearity and avoid overfitting.
  - Minimize number of default observations with missing data.
  - Where there are multiple definitions of a ratio, choose simplest one.
  - Expert judgment by model builders and / or field staff is often necessary.

Choice of Variables

KEAL BanCorp.

- Over 50 ratios to choose from.
- Using processes and criteria outlined above, after extensive testing, arrived at final list:
  1. Liquidity (working capital / total assets)
  2. Leverage (total liabilities / total assets)
  3. ROE
  4. Interest coverage (net operating income + income tax + interest expense / interest expense)
  5. Total debt / total capital (including rentals and capitalized leases)
  6. Firm size (Ln(Assets))
- All testing, results, criteria, and final choices should be fully documented.

Model Specification

- Observation window: 12 months. Model based on relationship between independent variables (ratios and size) in year ending December 31 and outcome (default or non-default) during the following 12 months.
- Censoring of ratio outliers: pro and con.
- Segmentation by industry grouping
  vs. single national model.
- Format of financial ratios: transformed (e.g., log, "binned," or ranked) or untransformed.

Modeling Techniques

KEAL considered large number of techniques, including OLS, ordered probit, decision tree (CHAID), and logit (both standard and nested).

- Different techniques entail different dependent variables and in some cases would require different independent variables in the sample dataset.
  - Ordered probit and CHAID can directly estimate the risk-rating category or bucket for individual corporate exposures.
  - OLS can estimate the score (log odds) based on dichotomous (0,1) outcomes data; or the risk-rating category.
  - Logit can estimate the score (log odds) based on dichotomous (0,1) outcomes data.

Choice of Modeling Technique

After initial testing, bank narrowed choices to two: nested logit, which can capture non-linear relationships, and standard logit (both with untransformed ratios).

- Results:
  - Both separated defaults from non-defaults effectively.
  - Based on obligor risk rating system with 10 grades of equal score width, standard logit produced more reasonable and appropriate distribution across ratings. (See Figure 1.)
  - Standard logit had slightly better CAP curve (see Figure 2) and Accuracy Ratio (85.7 vs. 81.2).
  - Although nested logit captured non-linear relationships, it was more difficult to interpret, and coefficients and outcomes can be statistically unstable.
  - Therefore, KEAL chose the standard logit as its final rating model. (See Figure 3.)

Risk Rating Distributions

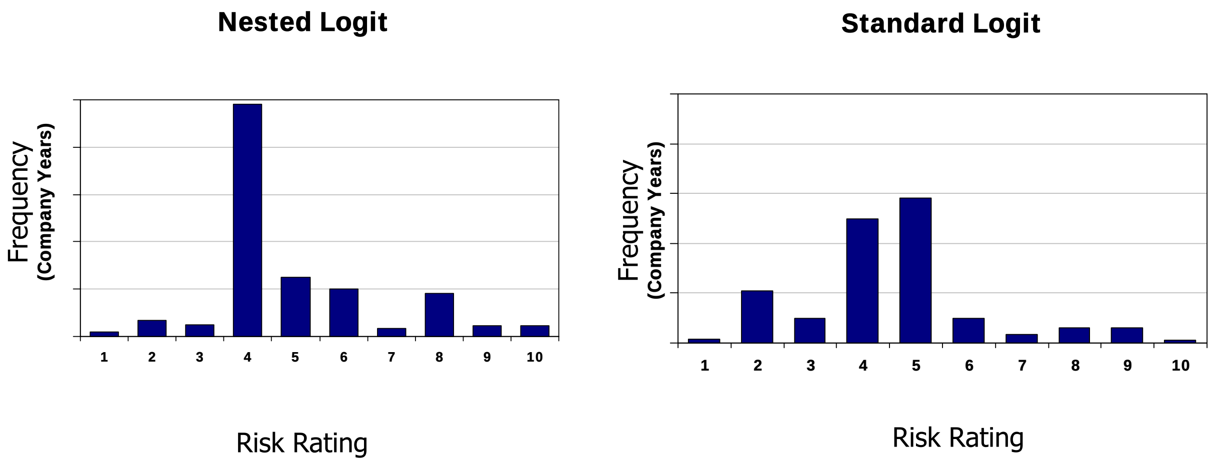


Figure 1

Results of Nested and Standard Logit Models

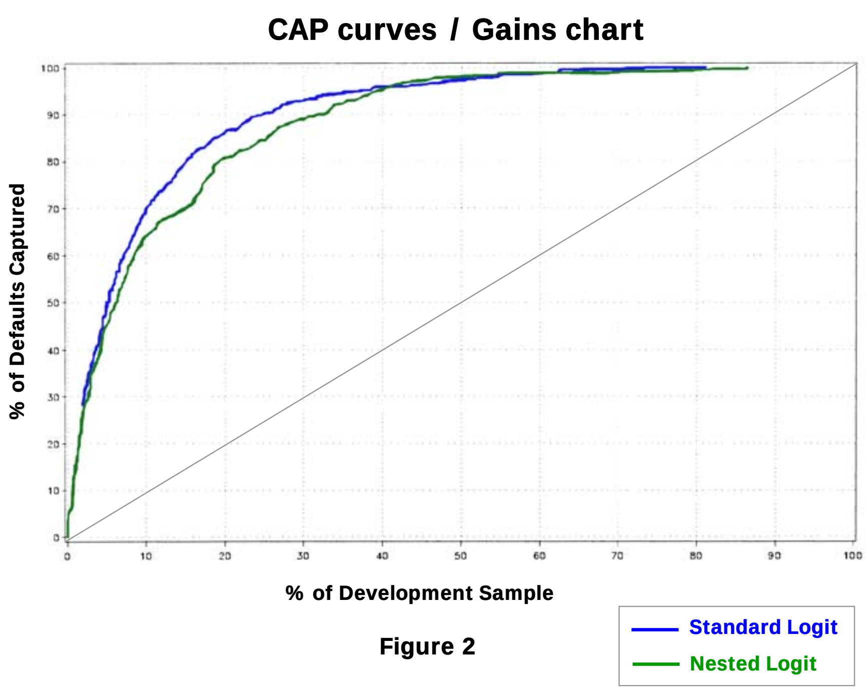


Figure 2

Final Risk-rating Model

KEAL Bancorp.

Variable	Coefficient	p-value Pr > Chi Sq
Intercept	- 4.501	<.0001
Liquidity	- 2.011	<.0001
Leverage	7.922	<.0001
ROE	- 0.093	<.0001
Interest Coverage	- 0.015	<.0001
Total debt / total capital	0.158	<.0001
Ln (assets)	- 0.106	<.0001

Figure 3

Qualitative Factors

Four general questions answered by line of business or risk managers, used to adjust, or supplement, results of scoring model.

- Each question can be answered Weak (-0.125 points); Neutral / Average (0); or Strong (+0.125).
- Point total (-0.5 to +0.5) added to score.

Qualitative Factors

Questions:

1. Regulation / supervision: Intensity of government supervision; prospects for added burden or deregulation.
2. Industry characteristics: Growth prospects (short- and long-term); vulnerability to natural disasters or cut-offs of supply (e.g., OPEC).
3. Managerial factors: Number of layers; encouragement of or opposition to innovation; succession planning.
4. Competition / concentration in industry, among suppliers and among customers.

- Loan officers and risk managers work with model developers. Based on their experience in the lending process, they play a significant role in choice of variables and qualitative questions.

Validation Issues: Modeling Techniques

- All techniques, model estimates, and results should be fully documented.
  - Bank provided CAP curves, Accuracy Ratios, and distributions by risk ratings for 2 "finalists,"
    - but no K-S statistics or divergence indices for the individual models.
  - Bank provided no testing or diagnostics at all for the "final" model including the scores as modified by the qualitative questions.

In-time / Out-of-time Sample Testing

Bank chose random sample that was not used in the development model.

- 1,739 firms; 26,085 company years; 170 defaults (all before data cleaning and scrubbing).
- Ran final model on this sample.
  - Reported Accuracy Ratio of 79.3 (vs. 86.7 for development sample).

In-time / Out-of-time Sample Testing

Validation issues:

- What does difference of 7.4 in AR mean? Bank has set no thresholds, no margins to trigger any particular processes (such as model review).
- Bank should report full results and diagnostics for out-of-time sample, to permit thorough cross-validation and analysis of indications of possible overidentification and/or misspecification.
  - Comparison of all individual coefficients (magnitude, sign, and significance).
  - Risk-rating distribution.
  - CAP curve.
  - K-S statistic and / or divergence indices.

One Last Step in Model Design and Building

Putting model into production

- Hand-off from model development / validation team to IT production.

- Hand-off from Developmental Evidence validation processes to Process Verification and Benchmarking.

- For both of those transitions:
  - Critical importance of documentation, transparency.
  - User training.

Conclusions

Validation is a central aspect of model development.

- Should be integral part of every stage.
- Should be planned from day one as part of design process.
- Not something you can put off thinking about until model is almost ready to roll.

- Despite differences in details and terminology, there are fundamental similarities between wholesale and retail in model design and validation.

References

Basel Committee on Banking Supervision, Newsletter No. 6 (September 2005), "Validation of low-default portfolios in the Basel II Framework."

Session IV
Process Verification and Data Maintenance

Mike Carhill

Director
Risk Analysis Division
Washington, DC 20219
Michael.Carhill@occ.treas.gov

2006 Validation of Credit
Scoring Models Workshop

Model Validation *re* OCC Bulletin 2000-16

- Abstract computer models have three components.
  - Inputs
  - Processing
  - Output

OCC VCRSM Workshop, February 2006

Model Validation *re* OCC Bulletin 2000-16

- Abstract computer models have three components.
  - **Inputs**
  - **Processing**
  - Output

Model Validation *re* OCC Bulletin 2000-16

- Each of which is validated by respecting these general principals:
  - Independence
  - Documentation
  - Cost versus benefits

Inputs

- Output from other models
- Internal raw data
- External raw data
- Constructed variables

Model Outputs as Inputs

- Output from another internal model
  - Which itself is validated according to these principles
  - Ongoing forecast-versus-actual comparisons
- Output from vendor models
  - Ongoing forecast-versus-actual comparisons
    - Generic bureau scores applied to bank's portfolio
  - Vendor documents its own validation process

Internal Data

- Reconciliation to general ledger or other MIS
  - Policy should specify error tolerances
  - Record of variances
  - Usually a strength of internal audit departments
- Test for accuracy of fields
  - Transaction testing

External Raw Data

- "Raw" observation from an external source (e.g., quarterly income growth)
  - Documentation:
    - User's guide for accessing the data
    - Rationale for choice of source
    - Caveats as to accuracy
    - Any tweaks done to the variable

Constructed Variables

- Variables formed from raw data via simple definitions
  - Modelers should maintain data dictionary
    - Many possible definitions of "leverage"
  - Most external "raw" data is actually constructed data
    - Care should be taken to ensure that use of variable is consistent with definition

Processing

- Coding
- Theory

Coding

- Simple models
  - Independent and Identical Construction (IIC)
    - Cheap
    - Should produce identical results
- IIC not practicable for complex models
  - Too expensive
  - Would never get identical results, anyway
- For gray areas independent inspection of code can work
  - But far from fool-proof

Validating Code in Complex Models

- Inspection
  - Probably won't work
  - Staff retention problematic

Validating Code in Complex Models

- Documentation
  - Internal code documentation
  - External technical documentation should cover interrelationships between modules, flow charts and "pseudo code"
    - Change control and documentation
  - Meet the test: Could an entirely new team use existing model to continue development or production?

Validating Code in Complex Models

- Comparison to other models
- Convergence to market
- Ongoing forecast-versus-actual comparison

Validating Theory

- Comparison to other models
- Convergence to market
- Ongoing forecast-versus-actual comparison

Validating Theory

- Documentation:
  - Reference to literature
  - Document internal applications and any innovations
  - Precise specification of question being answered

Conclusions

- Inputs and processing are the "perfectly" part of RAD's mantra "all models should be perfectly wrong."
- While the intellectual firepower goes to validating output, most of the expense goes to validating inputs and processing.

Comptroller of the Currency
Administrator of National Banks

US Department of the Treasury

ENSURING A SAFE AND SOUND
NATIONAL BANKING SYSTEM
FOR ALL AMERICANS

Session V
Monitoring and Benchmarking

M. Nazmul Hasan

Lead Expert – Credit Risk Modeling
Risk Analysis Division
Washington, DC 20219
Mohammad.Hasan@occ.treas.gov

2006 Validation of Credit
Scoring Models Workshop

Introduction

A comprehensive validation process requires:

- Evaluation of developmental evidence
- Analysis of outcomes
- Process verification
- **Ongoing monitoring and benchmarking**

Outline

- Motivation
- Monitoring and Benchmarking Tools
  - Front-end analysis of the score distribution
  - Back-end analysis of the performance measures
- Analysis of Risk Characteristics (Drivers)

Motivation: When does a model fail?

- A model may fail when
  - Credit profile of the current portfolio changes significantly from the development sample
  - Weights of risk characteristics to performance measure of the model changes
- Factors contributing to a change in portfolio credit profile or risk weights of individual characteristics
  - Poor pricing (adverse selection)
  - Change in underwriting standards
  - Change in business strategy
  - Change in macroeconomic conditions

Motivation: What can we do to reduce model risk?

- Cannot wait for backtesting results
  - Long time lag between developmental sample and validation sample for backtesting
- Assess model risk by close monitoring and benchmarking
  - Front-end analysis
  - Back-end analysis
- Perform characteristic analysis to explain the deviations from benchmark analysis

Monitoring and Benchmarking

Are they separate processes?

- Effective ongoing monitoring almost always involves benchmarking. Although they may appear as two distinct and independent processes they are closely linked. The most common benchmarks are
  - Development sample
  - Alternative models (cross-validation)
    - Internal models
    - Vendor models
    - Rating agencies
    - Peer institutions

Monitoring and Benchmarking

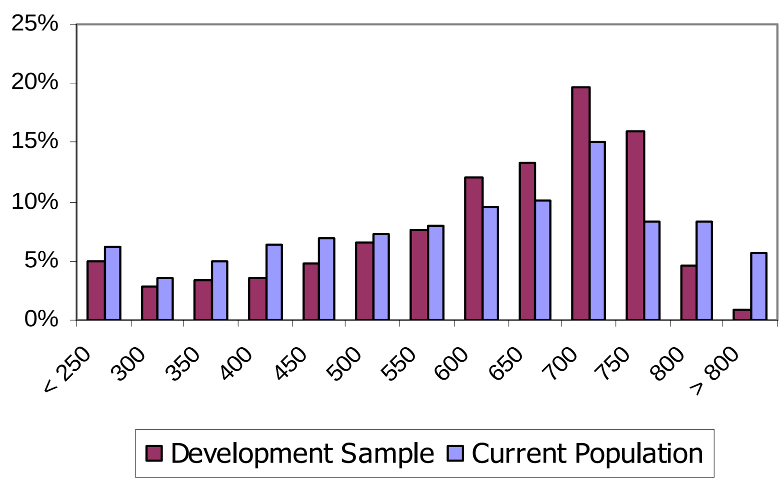
- Non-outcomes based evaluation: Front-end analysis of the score distribution
  - Population stability of the score distribution of the current portfolio (benchmarking to the development sample)
  - Ongoing comparison of the score distributions generated by competitive models (benchmarking to alternative models)

Monitoring and Benchmarking

- Outcomes based evaluation: Back-end analysis of the performance measures
  - Cross validation (Champion/Challenger: benchmarking to alternative models)
    - on a common reference data set at development
    - on the current portfolio
  - Trend analysis (benchmarking to development sample)
    - on different vintages/cohorts

Front-end Analysis
Population Stability: Score Distribution

- Current population is attracting a lot of risky customers
- We can investigate it in terms of borrower characteristics

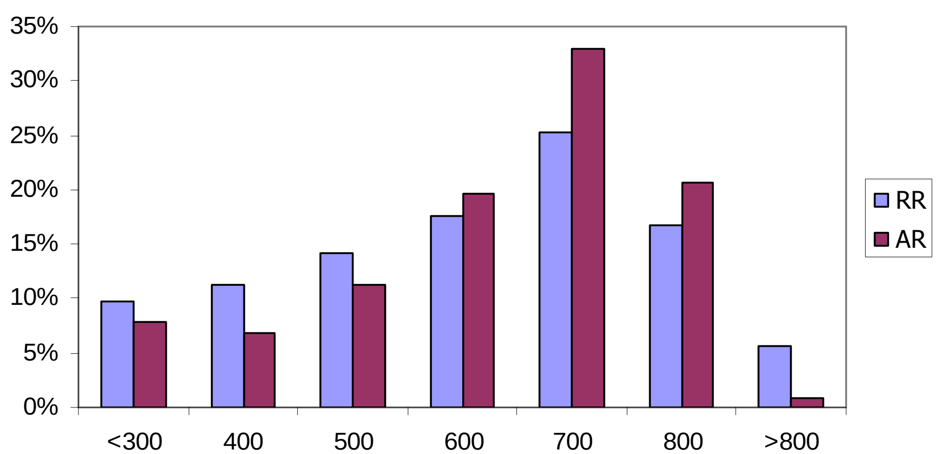


Front-end Analysis
Measures of Separation

- Various measures of separation are available:
  - Divergence index
  - K-S Statistic
  - ROC and Gini coefficient
  - Pearson's Chi-square test
- No single test is statistically powerful and robust enough to be sufficient. So apply multiple tests to confirm separation
- Create longitudinal reports to separate the transitory versus permanent shifts

Front-end Analysis: Competing Models
Score or Rating Distributions

Two Risk Rating Systems: RR and AR



RR1	RR2	RR3	RR4	RR5	RR6	RR7
AR1	AR2	AR3	AR4	AR5	AR6	AR7

Front-end Analysis: Competing Models Rating Distributions

- Analyze the off-diagonal elements to understand the differences in the models

	AR1	AR2	AR3	AR4	AR5	AR6	AR7
RR1							
RR2							
RR3							
RR4							
RR5							
RR6							
RR7							

Front-end Analysis: Competing Models Score or Rating Differences

- Effective benchmarking against alternative models requires a good understanding of differences in modeling methodology
  - Time horizon over which the risk is assessed
  - Differences in bad definition
  - Risk characteristics used in the models
  - Alternative risk measures PD versus EL (e.g. rating models)
  - Statistical methodology employed to estimate the models

Back-end Analysis
Cross Validation: Objective

- Cross-validation has much broader use. For example, it helps
  - Choose the best model by comparing the reliability and accuracy of the models
  - Assess if the internal ratings are punitive or overly optimistic
  - Identify process inefficiency through ongoing comparisons

Back-end Analysis:
Cross Validation (Champion/Challenger)

- Internal models based on alternative methodology
  - Scoring models built upon different statistical techniques (e.g. Logistic vs. Neural Network)
  - Rating models based upon different theoretical frameworks (e.g. Reduced form vs. Structural)
- Internal models vs. vendor models
  - Internal credit scoring vs. FICO model (retail)
  - Internal rating model vs. RiskCalc (middle market)
  - Internal rating model vs. MKMV EDF implied rating (large public corporate)
  - Internal models vs. rating agency

Back-end: Trend Analysis

- Provides a dynamic view of the changing portfolio when compared against the development sample

Vintage curve analysis

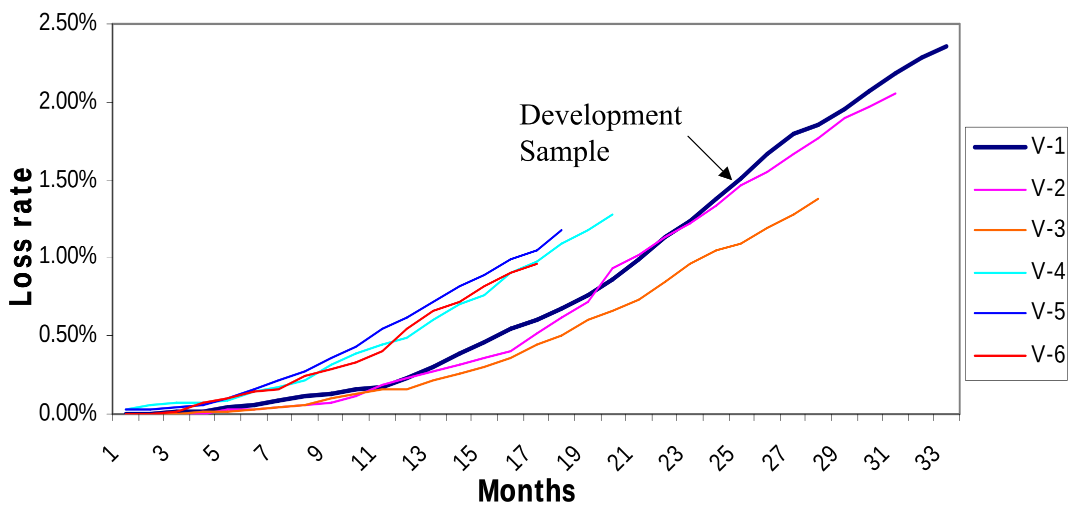
- Borrowers are fixed over time
- Vintage-specific delinquency curves that track the *cumulative* bad rate over time for each vintage
- Vintage curves by score band against some performance measure -- provide a more dynamic benchmark for backtesting the models

Portfolio trend analysis

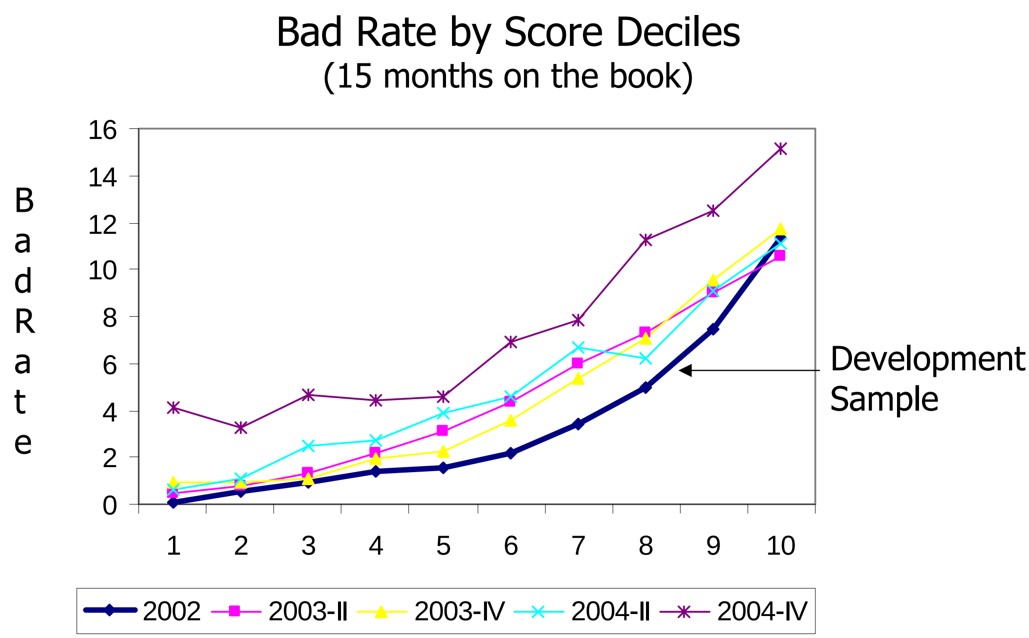
- Borrowers are changing over time
- Provides a dynamic view of the entire portfolio

Vintage Curve Analysis: Vintage Specific Cumulative Loss Curve

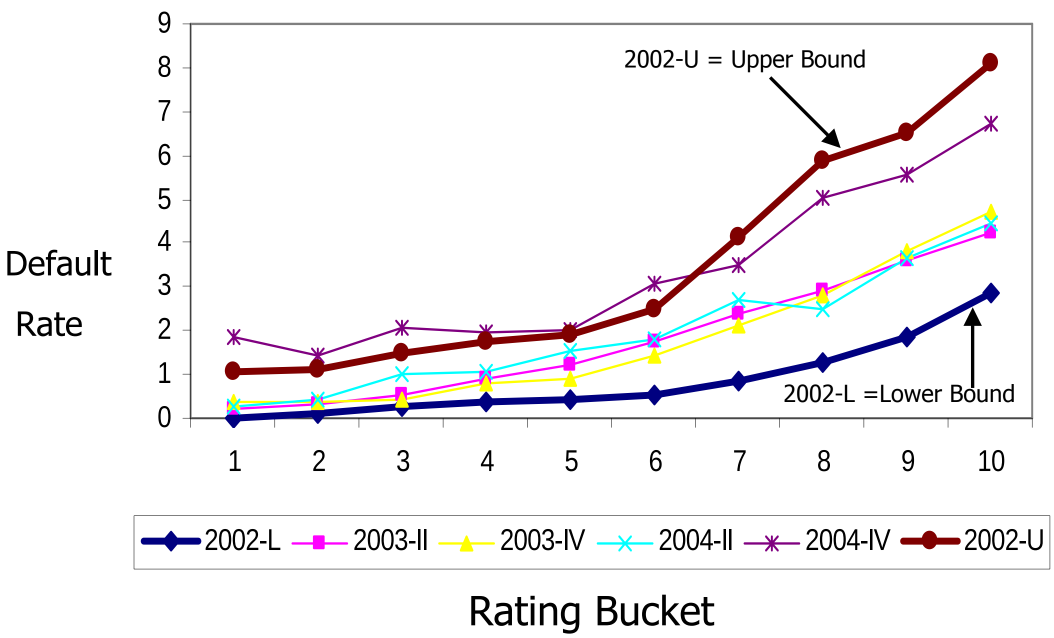
Tracks the cumulative loss rate over time



Vintage Curve Analysis: Dynamic Benchmarking for Back Testing



Back End: Portfolio Trend Analysis

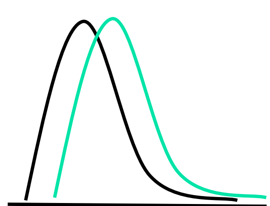


Analysis of Risk Characteristics (Drivers)

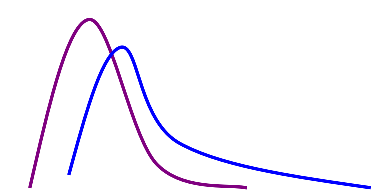
- Isolate the reasons for instability or deteriorating performance of the model
  - Is there any shift in the distribution of a risk characteristic?
  - Analyze how the change in distribution affects the score of a borrower on average
  - If performance data are available, assess the predictive or discriminating power of characteristics included or excluded from the model

Analysis of Characteristics

- Changes in characteristics reflect changes in the distribution of borrower attributes
- The distribution may change due to change in
  - Location parameters: mean, median, or mode
  - Shape parameters: variance, skewness, etc.



Location Shift



Shift in Shape Parameters

Analysis of Characteristics Consequences: Shift in Distribution

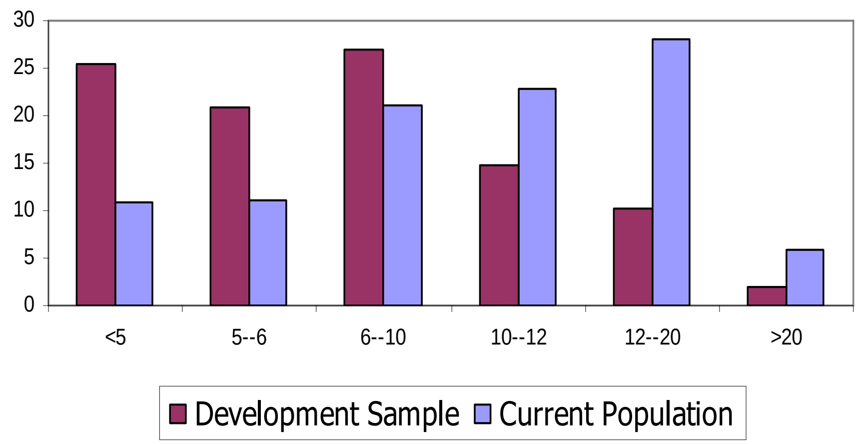
- Location shift
  - In a regression context, location shift affects only the intercept parameter, and the relationship between the attribute and log-odds remains unchanged
  - Rank-ordering remains stable, with similar magnitude of inflation or deflation of log-odds for all borrowers
  - Cut-off points may need to be adjusted

Analysis of Characteristics Consequences: Shift in Distribution

- Shift in shape parameters
  - Affects both intercept and slope parameters
  - Rank-ordering as well as accuracy will be affected
  - Unlike location shift, no easy fix to cut-off strategy without rebuilding the model or making some serious adjustment to scorecard calibration of score-to-odds relationship

Analysis of Characteristics
An Example: Debt Service to Income

- Compare the percentage of the most recent accounts that fall within the same attribute category as those of the development sample



Analysis of Characteristics
An Example: Debt Service to Income

Attributes	Development (%)	Current (%)	Difference	Score	Weighted Difference
Below 5%	25.40	7.30	-18.10	83.00	-15.02
5 -- 6%	20.80	11.10	-9.70	73.00	-7.08
6 -- 10%	26.90	21.10	-5.80	65.00	-3.77
10 -- 12%	14.70	22.90	8.20	55.00	4.51
12 -- 20%	10.20	28.10	17.90	51.00	9.13
Over 20%	1.96	5.90	3.94	48.00	1.89
Missing	0.04	3.60	3.56	65.00	2.31
			Total Change in Points		-8.03

What does this 8 point drop mean?

If the scorecard is calibrated so that odds double for every 20 points and the initial average odds is 20:1 (bad rate 5%), then an 8 point drop will lead to a rise in the bad rate to almost 6.4%

Analysis of Characteristics: Predictive or Discriminating Power of Characteristics

- Measures of predictive or discriminating power, e.g.
  - Chi-square statistic
  - Information statistic
  - Somer's D concordance statistic
- Analysis may reveal that
  - The relationship of the attributes of a characteristic to the score-weight may need to change
  - Characteristics excluded from the model are more predictive or discriminatory than those included
  - The predictive or discriminatory power of the model in production is deteriorating relative to alternative models

Conclusions

- Monitoring and benchmarking are closely linked processes
- An effective monitoring-benchmarking process requires:
  - Continuous assessment of borrowers' characteristics in development sample versus current portfolio
  - Trend analysis of various performance metrics
  - Comparison against alternative models
  - Application of a variety of quantitative and statistical tools

Session VI
Validation as a Control Function Under Basel II

Mark Levonian

Deputy Comptroller
Modeling & Analysis
Washington, DC 20219
Mark.Levonian@occ.treas.gov

2006 Validation of Credit Scoring Models Workshop

Outline

- Basel II emphasizes validation
- Credit risk under Basel's IRB approach
- Validation and other control processes for IRB
- A validation example: LGD for Basel II
- New challenges likely require new tools

The Basel Connection

- Basel II has enhanced interest in validation
- Basel Committee's Accord Implementation Group (AIG) has established a validation subgroup, which has published validation principles.
  (Basel Committee Newsletter No. 4, January 2005)
- Basel II brings a new focus
  Aspects previously regarded as arcane quantitative issues may become central concerns for both bank management and bank supervisors

Basel II on Validation for Credit Risk

- Basel framework includes specific language requiring validation:

500. Banks must have a robust system in place to validate the accuracy and consistency of rating systems, processes, and the estimation of all relevant risk components. A bank must demonstrate to its supervisor that the internal validation process enables it to assess the performance of internal rating and risk estimation systems consistently and meaningfully.

(Source: Basel Committee, November 2005, page 105)

Credit Risk Under Basel II

- Broad outlines of credit risk under Basel II likely are familiar by now
- Under the Internal Ratings-Based (IRB) approach, banks must:
  - Differentiate obligors and exposures according to credit risk
  - Quantify credit risk for obligors and exposures within a particular modeling framework

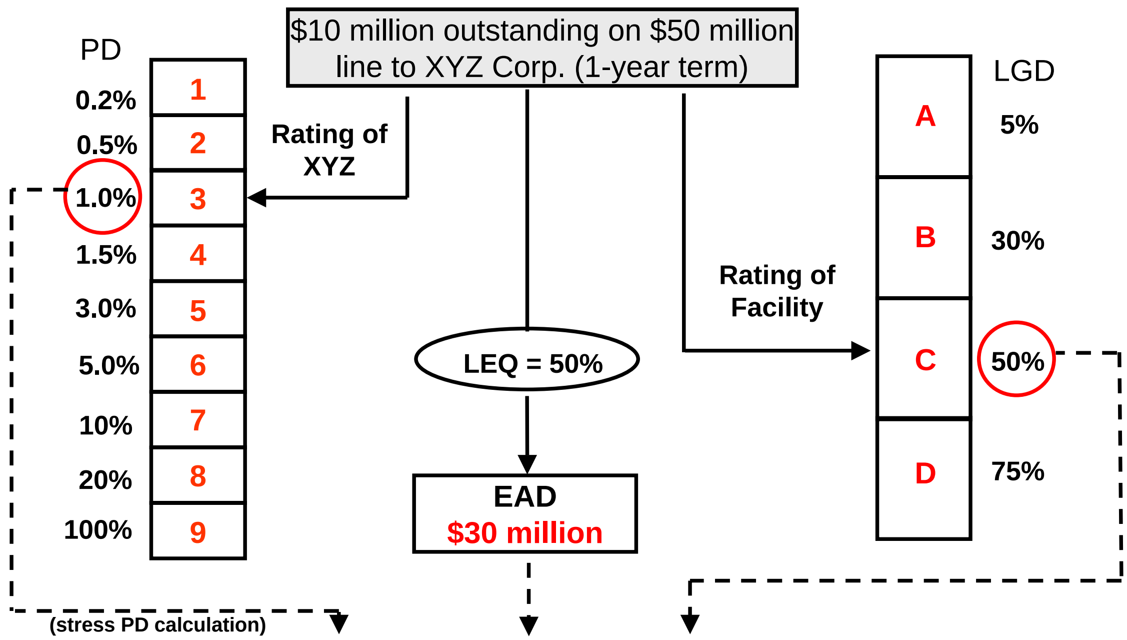
Risk Differentiation for IRB

- Banks are required to assign exposures to groupings with roughly homogeneous risk
  - Obligor ratings linked to default frequency
  - Severity grades linked to default losses
  - Segmentation for retail exposures
- Traditional credit rating and scoring methods may be used, or "models" may be less explicit (for example, ratings assigned using expert judgment)

Risk Quantification for IRB

- Banks estimate certain parameters of the credit risk model Basel II uses for capital calculations
  - PD: probability of default
  - LGD: loss given default
  - EAD: exposure at default
  - M: effective maturity (for wholesale exposures)
- Parameter estimates are assigned to grades, segments, or exposures as relevant

IRB for Wholesale Credit



(13%) x ($30 million) x (50%) = **$2.0 million capital**

Example: Exposures for Large Bank

		LGD							
	PD	**10%**	**15%**	**20%**	**23%**	**30%**	**33%**	**34%**	**38%**
1	**0.03%**	3,048	865	1,258	0	0	0	0	8
2	**0.05%**	31	207	2	0	12	40	69	247
3	**0.29%**	43	179	0	0	115	4	0	913
4	**0.32%**	70	578	0	0	51	224	0	1,144
5	**0.77%**	59	539	48	75	60	1,002	0	2,187
6	**7.77%**	17	81	0	6	10	1,500	0	511
7	**19.74%**	2	83	0	16	2	370	0	166
8	**27.17%**	1	42	0	0	4	436	0	271

Corporate-Bank-Sovereign exposure (in $ millions)

Control Processes for IRB

- Integrity of internal risk estimates must be ensured through adequate governance around processes
  - Formalized, approved policies and procedures
  - Independent review
  - Effective internal audit
  - Incentives inherent in the system
  - Documentation and transparency
- Validation is another element of the control environment
  - Quantitative nature of IRB may make validation a particularly important control
- The control environment, including validation, should be viewed as a whole

Validation in the Basel Context

- Recall key elements of validation from earlier talks
  - Developmental evidence
  - Ongoing monitoring, process verification, and benchmarking
  - Analysis of outcomes
- For Basel II, the specifics of validation may change, but concepts or principles remain the same
- Where is validation needed?
  - Explicit models may be used to differentiate and quantify risk
  - But there are also "models" in a broad sense: transforming information as input into output for making a decision
    - These "models" may not be captured in computer code

Validating Risk Groupings

- Assignment of obligors and exposures to internal rating grades or segments must be validated
  - Methods span a spectrum from explicit, statistically based quantitative scores to judgmental approaches
  - Homogeneous risk within groupings is crucial
    - Models used may be designed to rank-order, but this might not be the most important feature for IRB rating assignments
- Validation elements in this context include:
  - Developmental evidence for the risk grading system
  - Benchmarking in the form of comparison to alternatives
  - Process verification through transaction testing
  - Ex-post analysis of credit outcomes

The Relevance of Rating Philosophy

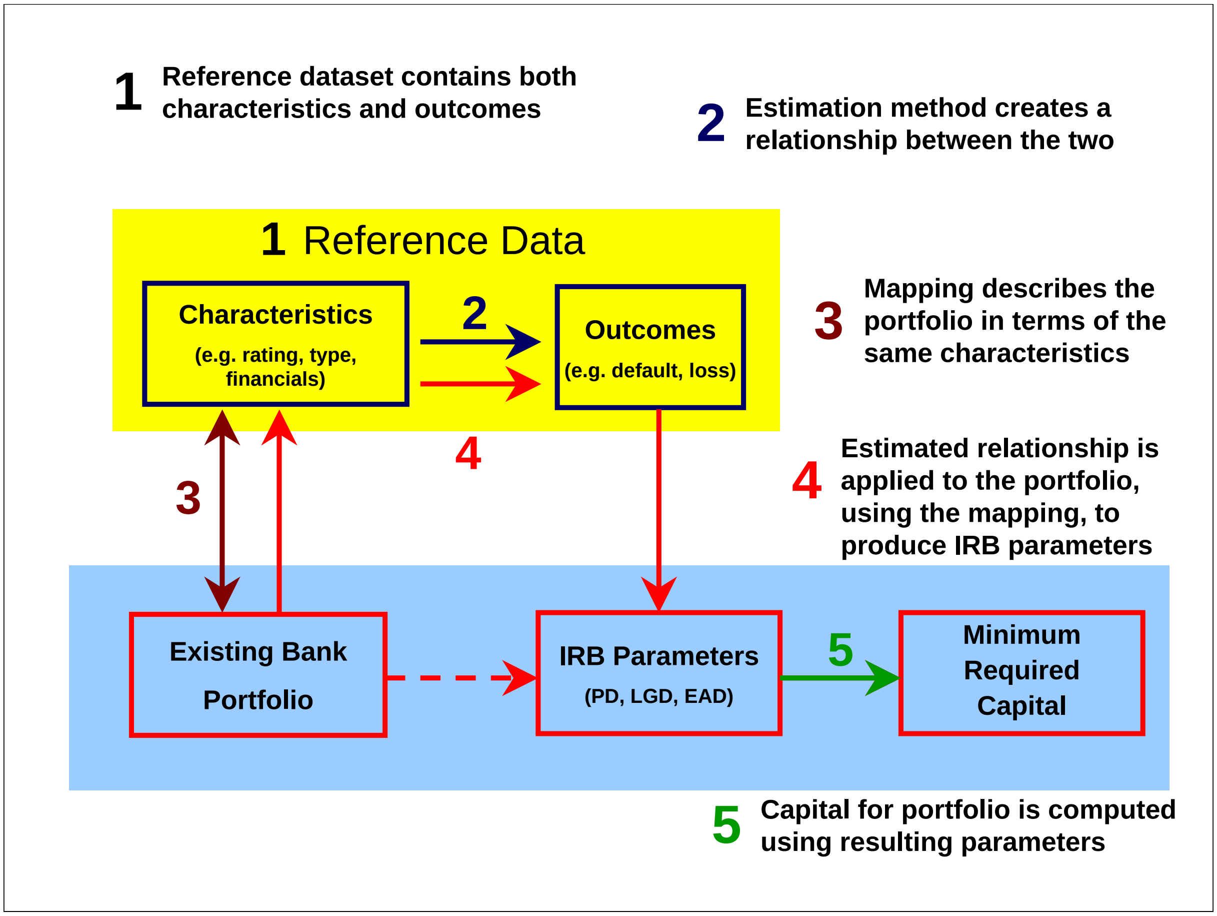
- Different rating systems aim to reflect cyclical or systematic effects in different ways
  - Primarily an issue in corporate credit
  - Commonly discussed in terms of "point-in-time" and "through-the-cycle" (whatever those mean…)
- Differences in "philosophy" have implications for validation of IRB systems
  - Philosophy or approach affects interpretation of outcomes analysis for risk-grading systems
  - Estimation and mapping must take into account possible differences between a bank's current approach and the philosophy embedded in reference data

Risk Quantification: The Big Picture

- After homogeneous risk groups are identified, risk must be quantified, and quantification also must be validated
- Details of quantification vary between retail and wholesale, and across parameters (PD, LGD, EAD)
- However, all practical approaches to quantification include identifiable conceptual steps or stages
- Each stage can and should be subject to validation

Risk Quantification: Four Stages

- **Reference Data**: a dataset with known outcomes, and information on characteristics related to risk
  - In some settings this is called a "developmental sample"
- **Estimation**: methods that relate observed outcomes to the characteristic variables in the reference data
- **Mapping**: a process to link observable features of obligors or exposures in the existing portfolio to similar variables used in the estimation
- **Application**: use the established mapping to apply the estimates to the existing portfolio



Validation for IRB Quantification

	Developmental Evidence	Process Verification and Benchmarking	Outcomes Analysis
Reference Data	X	X	
Estimation	X	X	?
Mapping	X	X	?
Application	X	X	X

Example: LGD Quantification

- Bank has internal data on all defaulted loans, with timing and amounts of recoveries, back to 1996 (net of workout costs)
- For each loan, data include collateral type (e.g. real estate, inventories, cash), and collateral coverage as "high, medium, low"
- Apply discount rate to value recoveries, then estimate LGD from average recovery rate, for each of 12 combinations

LGD (percentage of EAD lost in default)			
	Low coverage	**Medium coverage**	**High coverage**
Collateral Type 1	**25**	**12**	**5**
Collateral Type 2	**5**	**3**	**1**
Collateral Type 3	**16**	**16**	**14**
Collateral Type 4	**90**	**40**	**25**

Example (continued)

- Bank has more detailed information on collateral types and coverage for the exposures in its existing portfolio, but divides the portfolio into 12 categories to match the available reference data
- Any exposure with multiple types of collateral receives an average of the LGD values for those collateral types
- All LGD estimates adjusted upward by 10% to account for "benign environment" represented in reference data

Dissecting the Example

- Internal risk-rating system for loss severity, based on established criteria related to loss rates
- Reference data set of internal defaults, with some observable risk-related characteristics (collateral)
- Estimation is simple averaging within categories
- Mapping requires determination of relationship between collateral information for existing portfolio and less-detailed information in reference data
- Application stage involves some adjustments for special cases (multiple collateral types) and conservatism

Example: Illustrative Validation Questions

	Developmental Evidence	Process Verification and Benchmarking	Outcomes Analysis
Reference Data	Was there available information that was excluded from the reference data set?	How does the discount rate compare to what others use?	
Estimation	Did the bank consider other factors that might affect losses?	How do these LGDs compare to other available estimates?	How do realized loss rates compare to LGD estimates?
Mapping	How did the bank establish the relationship for the collateral variable?	Does the approach resemble current sound practice?	
Application	How did the bank determine that 10% was an appropriate adjustment?	How does the 10% adjustment compare to other banks' practices?	Is there evidence that the adjustment accomplished its objectives?

Basel Validation: New Tools Needed

- From the LGD example:
  - Outcomes analysis when distribution is multimodal
  - Benchmarking when workout practices differ across banks
- Challenges for assessing PD
  - Small samples, small probabilities
  - Statistical tests can be difficult if default rates vary over time
- Requirement to validate *all* parts of the process
  - For risk quantification, validation can be organized around the four "stages" discussed above
- Likely need for better data – data have not necessarily been collected in the form now needed

OCC VCRSM Workshop, February 2006

Prominent Basel Validation Issues

- Dialogue among regulators and with industry representatives highlights a number of issues
  - Expectations for validation of vendor models used for IRB
  - Expectations for independence in validation or other aspects of IRB
  - Validation of "low-default" portfolios
  - Expectations for "conservatism" in various areas and the impact on validation
- These and many other issues are the subject of continuing work and development
  - Validation for so-called "low-default" portfolios is discussed in a recent Basel Committee newsletter (No. 6, September 2005)

Conclusions

- Validation is a process, not an event
  - Process must specify who, what, when – and include responses linked to established "tolerances"
- Now is the time to consider the validation strategy
  - Models used for IRB should be validated according to the principles of OCC 2000-16
  - Validation should be built into the development process
- Validation should be designed and evaluated in the context of other controls around the IRB system
- Creative thinking and new tools and data are needed

References

- Basel Committee on Banking Supervision, "Update on work of the Accord Implementation Group related to validation under the Basel II Framework," *Basel Committee Newsletter* No. 4, January 2005.
- Basel Committee on Banking Supervision, "Validation of low-default portfolios in the Basel II Framework," *Basel Committee Newsletter* No. 6, September 2005.
- Basel Committee on Banking Supervision, *International Convergence of Capital Measurement and Capital Standards: A Revised Framework,* updated November 2005.

Expert Profiles

MIKE CARHILL

Mike Carhill joined the Office of the Comptroller of the Currency as a financial economist in the Department of Economic and Policy Analysis in 1991. After serving as Deputy Director from 1995-2003, in December 2003 he was appointed Director of the Risk Analysis Division (RAD). Before joining the OCC, Carhill served as a staff economist with the Federal Home Loan Bank of Atlanta, and researched the role that interest-rate risk plays in thrift profitability. Carhill holds a Ph.D. in monetary theory from Washington University.

DENNIS GLENNON

Dennis Glennon is Deputy Director for Credit Risk Modeling in the Risk Analysis Division at the Office of the Comptroller of the Currency. His research interests are in the areas of credit scoring, credit risk modeling, and bank failure analysis. His current research projects include the application of maximum entropy econometric techniques to bank-failure and loan-default modeling, the development of competing-risk models of SBA loan default and prepayment behavior, and the development of alternative methods of assessing model risk. He actively participates in the supervision of retail lending and mortgage credit-scoring and pricing models. Glennon received his doctorate in economics from the University of Missouri - Columbia.

M. NAZMUL HASAN

M. Nazmul Hasan is Lead Expert for Credit Risk Modeling in the Risk Analysis Division at the Office of the Comptroller of the Currency. In his supervisory experience, Hasan has worked on such subjects as credit scoring, risk rating, pricing, and asset valuations. Prior to joining the OCC, Hasan was a senior manager in the Customer Information Management Division at American Express and an assistant professor at Illinois State University. His research interests include both theoretical and applied topics in economics. Hasan holds a master's degree in statistics and a Ph.D. in economics from the University of Illinois at Urbana-Champaign.

NICHOLAS M. KIEFER

Nicholas M. Kiefer is the Ta-Chung Liu professor of economics at Cornell University, where he is a member of the graduate field faculties in economics, statistics and hospitality administration. He is widely known for his theoretical and applied contributions in the econometric modeling of duration data, the estimation of dynamic programming models under learning, and financial market microstructure. Kiefer's current research includes applications in financial economics, credit scoring and risk management, consumer trend forecasting, and development of quantitative management techniques for the restaurant and retail industries. Kiefer is an internationally recognized expert, having published in excess of 100 journal articles. books, and reviews. He is a Fellow of the Econometric Society, and a past recipient of the Guggenheim Memorial Fellowship.

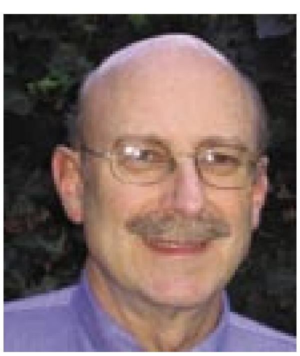
C. ERIK LARSON

C. Erik Larson is Lead Expert for Enterprise Risk the Risk Analysis Division of the Office of the Comptroller of the Currency. His present work involves traveling to the largest of our nation's banks and examining the models developed and used by these institutions to measure and manage risk. Larson's research focuses on statistical and econometric issues in the modeling of credit risk and economic capital. Prior to joining the OCC, Larson analyzed and developed individual and corporate income tax policy in the Office of Tax Analysis at The Treasury. He also taught courses on probability, statistics, and econometrics while on the faculty of the University of Southern California School of Business Administration, and served as a private consultant, specializing in the development of methods to value financial assets. Larson holds a Ph.D. in economics from Cornell and a B.A. from Syracuse University.

MARK LEVONIAN

Mark Levonian is Deputy Comptroller for Modeling and Analysis at the Office of the Comptroller of the Currency. Previously, he served as vice president of Banking Supervision and Regulation at the Federal Reserve Bank of San Francisco. In addition to publishing extensively, Levonian has taught finance at San Jose State University and the University of California at Berkeley and has been an advisor or consultant to the International Monetary Fund, the World Bank, and various foreign central banks. He holds a Ph.D. in economics from M.I.T. and a B.A. in economics from the University of California.

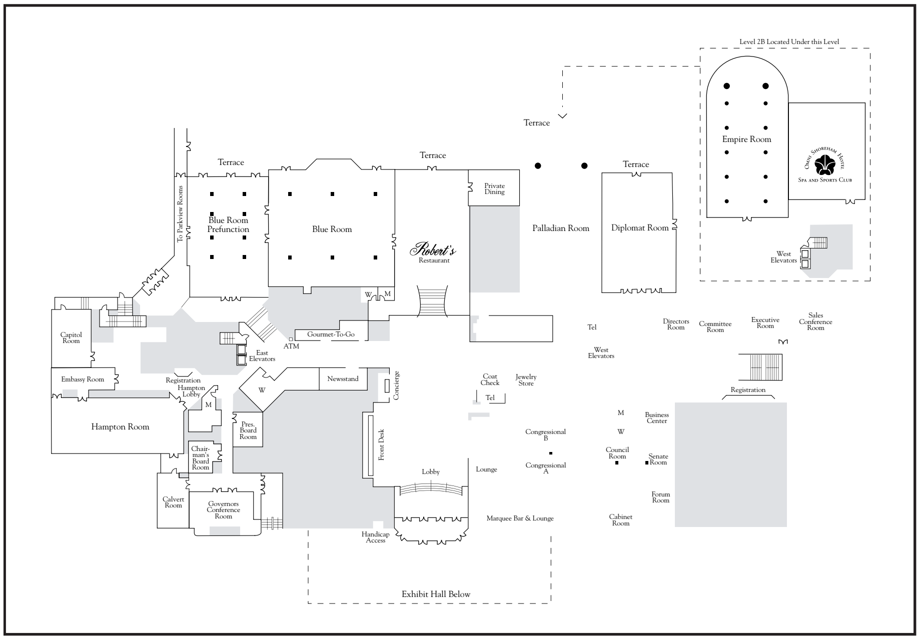
MITCHELL STENGEL

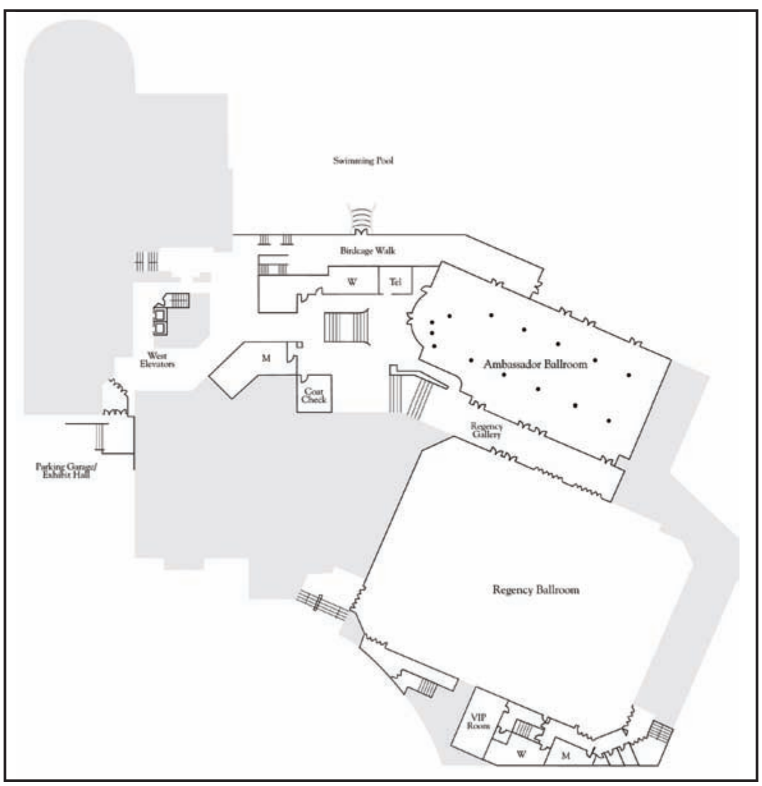


Mitchell Stengel is Lead Expert for Basel II in the Modeling & Analysis department of the Office of the Comptroller of the Currency (OCC). He has served on numerous Basel committees and task forces at the international and U.S. levels since 2000, with a primary focus on the retail credit aspects of the new Accord. Prior to his Basel work, he was Deputy Director of the Credit Risk Modeling unit in the OCC's Risk Analysis division (RAD); in the early 90's he did research on fair lending issues in residential mortgage markets. Stengel holds an A.B. from the University of Michigan and a Ph.D. from Harvard University. Before coming to Washington, he taught economics at the University of Michigan and Michigan State University.

OMNI SHOREHAM HOTEL

2500 Calvert Street NW
Washington, D.C. 20008
Telephone: (202) 234-0700
Facsimile: (202) 265-7972





Office of The Comptroller of the Currency
250 E Street, SW
Washington, DC 20219
http://www.occ.treas.gov