

Session V

Monitoring and Benchmarking

M. Nazmul Hasan

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

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)

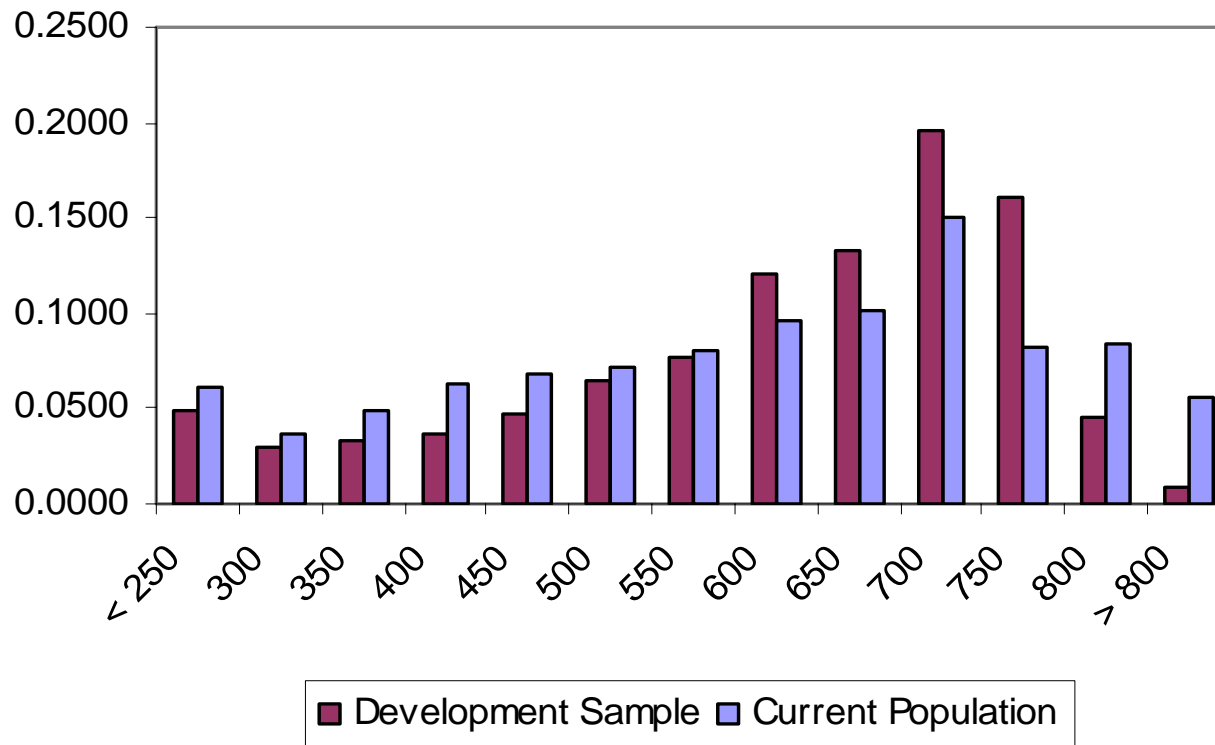
Monitoring and Benchmarking

- 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



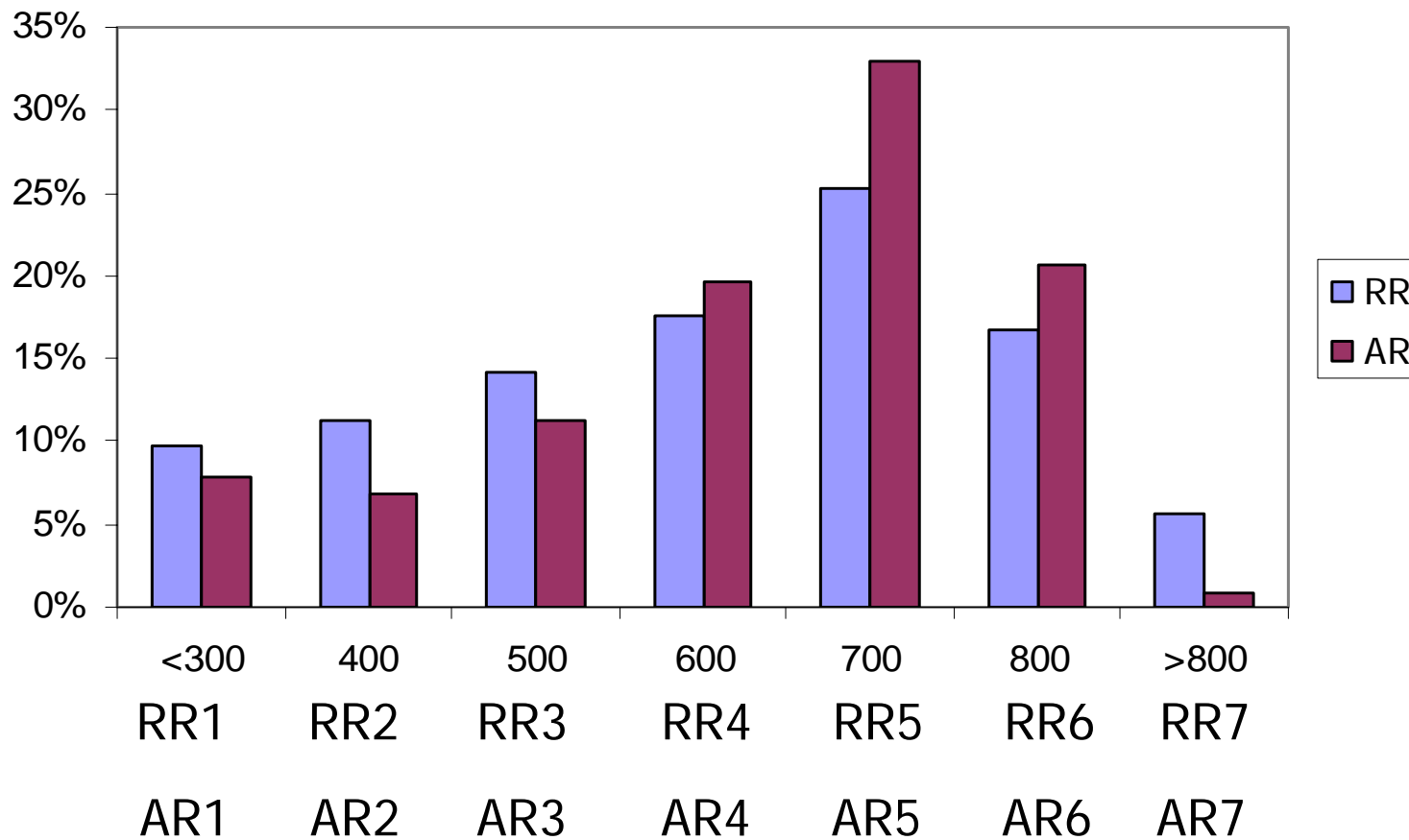
Front-end Analysis

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



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

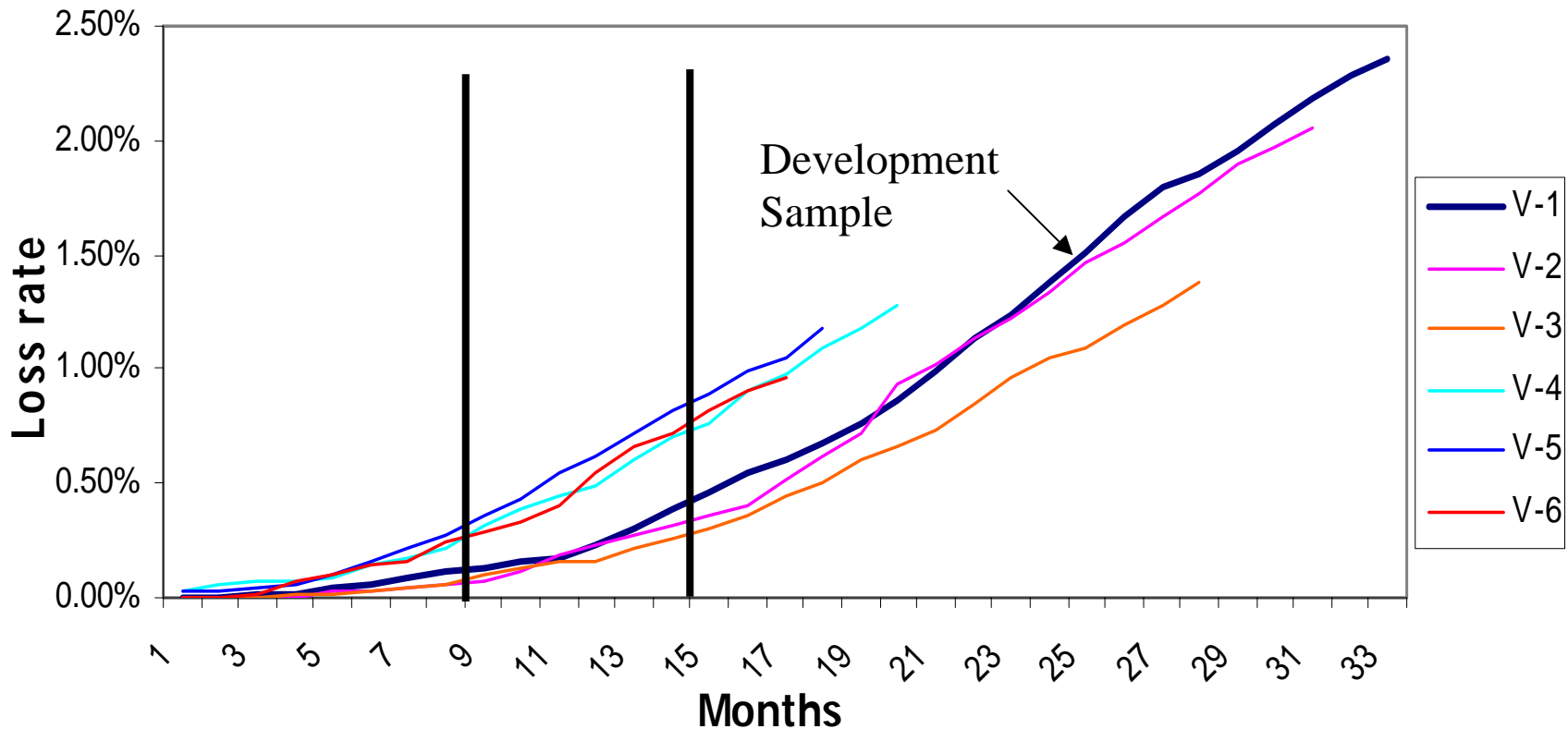
- Borrowers are fixed over time
- Vintage-specific delinquency curves that track the *cumulative* bad rate over time for each vintage
- Vintage curves by score band against some performance measure -- provide a more dynamic benchmark for backtesting the models

Portfolio trend analysis

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

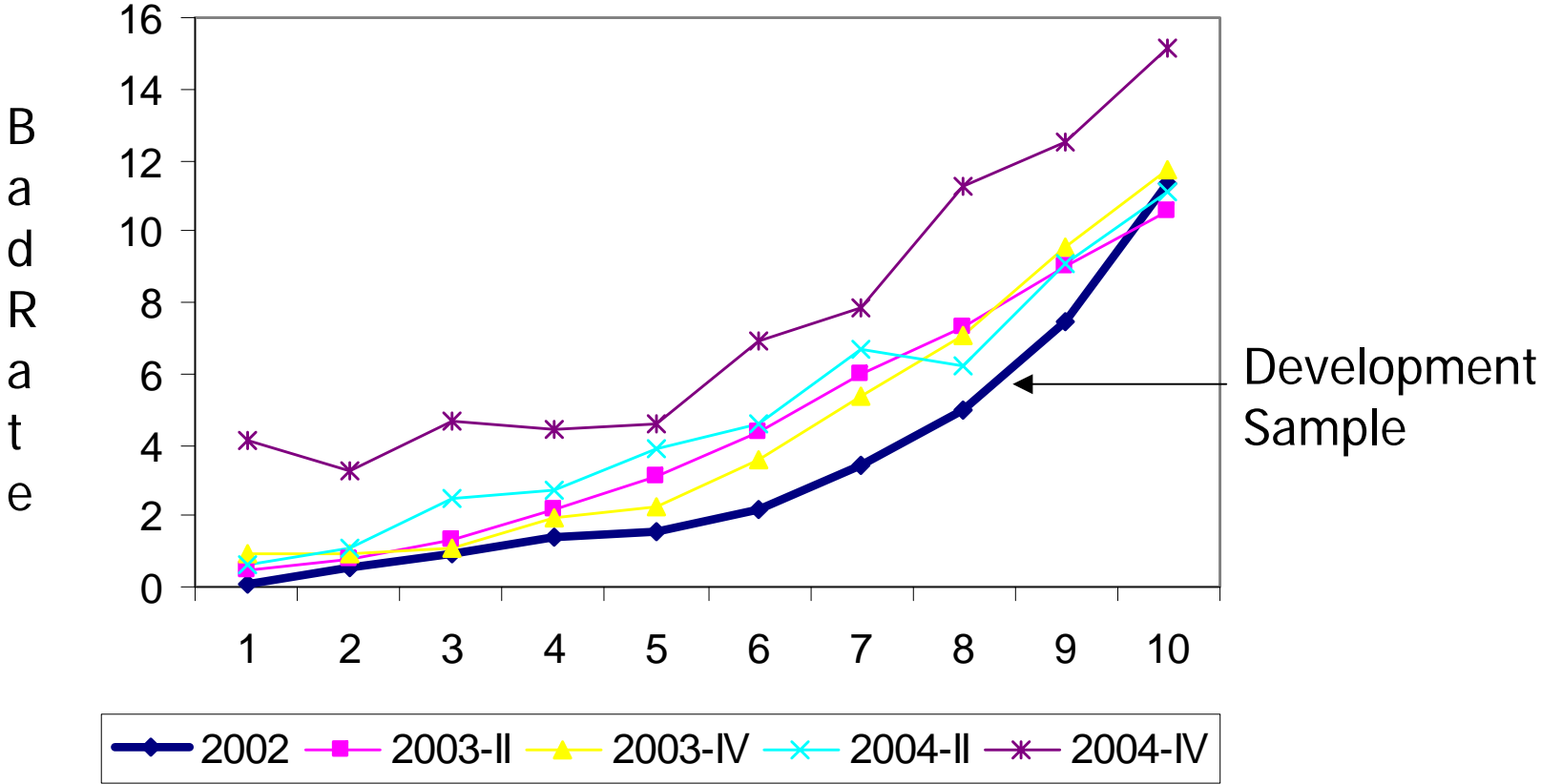
Vintage Curve Analysis: Vintage Specific Cumulative Loss Curve

Tracks the cumulative loss rate over time

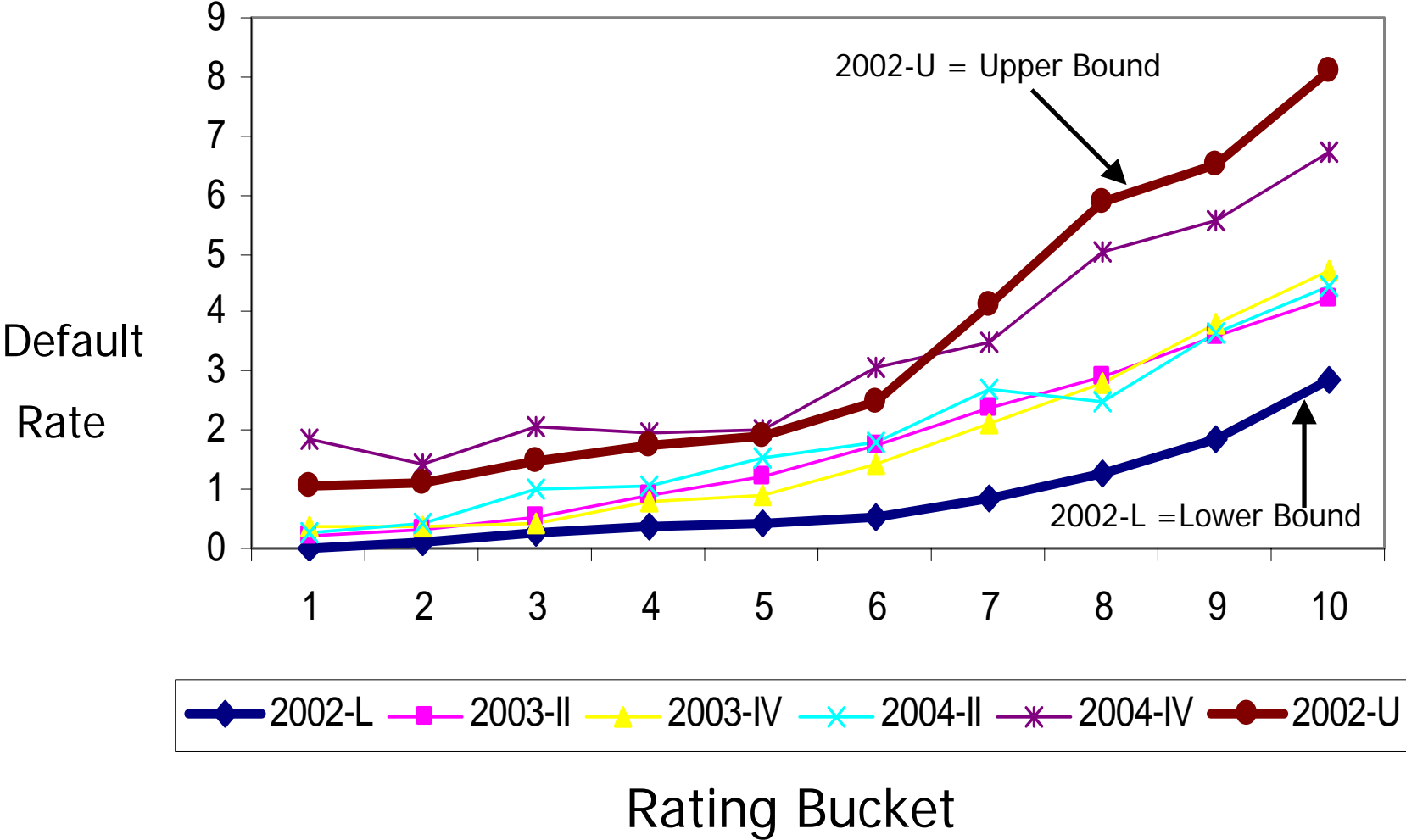


Vintage Curve Analysis: Dynamic Benchmarking for Back Testing

Bad Rate by Score Deciles (15 months on the book)



Back End: Portfolio Trend Analysis

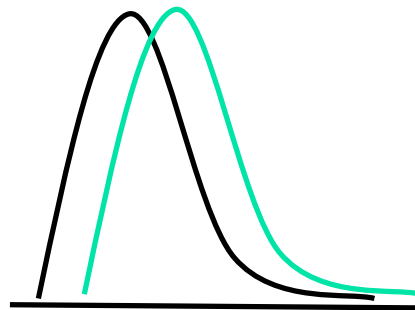


Analysis of Risk Characteristics (Drivers)

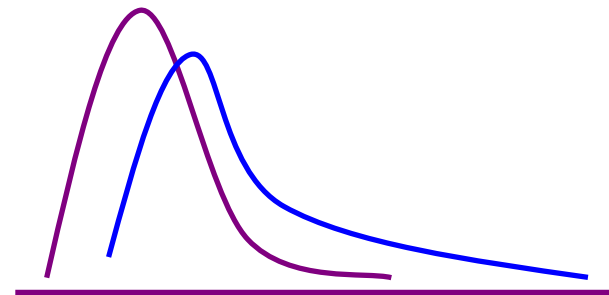
- Isolate the reasons for instability or deteriorating performance of the model
 - Is there any shift in the distribution of a risk characteristic?
 - Analyze how the change in distribution affects the score of a borrower on average
 - If performance data are available, assess the predictive or discriminating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

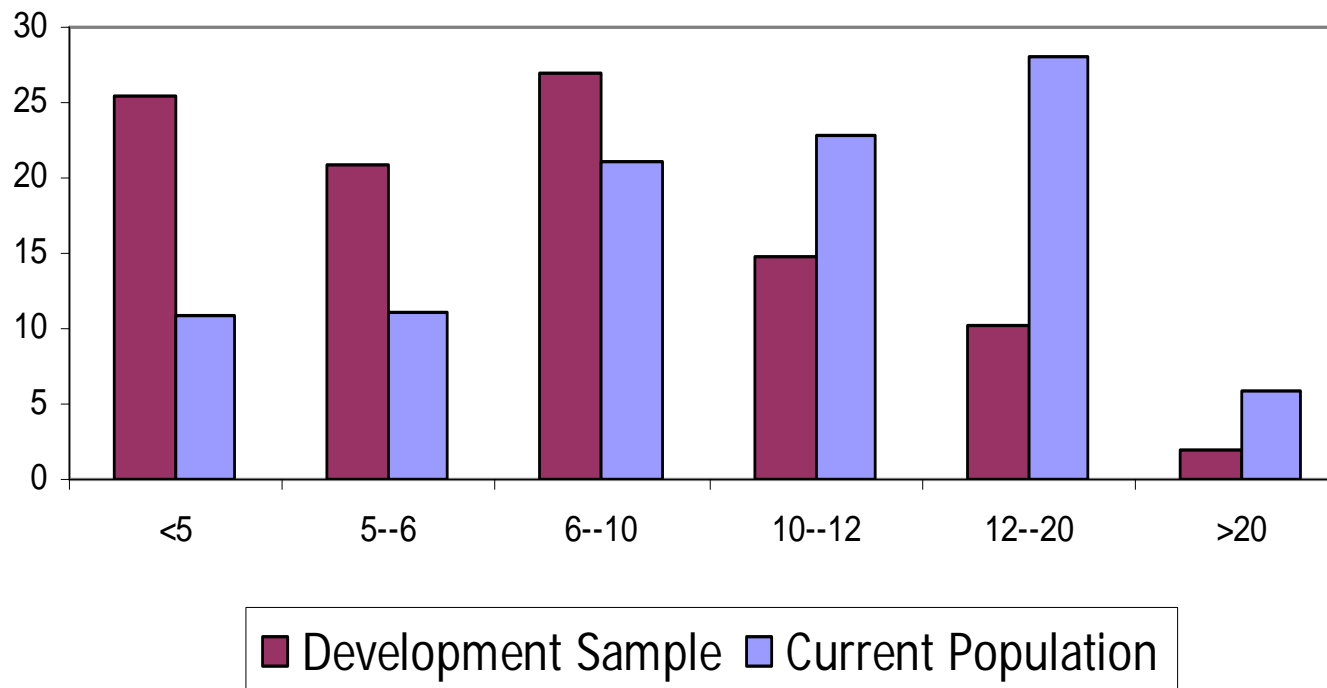
Analysis of Characteristics Consequences: Shift in Distribution

- 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



Comptroller of the Currency
Administrator of National Banks

Validation of Credit Rating and Scoring Models

15 minute Break
The Ambassador Ballroom