Advances in Loss Data Analytics: What We Have Learned at ORX

Federal Reserve Bank of Boston: New Challenges For Operational Risk Measurement and Management

May 14, 2008

Bank of America
Bank of Opportunity™

Regulatory and Management Context

- U.S. Final Rule defines external operational loss data as "gross operational loss amounts, dates, recoveries, and relevant causal information for operational loss events occurring at organizations other than the bank"
- Banks must establish a systematic process for incorporating external loss data into their AMA system
  - Supplement internal data in quantitative models
  - Inform scenario analysis
  - Validate adequacy of internal data and capital
- To be useful in quantitative modeling external loss data should
  - Reduce sampling error when combined with internal data
  - Introduce minimal bias in parameter and/or quantile estimates
  - Data should be stationary for the unit-of-measure under consideration

Bank of America
Bank of Opportunity™

Operational Risk Data Exchange (ORX)

- ORX currently has 42 members in 14 countries and more than 90,000 loss events totaling more than €30 billion
  - Improve understanding of operational risk and key drivers of operational losses
  - Provide peer benchmarks
  - Enhance efforts to measure operational risk exposure
  - Develop and propagate best practices
- Key questions for the use of external data are:
  - Is external loss data relevant to the institution?
  - Do the internal and external loss data come from the same underlying probability distribution?
  - How do we control for regional variation and differences in the size of institutions contributing losses to the database?
- To address these issues, core members of the ORX Analytics Working Group engaged in three work-streams with IBM Research serving as analytics agent
  - Homogeneity analysis
  - Scaling Analysis
  - Combined Homogeneity & Scaling Model

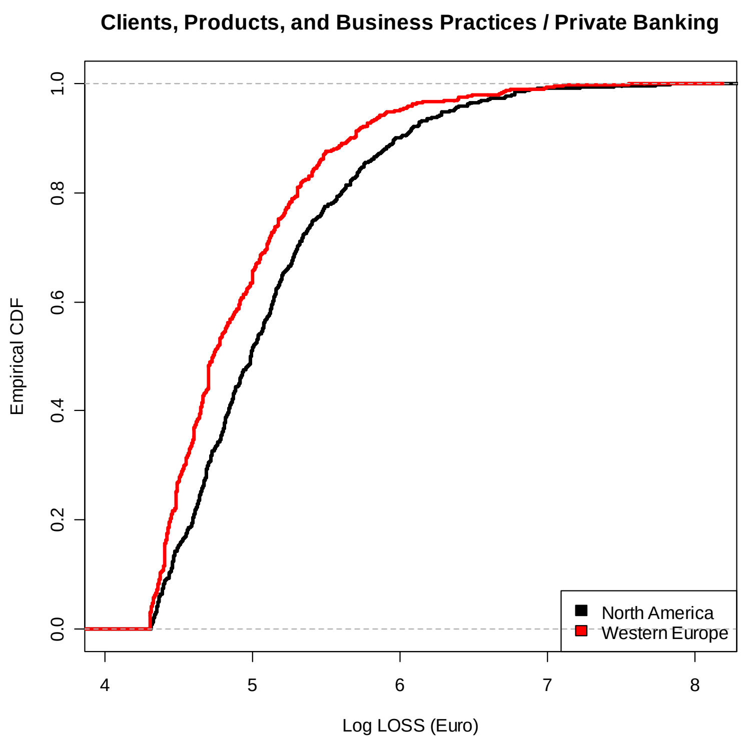
Bank of America
Bank of Opportunity™

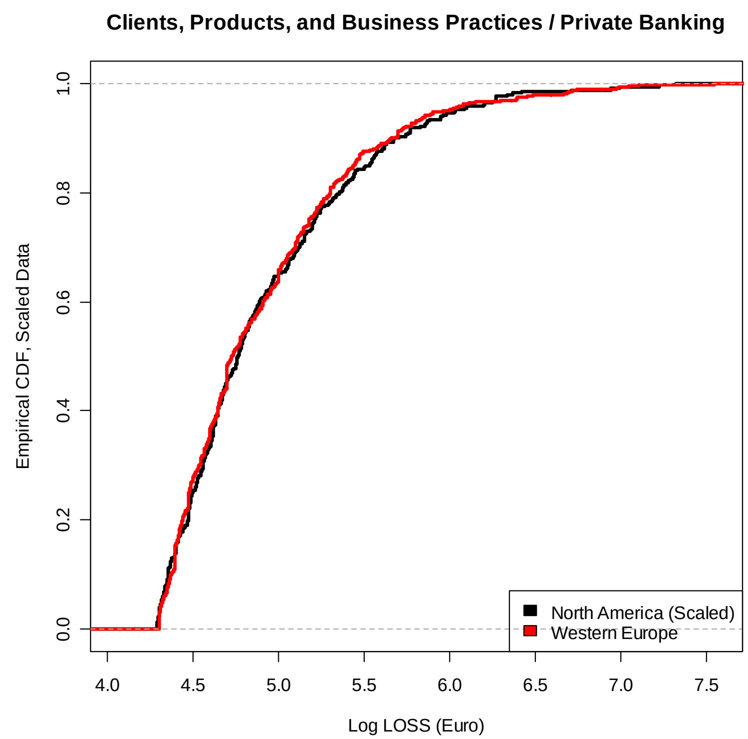
Testing for Homogeneity in External Data

- Analysis explored similarities in the size of losses and shape of loss distributions between ORX members
- Similarity was assessed in terms of statistical measures of goodness-of-fit among loss distributions
- Success was determined by reduction of error in the predicted value of large losses resulting from use of pooled data rather than internal data alone
- Clustering techniques were used to determine groupings of banks with similar loss distributions
- Overall results:
  - Simple transformations of location and scale were effective in aligning many loss distributions
  - A high level of homogeneity was evident in the shapes of various loss distributions across all levels in the sample
  - Groupings were often correlated with firm size and region
  - Pooling losses among banks with similar loss distributions resulted in significant error reductions (20-30%) when estimating high quantiles of the loss severity distribution

Example: Scale Shift Model

- Multiplying North American Private Banking losses by a scale factor results in a distribution almost exactly like that seen in Western Europe



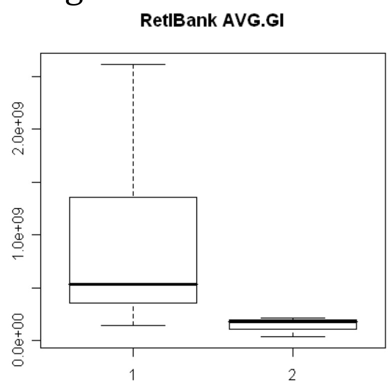


Bank of America
Bank of Opportunity™

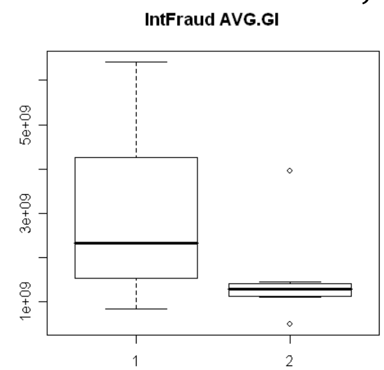
Homogeneous Clusters – Four Examples

Losses averaged across banks are frequently correlated with gross income

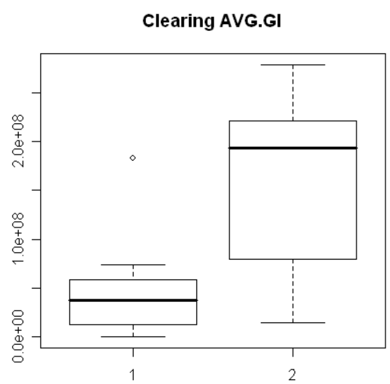
Retail Banking – Cluster sizes: 25, 3; *p* = .01



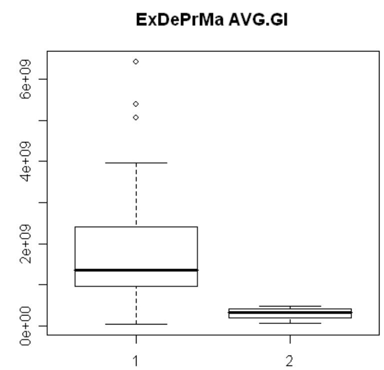
Internal Fraud – Cluster sizes: 12, 8; *p* = .025



Clearing – Cluster sizes: 8, 5; *p* = .065



Execution, Delivery, Process Mgmt: 25, 3; *p* = .025



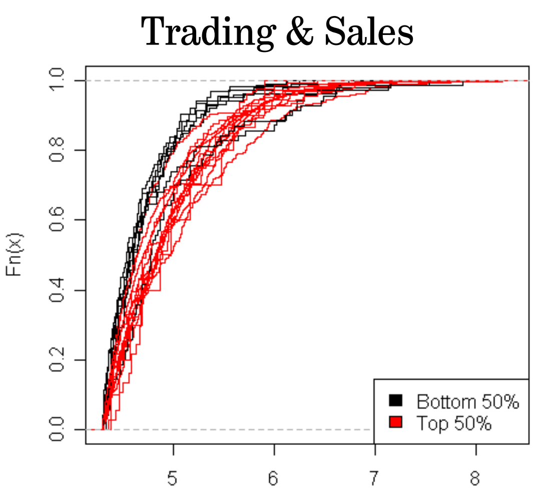
Bank of America
Bank of Opportunity™

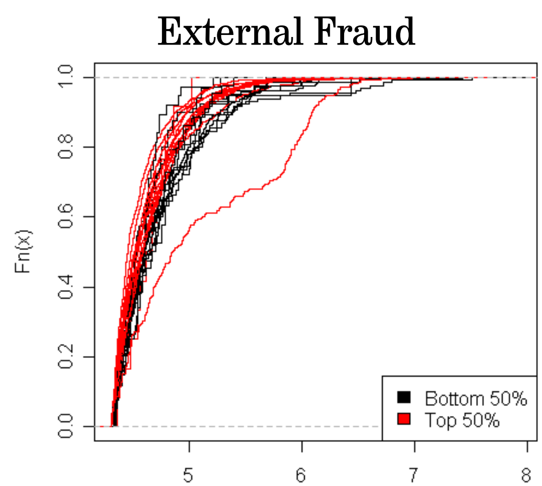
Scaling Methodology

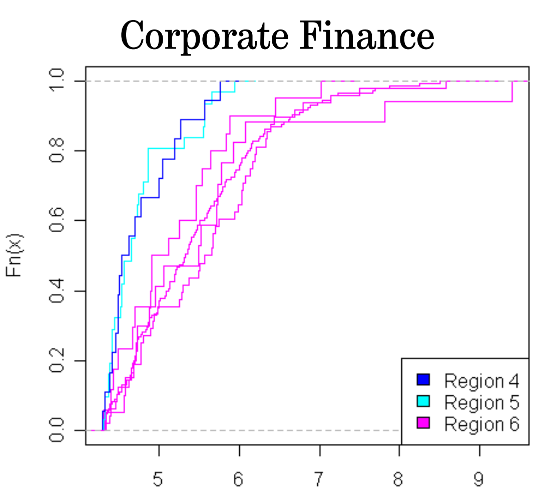
- Determine if indicators such as region and firm size influence the size of small, medium and large losses
  - Provide the distribution location and scale transformations with an economic interpretation
  - Quantile regression methods were used to estimate how losses at each level of the distribution changed with exposure indicators
- Overall Results
  - In many loss categories, the scale of the loss distribution was strongly correlated to the exposure indicators
  - Both increasing and decreasing relationships between loss amount and firms size were observed
  - Large differences were seem between Western European and North American losses in several categories
  - In some cases, large losses scaled differently from small or medium sized losses
- Results submitted for publication in Operational Risk Journal (Spring 2008)

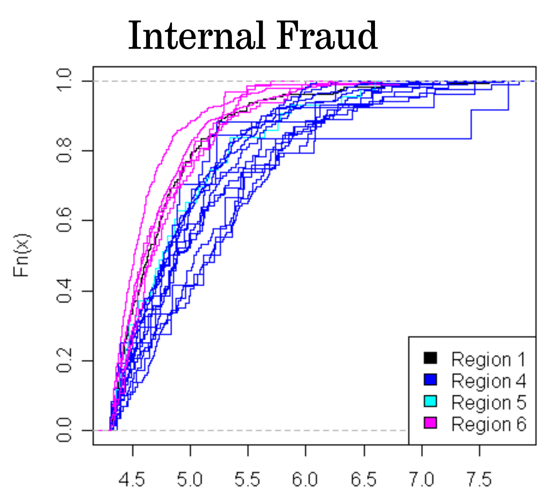
Bank of America
Bank of Opportunity™

Distribution Clusters









Bank of America
Bank of Opportunity™

Combining the Homogeneity and Scaling Methodology

- Derive statistical models for each event type and business line combination
  - Include regional and firm/business line size as appropriate
  - Pool loss data from categories with similar distributions
  - Validate that models can accurately predict future losses

- Developed decision tree methods for characterizing loss distribution shape, scale and location
  - Partitioned data according loss distribution shape
  - Within partitions, we characterize distributional differences by location and scale shifts using quantile regression models

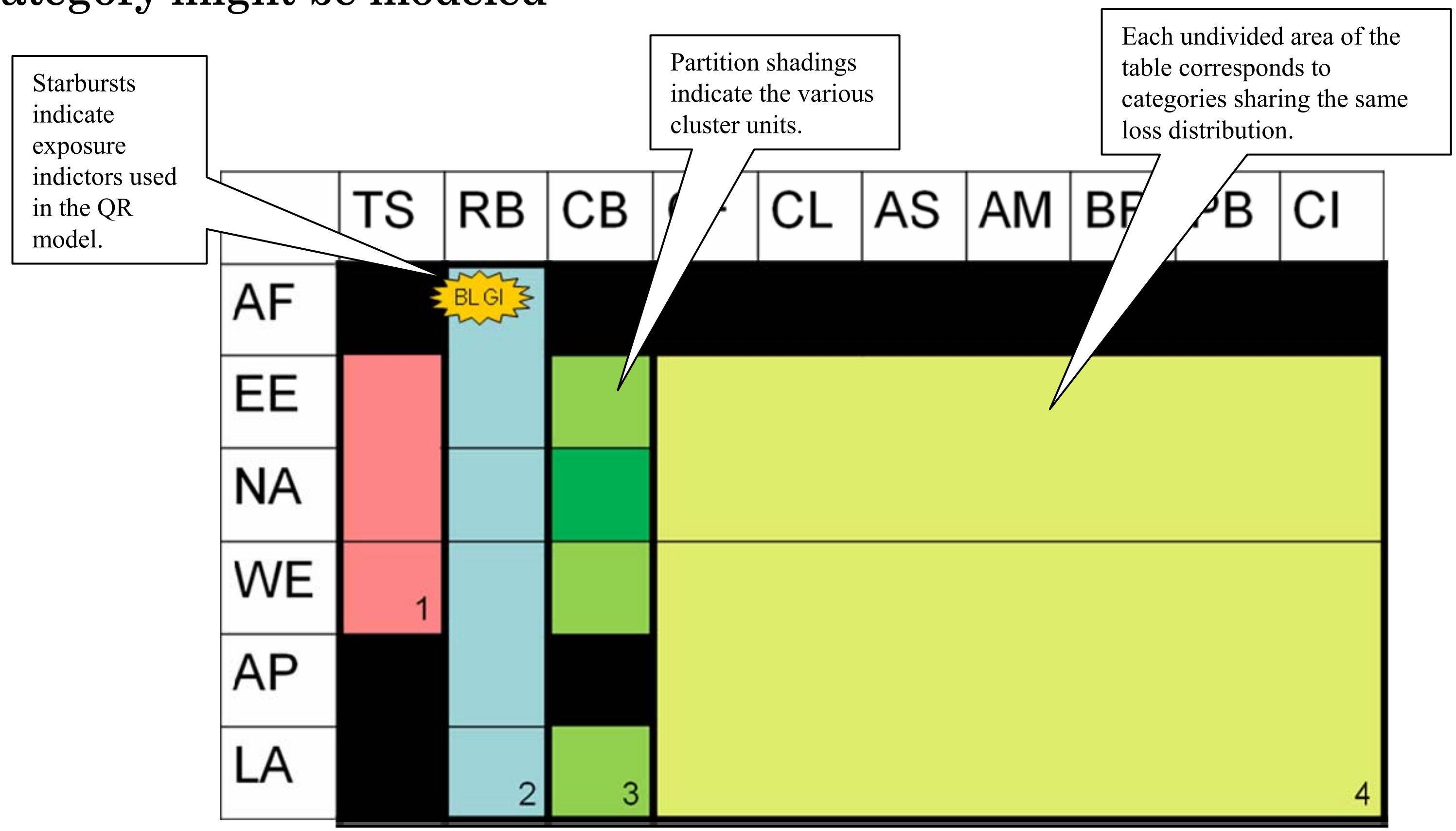
- The quantile regression models returned two sets of results:
  - An *estimated base distribution*, which indicates the overall shape of the distribution
  - *Location- and scale-shift factors*, which indicate how the base distribution should be scaled and shifted in response to different loss subcategories
  - A typical location-shift model is of the form
    - $\log(\text{LOSS}_i) = b_1 \cdot 1\{\text{RB, CB}\} + b_2 \cdot 1\{\text{AP, EE}\} + b_3 \cdot 1\{\text{NA}\} + b_4 \cdot \text{QTR.TOT.GI} / 10^9 + \varepsilon_i$
    - The parameters b_1, b_2, b_3, b_4 are estimated location-shift parameters, and an estimated CDF is provided for the base distribution ε_i

Bank of America

Bank of Opportunity™

Partioning Diagram

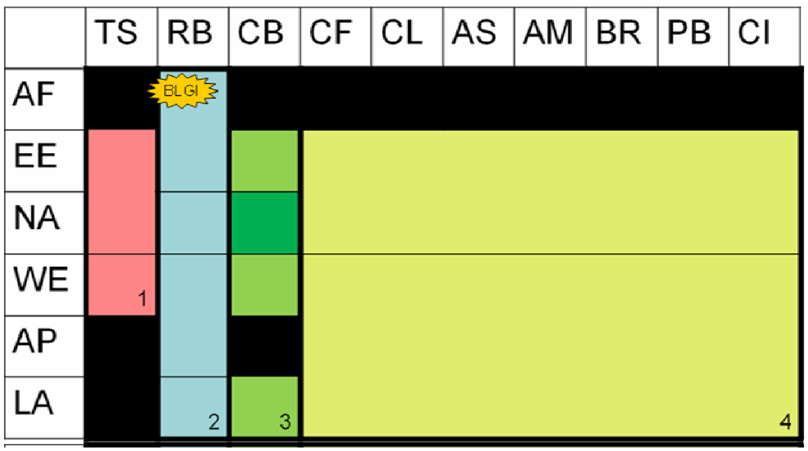
The diagram below gives an example of how losses in a single event category might be modeled



Examples of Loss Distribution Partitions

External Fraud and Clients, Products & Business Practices loss distributions have partitions with significant Gross Income and Regional effects for some banks

External Fraud



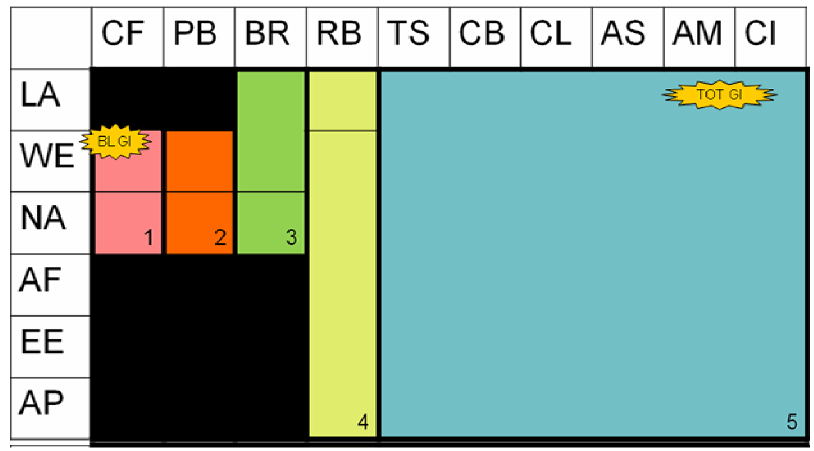
1: $Y_i = -\beta_1 \cdot 1\{WE\} + \varepsilon_i$

2: $Y_i = -\beta_1 \cdot 1\{AP \text{ or } WE\} - \beta_2 \cdot 1\{LA\} - \beta_3 \cdot 1\{NA\} - \beta_4 \cdot QTR.BL.GI / 10^9 + \varepsilon_i$

3: $Y_i = -\beta_1 \cdot 1\{NA\} + \varepsilon_i$

4: $Y_i = \beta_1 \cdot 1\{AP, WE, \text{ or } LA\} + \varepsilon_i$

Clients, Products & Business Practices



1: $Y_i = -\beta_1 \cdot QTR.BL.GI / 10^9 + \beta_2 \cdot 1\{NA\} + \varepsilon_i$

2: $Y_i = \beta_1 \cdot 1\{NA\} + \varepsilon_i$

3: $Y_i = \beta_1 \cdot 1\{NA\} + \varepsilon_i$

4: $Y_i = -\beta_1 \cdot 1\{LA\} + \varepsilon_i$

5: $Y_i = \beta_1 \cdot QTR.TOT.GI / 10^9 + \varepsilon_i$

Bank of America
Bank of Opportunity™

Scaling Analysis Results

- The scaling model can control for large differences between Western European and North American losses in several categories
  - Corporate Finance losses in Clients, Products, and Business Practices were higher in North America
  - Retail Banking losses for Internal and External Fraud were higher in Western Europe
  - A possible economic explanation for this may be differences in legal and regulatory environments
- It also provides a mechanism to control for differences in loss characteristics between smaller and larger banks
  - Risk management practices
  - Product Complexity
  - Transaction Size
- Significant business utility is realized from using external loss data
  - Homogeneity analysis suggests that model accuracy can be improved by an average of 20% by pooling industry data
  - Larger improvements are expected using scaled data