LDA at Work: Deutsche Bank’s Approach to Quantifying Operational Risk

Workshop on Financial Risk and Banking Regulation

Agenda

1. AMA at DB
2. Main components of an LDA model
3. Appendix: model validation

For more information:
AMA Model Development at Deutsche Bank

Timeline
1999  Systematic collection of loss data
2000  Economic capital with LDA
- Top-down model: loss distribution at Group level, capital allocation with risk indicators
- Internal and external loss data
- Qualitative adjustment with Incentive Scheme
2001  AMA project
2002  Development of AMA model
2003  Implementation of prototype
2004  EC test calculations with AMA model
2005  Official EC calculation with AMA model (starting Q2 05)
2006  Implementation of production engine
AMA application submitted in September
2007  Regulatory approval
2008  Joint EC and RC calculation with AMA model

AMA at DB: Calculation flow

Group-Level
- Net Losses
- Gross Losses
- Insurance
- Frequency/ Dependancy
- Severity

Event Types
- ORX
- External Data
- Internal Data

Allocation
- Group-Level
- Business Division
- RC before QA
- RC after QA
- Qualitative Adjustment (BE&ICF)
- Scenario analysis
### DB’s Business Line / Event Type Matrix

<table>
<thead>
<tr>
<th>Basel Level 1</th>
<th>Internal Event Types</th>
<th>Business Lines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal Fraud</td>
<td></td>
<td>BL1</td>
</tr>
<tr>
<td>External Fraud</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manage to physical assets</td>
<td>Internal Fraud</td>
<td></td>
</tr>
<tr>
<td>Business disruption</td>
<td>Clients, Products, Business Practices</td>
<td></td>
</tr>
<tr>
<td>Executions, delivery, process management</td>
<td>Clients, Products, Business Practices</td>
<td></td>
</tr>
<tr>
<td>Employment practices, workplace safety</td>
<td>Employment practices, workplace safety</td>
<td></td>
</tr>
</tbody>
</table>

- Design criteria
  - comparable loss profile
  - same insurance type
  - same management responsibilities
  - availability of data
  - relative importance of cells

- Treatment of losses that cannot be assigned to a single cell
  - Group losses
  - Split losses

---

### Agenda

1. **AMA at DB**
2. **Main components of an LDA model**
3. **Appendix: model validation**
Data for Modelling Loss Distributions

Data sources
- Internal loss data
- Consortium data
- Commercial loss database
- Scenarios

Internal loss data is the most important data source
- Each firm's operational losses are a reflection of its underlying operational risk exposure
- Internal losses are used for
  - modelling frequencies (exclusively)
  - modelling severities
  - estimating correlations

Motivation for using external data and scenarios
- Additional information on severity profile, in particular on risk of unexpected losses (tails of severity distributions)
Creating a Relevant Loss Data Set

Scenarios are added as individual data points to relevant external losses
Biased External Loss Data

**Scale Bias**
- Operational risk is dependent on the size of the bank, i.e. the scale of operations
- The actual relationship between the size of the institution and the frequency and severity may be stronger or weaker depending on the particular OR category

**Truncation Bias and Data Capture Bias**
- Collection thresholds are not uniform for different data sets
- Data is often captured with a systematic bias. This problem is particularly pronounced with publicly available data: there exists a positive relationship between the loss amount and the probability that the loss is reported
- The disproportionate number of large losses could lead to an estimate that overstates a bank’s exposure to operational risk

**Scaling in AMA at DB**
- No correction of Scale Bias since it is considered less relevant for severity modeling
- Correction of Truncation Bias and Data Capture Bias

---

**Frequency Modelling**

- Group-Level
- Business Division
- Gross Losses
- Net Losses
- RC before QA
- Correlation / Diversification
- Insurance
- Scenario analysis
- Severity
- Frequency
- Aggregate distribution
Frequencies in AMA at DB

Data
Only internal loss data is used for calibrating frequency distributions:
- Internal loss data reflects DB’s loss profile most accurately
- Difficult to ensure completeness of external data (essential for application in frequency calibration)
- Lower data requirements in frequency modeling (compared to severity modeling)

Implemented distributions
- Poisson (no dependence between occurrence of events in a cell)
- Negative Binomial (positive dependence)
- Selection algorithm based on statistical tests

Frequency distributions in official capital calculations
- Poisson in all cells
- Reason: negligible difference to combination of Poisson and Negative Binomial cells

Severity Modelling

Group-Level

Correlation / Diversification

Insurance

Gross Losses

Net Losses

Scenario analysis

Frequency

Severity

Deutsche Bank's AMA Model · page 13

Deutsche Bank's AMA Model · page 14
Modelling Decisions

Range of distribution
- One distribution for the entire severity range
  or different distributions for small, medium and high losses?

Choice of distribution family
- Two-parametric distributions like lognormal, GPD
  or more flexible distribution families, i.e. three- or four-parametric,
  or even empirical distributions?
- One distribution family for all cells
  or selection of “best” distribution based on quality of fit?

Mixing internal and external data
- How much weight is given to internal and external data?
- How to combine internal and external data?

Severities in AMA at DB

Range of distribution and choice of distribution family
- In many cells, data characteristics are different for small and big losses
- Different distributions for body and tail
  - Body: non-parametric (empirical) distribution
  - Tail: modified technique from Extreme Value Theory for tail modelling
- Empirical and parametric distributions are combined via a weighted sum
  applied to the cumulative distribution functions

Mixing internal and external data
- Internal data for calibrating body of distribution
- Internal and external data for calibrating tail
Core Idea: Piecewise Defined Severity Distributions

**First section:** given by empiric distribution of cell specific internal data

**Mid section:** given by weighted average of
- empiric distribution of cell specific internal data
- empiric distribution of cell specific external and scenario data

**Tail section:** given by weighted average of
- empiric distribution of cell specific internal data
- empiric distribution of cell specific external and scenario data
- parametric distribution calibrated on all data >= 50mn
Insurance in AMA at DB

Insurance calculation process
- OR event types are mapped to insurance policies
- Insurance policies are modelled individually, e.g. by specifying deductible, limits and haircut
- Insurance payment is calculated for each of the simulated gross losses separately

Insurance Mapping

<table>
<thead>
<tr>
<th>OR event types</th>
<th>Insurance policies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fraud</td>
<td>Fidelity</td>
</tr>
<tr>
<td>Infrastructure</td>
<td>Burglary, Theft, Robbery</td>
</tr>
<tr>
<td>Execution, Delivery &amp; Process Mgmt</td>
<td>Property Damage</td>
</tr>
<tr>
<td>Clients, Products &amp; Business Pcts</td>
<td>Service &amp; Elec. Break-Down</td>
</tr>
<tr>
<td>Employment Practices &amp; Workplace</td>
<td>General Liability</td>
</tr>
<tr>
<td>Safety</td>
<td>Professional Liability</td>
</tr>
<tr>
<td></td>
<td>Employers Practice Liability</td>
</tr>
<tr>
<td></td>
<td>Not insured</td>
</tr>
</tbody>
</table>
Modelling Insurance Contracts

- Deductible: amount the bank has to cover by itself
- Cap: maximum amount compensated by the insurer

\[ \min(c, \max(x - d, 0)) \]

Additional features
- Aggregate caps
- Haircuts (regulatory requirements)

Modelling Dependence

- RC before QA
- Aggregate distribution
- Correlation / Diversification
- Net Losses
- Gross Losses
- Insurance
- Frequency
- Severity
- Scenario analysis

Deutsche Bank's AMA Model - page 21

Deutsche Bank's AMA Model - page 22
Analyzing Dependence

Dependence in a bottom-up LDA

- Within cells
  - Dependence between the occurrence of loss events
  - Dependence between the frequency distribution and the severity distribution
  - Dependence between the severity samples

- Between cells
  - Dependence between the frequency distributions
  - Dependence between the severity distributions

Statistical analyses performed at Deutsche Bank

- Based on internal loss data
- Identification of dependence between
  - occurrence of loss events within a cell → Frequency distribution not Poisson
  - frequency distributions in different cells → Copula applied to frequencies

Dependence in AMA at DB

Frequencies

- Gaussian copula applied to frequency distributions

Severities

- Sum of split losses
- Severities of different loss events are independent

Example: Gaussian copula applied to a Poisson and a Negative Binomial distribution

uncorrelated

Correlation factor $\sqrt{0.5}$ in Gaussian copula
Calculation of Capital

<table>
<thead>
<tr>
<th>Event Type</th>
<th>Cons. Data</th>
<th>External Data</th>
<th>Internal Data</th>
<th>Staging area</th>
<th>dbIRS</th>
</tr>
</thead>
</table>

Calculation and Allocation of Risk Capital

**Aggregate loss distribution:** Monte Carlo simulation

**Economic Capital:** 99.98% Quantile minus Expected Loss

**Regulatory Capital:** 99.9% Quantile minus Expected Loss

**Capital allocation**
- **Cell level:** Expected Shortfall allocation
- **Divisional level:** Aggregation of EC in divisional cells plus proportional contributions of Group cells
Qualitative Adjustment

- Risk indicators and self assessment are main components
- QA score (applied on BL/ET level) plus penalty component (inappropriate loss data collection, KRI/SAT minimum standards, etc.)
- Insurance OR capital may be adjusted by +40% to -40%
- Measurement of risk sensitivity and coverage determines range

Key risk indicators
- Global KRI: HR, BCM, open issues (Audit, SOX, db-Track), NPA, Technology Risk
- Business specific KRI: e.g. nostro reconciliations, outstanding confirmations, average processing time of customer complaints
Agenda

1. AMA at DB
2. Main components of an LDA model
3. Appendix: model validation

Validation

Basic properties of LDA model
- Variance analysis
- Loss distributions for heavy-tailed severities

Sensitivity analysis of basic components of LDA models
- Frequencies
- Severities
- Dependence
- Insurance

Impact analysis of stress scenarios

Backtesting and benchmarking
- Benchmarking the tail of the aggregate loss distribution against individual data points
Variance Analysis

Cell level

- Variance analysis
  - does not provide information on quantiles of loss distribution
  - but: quantifies impact of frequencies and severities on volatility of aggregate losses
  - is independent of specific distribution assumptions
- Variance of aggregate losses ($F$ and $S$: frequency and severity distribution):

\[ E(F) \cdot \text{Var}(S) + \text{Var}(F) \cdot E(S)^2 \]

Conclusion

- Importance of frequency distribution depends on relationship of $\text{Var}(F)/E(F)$ (frequency vol) and $\text{Var}(S)/E(S)^2$ (severity vol)
- In high impact cells, the volatility of severities dominates and the actual form of the frequency distribution is of minor importance:

\[ E(F) \cdot \text{Var}(S) + \text{Var}(F) \cdot E(S)^2 \]

Variance Analysis

Group level

Frequency correlations

- Variance of loss distribution at Group level

\[ \sum_{j=1}^{m} E(F_j) \cdot \text{Var}(S_j) + \text{Var}(F_j) \cdot E(S_j)^2 + \sum_{j,k=1}^{m} \text{Cov}(F_j, F_k) \cdot E(S_j) \cdot E(S_k) \]

- Variance in the homogeneous model ($c$: homogeneous correlation coefficient)

\[ m \cdot (E(F) \cdot \text{Var}(S) + \text{Var}(F) \cdot E(S)^2 \cdot (c \cdot (m - 1) + 1)) \]

Impact of frequency correlations depends on

- number of (relevant) cells $m$ and
- relationship of $\text{Var}(F)/E(F)$ (frequency vol) and $\text{Var}(S)/E(S)^2$ (severity vol)

In general, the impact of frequency correlations is rather limited and less significant than the impact of correlations of severities or loss distributions.
Loss Distributions for Heavy-Tailed Severities

Subexponential distributions
- Heavy-tailed: tail decays to 0 slower than any exponential $\text{Exp}[a \cdot x], \ a < 0$
- Tail of the sum of subexponential variables has the same order of magnitude as tail of the maximum:

$$\lim_{x \to \infty} \frac{P(X_1 + \ldots + X_n > x)}{P(\max(X_1, \ldots, X_n) > x)} = 1$$

Aggregate loss distributions of subexponential severities
- Let $F$ be a frequency distribution
- $S$ the distribution function of a subexponential severity
- $G$ the distribution function of the aggregate loss distribution
- Under general conditions on $F$ (satisfied by Poisson and Negative Binomial):

$$\lim_{x \to \infty} \frac{\mathbb{E}(G(x))}{S(x)} = E(F), \quad \text{where } S(x) := 1 - S(x)$$

Sensitivity Analysis of Basic LDA Components

Based on theoretical results and experience with Deutsche Bank’s LDA model

- Frequency distributions
  - Mean of frequency distribution is important
  - Shape has limited impact on capital in cells with fat-tailed severities
  - Shape has limited impact on Group capital

- Severity distributions
  - Weights and techniques for combining different data sources are important
  - Significant impact of distribution assumptions for severity tails and tail probabilities

- Dependence
  - Impact depends on the level where dependence is modelled, e.g. frequencies, severities or aggregate losses
  - Limited impact of frequency correlations
**Sensitivity Analysis of Insurance Model**

**OR event types**

- Fraud
- Infrastructure
- Execution, Delivery & Process Management
- Clients, Products & Business Practices
- Employment Practices & Workplace Safety

**Insurance policies**

- Fidelity
- Burglary, Theft, Robbery
- Property Damage
- Service & Elec. Break-Down
- General Liability
- Professional Liability
- Employers Practice Liability
- Not insured

- Clients, Products & Business Practices consumes most of the capital
  - Impact of mapping percentages to insurance contracts
  - Most severe losses fall under Professional Liability: single limit of PL is particularly important
- Higher reduction (in percentage) for median (EL) than for high quantiles (EC and RC)
- Insurance may cause reallocation of capital between different event types

**Stressing Loss Data**

Methodology: Add (remove) internal and/or external losses and analyze impact on capital

**Stress Scenario**

Add 200mn loss in a Fraud cell

**Impact on capital**

- Fraud / BL 4: +50mn
- BL 4: +35mn
- Group: +15mn

**Impact on divisional capital**

Bar charts showing the impact on cell capital and divisional capital.
Backtesting and Benchmarking

- **Backtesting**
  - Sequential testing of a model against reality to check the accuracy of the predictions
  - Backtesting is frequently used for the validation of market risk models
  - In credit and operational risk, the inherent shortage of loss data severely restricts the application of backtesting techniques to capital models

- **Benchmarking**
  - Comparison of a bank's operational risk capital charge against a bank's close peers
  - Comparison of the AMA capital charge against the BIA or TSA capital charges
  - Comparison of the LDA model outputs against adverse extreme, but realistic, scenarios

These tests help to provide assurance over the appropriateness of the level of capital but there are obvious limitations

---

Benchmarking

**Tail of aggregate loss distribution versus individual data points**

- Based on assumption that these tails have the same order of magnitude:
  - Tail of aggregate loss distribution calculated in a bottom-up LDA model
  - Tail of loss distribution directly specified at Group level

- Loss distribution specified at Group level:
  - Take all losses (across business lines and event types) above a high threshold, say 1m, for the specification of a severity distribution $S$
  - Calculate the bank's average annual loss frequency $n$ above 1m

- Under the assumption that $S$ is subexponential, identify

$$
\alpha - quantiles \text{ of } \text{loss distribution } S_1 + \ldots + S_n \text{ with } \\
\alpha - quantiles \text{ of } \text{maximum distribution } \max(S_1, \ldots, S_n) \text{ with } \\
1 - ((1 - \alpha)/n) - quantiles \text{ of severity distribution } S
$$
Benchmarking Result

- $1 - \frac{(1-\alpha)}{n}$ – quantiles of the severity distribution correspond to individual losses for appropriate alpha and n.
- The amount of loss data provides a limit for the confidence level that can be derived directly from the data.

Application of this method to DB’s LDA model