Integrating External Data Into Operational Risk Management

Eric Rosengren
Senior Vice President
Federal Reserve Bank of Boston

June 9, 2003

PRMIA

The views expressed in this paper do not necessarily reflect those of the Federal Reserve System.

Outline

- Overview of operational risk
- The need for external data
- Adjusting for external data bias
- Integrating into self-assessments and scenarios.
- Technical paper can be found in a Federal Reserve Bank of Boston working paper `http://ssrn.com/abstract=395083`

Operational risk

- “The risk of loss resulting from inadequate or failed internal processes, people, and systems or from external events.”
- A variety of loss types:
  - Fraud.
  - Discrimination and wrongful termination.
  - Sales practices.
  - Damage to physical assets.
  - Systems problems, data entry, paperwork.
  - Compliance, regulatory, antitrust.

Importance of Operational Risk.

- More than 100 losses exceeding $100 Million over the last decade.
- Banks recognize the importance of oprisk:
  - On average, they hold 15% of their capital for Op Risk.
- Cause for regulatory concern.
  - deposit insurance.
  - Systemic issues.

Integrating external data into LDA

- Set frequency and severity based on self assessments and scenarios
- External data provides a cross-check for limited data from internal losses
- Because the tail is important for capital calculation - these choices are important in evaluating capital for operational risk

Business line risk assessments

- Compare internal loss experience with external loss experience in the same business line
- How do business mix and control environment impact loss experience
- How would changes in business impact future loss experience

Scenario

- Make loss analysis forward looking and tailored to bank
- Ask business lines for frequency of loss, by size of loss
- Model large losses, with help of specialists
- Compare results with external loss experience - are frequency and severity estimates consistent with industry experience?

Data strategies for LDA.

- High frequency low severity events.
  - Internal data should suffice.
- Low frequency low severity events.
- High frequency high severity events.
- Low frequency high severity events.
  - Internal data insufficient.
  - Need for external data.

External data.

- Vendors are OpRisk Analytics, OpVantage.
- Collect data from public news sources.
- Events over $1M from the past 10+ years.
- Vendors provide scaling data.
- Potential difficulties:
  - Business line classification.
  - Non-finalized loss amounts.
  - Non-monetary losses.
  - Intentional actions and fraudulent firms.

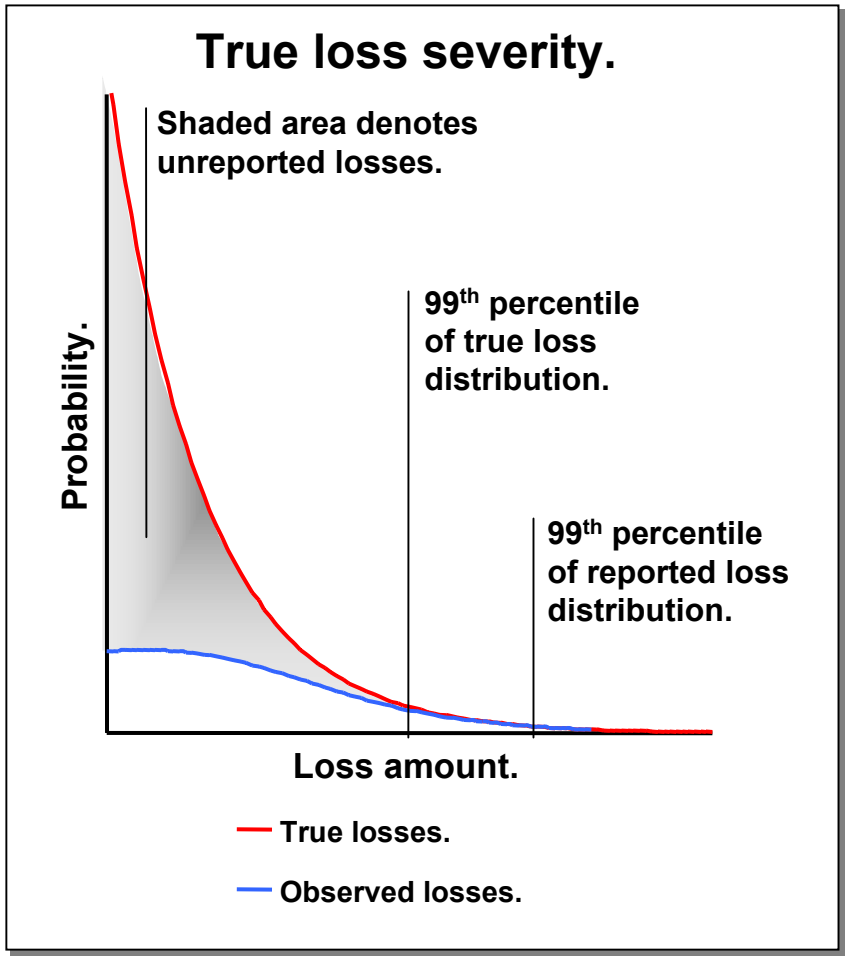
Basic statistics for U.S. losses.

	% of Losses		3rd Qrt. ($B)	
	OpR	OpV	OpR	OpV
Corp. Fin.	6%	4%	23	23
T&S	9%	9%	44	27
Ret. Bank.	38%	39%	11	12
Com. Bank.	21%	16%	24	28
P&S	1%	1%	11	11
Agency Svc.	2%	3%	110	28
Asset Mgmt.	5%	6%	20	22
Ret. Brok.	17%	22%	12	13
Total			17	17

- Most business lines are well populated.
- Apparent variation across business lines.
- Similarity across databases.
- Non-US losses are larger, less agreement across databases.
- Only 2 event types with many observations.

The 99.9th percentiles are $1.4B and $1.2B.
Are losses really that heavy-tailed?

Reporting bias.



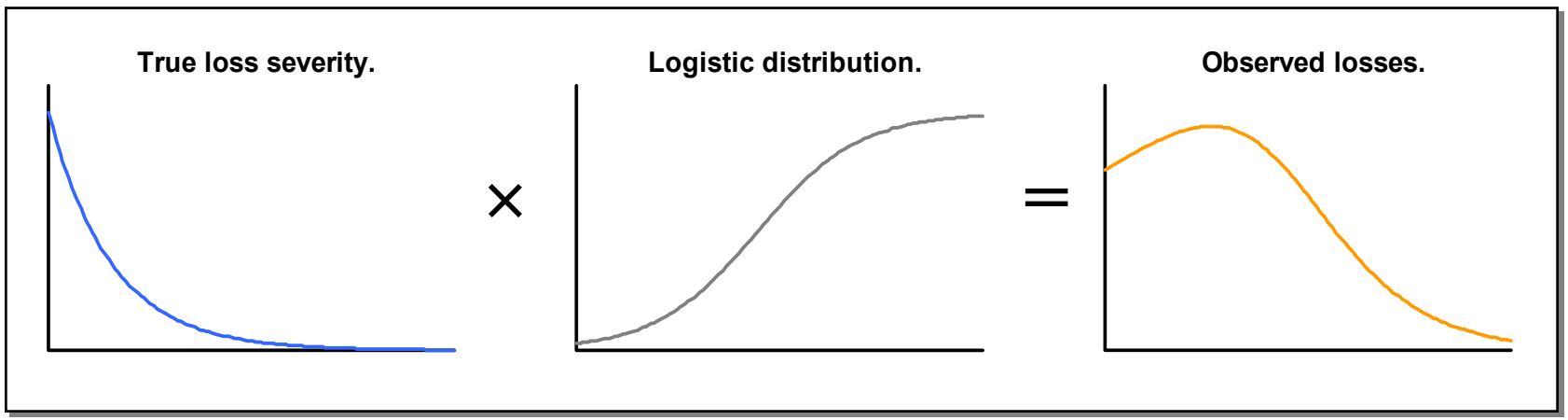
- Not all losses are reported.
- Reporting probability increases with loss amount.
- Loss severity estimates are biased upwards.
- Percentiles from the severity distribution also biased upwards.
- Capital estimates will likely be too high.

Stochastic truncation.

- Assume that loss x is observed only if it exceeds an unobserved (random) truncation point y.
- Density of observed losses equals the “true” density times the probability of being observed:

$$f(x \mid x > y) \sim f(x)\,\mathrm{G}(x).$$

- many factors influence reporting



Modeling the loss distribution.

- We use extreme value theory (EVT)
- EVT shows that “excesses” over a high threshold converge to a generalized pareto distribution, as the threshold increases, regardless of the underlying distibution

$$\mathrm{GPD}(x) = \begin{cases} 1 - (1 + \xi x/b)^{-1/\xi} & \xi > 0 \\ 1 - \exp(-x/b) & \xi = 0 \end{cases}$$

Model specification.

- Parameters:

  b ~ tail thickness parameter.

  τ ~ the $ loss with a 50% reporting probability.

  β ~ rate of increase in reporting probability.

Model results.

Table 3. Results for model 1.

	OpR	OpV
Exponential b	0.64	0.66
	(0.08)	(0.07)
Logistic β	0.78	0.82
	(0.10)	(0.09)
Logistic τ	4.45	4.46
	(0.36)	(0.32)
exp(τ)	\$86M	\$86M

- Data sets have remarkably similar parameter estimates.
- 99th and 99.9th percentiles are about \$20M and \$90M, respectively.
- The estimates of τ seem reasonable.

Sensitivity to threshold choice.

Results are not overly sensitive to threshold choice.

Table 4. Estimation results for various POT thresholds.

	Threshold value				
	$1M	$2M	$3M	$5M	$10M
OpRisk Analytics:	0.62	0.70	0.67	0.67	0.65
	(0.08)	(0.09)	(0.10)	(0.13)	(0.12)
OpVantage:	0.66	0.77	0.81	0.61	0.69
	(0.07)	(0.09)	(0.10)	(0.10)	(0.11)

Implications for capital

Including losses under $1 Million may yield capital estimates of about $2 Billion.

Table. 99.9 percentiles from simulated aggregate loss distributions.

	λ = 30 (Low freq.)	λ = 60 (Large bank.)	λ = 100 (High freq.)
b = 0.55 (Lower)	$0.4B	$0.6B	$0.8B
b = 0.65 (Est.)	$0.9B	$1.4B	$2.1B
b = 0.75 (Upper)	$2.4B	$4.0B	$6.0B

Note. Additional capital required to cover losses below $1 Million.

Implications for supervision

- External data important for supervisors to understand modeling
- Distinguishing bank from industry results will require modeling of external data
- Best practices is an important component of bank supervision - bank supervisors will need to continue to work with external and internal operational risk data