A review of the key issues in operational risk capital modeling

Mo Chaudhury
Desautels Faculty of Management, McGill University, 1001, Sherbrooke Street West, Montreal, QC, Canada H3A 1G5; email: mo.chaudhury@mcgill.ca

In an effort to bolster soundness standards in banking, the 2006 international regulatory agreement of Basel II requires globally active banks to include operational risk in estimating the regulatory and economic capital to be held against major types of risk. This paper discusses practical issues faced by a bank in designing and implementing an operational risk capital model. Focusing on the use of the loss distribution approach in the context of the Basel advanced measurement approach, pertinent topics for future research are suggested.

1 INTRODUCTION

According to the Basel Committee on Banking Supervision (2006, paragraph 644, p. 144), operational risk is defined as "...the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. This definition includes legal risk, but excludes strategic and reputational risk." Operational risk is highly company and operations specific, and unlike market, credit, interest rate and foreign exchange risks, a higher level of operational risk exposure is not generally rewarded with a higher expected return. Given the company and operations specific nature of operational risk, most often the exposure cannot be hedged with liquid instruments or in a cost effective manner. Although insurance is available for some types of operational risk (eg, damage to physical assets, business disruption and system failure, etc), the insurance policies can be quite expensive, may entail risks of cancellation or lack of compliance by the insurer, and there is a cap on regulatory capital relief for insurance of operational risk.

Examples of large and well-publicized operational risk events in recent times include: Barings Bank 1995 (US$1 billion), Long-Term Capital Management 1998 (US$4 billion), Société Générale 2008 (US$7 billion), and the terrorist attack of

The author is indebted to former colleagues at the State Street Corporation, Boston, MA, and to Shane Frederick for numerous discussions on operational risk issues. However, the contents of this paper reflect solely the opinion of the author. The valuable comments of an anonymous referee and seminar participants at the 2009 Pacific Basin Finance Economics and Accounting Association Conference in Bangkok, Thailand are also gratefully acknowledged.

September 11, 2001. These types of operational risk events have drawn attention to the fact that the exposure of financial institutions to operational risk could be as important, if not more, as their exposures to market, credit, interest rate and foreign exchange risks.[1] Concerns about the operational risk exposure of major financial institutions have further escalated due to the globalization of financial services, increasing complexity of financial products and the explosion in electronic trading and settlements. Accordingly, regulators of financial institutions, as embodied in the Basel II Accord, now require that financial institutions properly measure and manage their operational risk exposure and hold capital against such exposures. de Fontnouvelle *et al* (2003) find that the capital requirement for operational risk at large US financial institutions often exceeds the capital requirement for their market risk.

Despite the financial importance of operational risk exposure and the well-publicized incidences of operational risk events, operational risk related research remains at a meager level in the mainstream finance and management literature.[2] Although the array of statistical tools and the associated literature are rather extensive, the banks are, nonetheless, facing numerous implementation issues in their effort to comply with the Basel II regulatory framework. The goal of this paper is to identify and articulate a range of these issues in order to encourage further research that is directed specifically toward developing sound operational risk capital models for the banks. The importance of sound capital models in risk management of individual banks and in containing systemic risk can hardly be overemphasized.

We discuss the operational risk capital modeling issues for banks using the loss distribution approach (LDA).[3] Quantification of operational risk using the LDA under the advanced measurement approach (AMA) is a cornerstone of the Basel II Accord on the regulation and supervision of internationally active banks. Within some broadly defined guidelines and subject to the approval of the supervisory authority, the LDA allows a participant bank to use its internal model to characterize the probability distribution of potential aggregate operational losses over a one-year horizon. The difference between the 99.90th quantile and the expected loss, both calculated according to this distribution, constitutes the risk-based regulatory capital charge (RCAP) estimate. The economic capital is estimated the same way except that the quantile is the empirical probability of survival corresponding to a target credit rating.

Under the LDA, the severity distribution of loss from a single event is coupled with a frequency distribution of events over a given horizon, typically one year, to

[1] The market value impact of the operational risk events appears substantial (Cummins *et al* (2006), Perry and de Fontnouvelle (2005)).

[2] Cummins and Embrecht (2006) provide a summary review of the work in this area. Netter and Poulsen (2003) review the implications of operational risk to financial services firms, approaches to operational risk measurement and the role of the Basel II regulatory framework.

[3] Tripp *et al* (2004) discuss operational risk modeling for insurers.

arrive at the aggregate loss distribution for a given type of event over the horizon. The loss distributions for various types of operational risk events are then aggregated through the modeling of their dependence structure to generate the aggregate loss distribution for the bank as a whole. Rather than surveying the LDA-based operational risk literature, we provide an overview of the LDA in Section 2 and then highlight a range of modeling challenges faced by a practicing bank in implementing the LDA in the remaining sections. As such, we keep theoretical expositions at a minimum and focus more on the practical issues. According to the Basel II framework, banks need to make direct or indirect use of four types of datasets in estimating and/or validating their operational risk measures. The issues concerning operational risk datasets are hence discussed in Section 3, followed by the loss frequency and severity distribution matters in Section 4. Challenges in dependence modeling are taken up in Section 5. Finally, Section 6 contains a summary of key issues and some concluding remarks.

2 OVERVIEW OF THE LOSS DISTRIBUTION APPROACH

The LDA for operational risk is discussed in detail, among others, by Frachot *et al* (2001, 2003) and Yu (2005). Aue and Kalkbrener (2006) illustrate in detail how the LDA is applied to operational risk capital measurement at Deutsche Bank. Here we provide a summary overview of the LDA.

Consider a particular type of operational risk event, say processing error, in the retail banking business of a bank. The number of such errors, n, in a given a year is a random variable, commonly referred to as the frequency of an operational risk event. The dollar loss amount for the bank, S, when a processing error occurs, is also a random variable, and is called the severity of an operational loss event. The aggregate loss in a year due to processing errors in the retail banking business of the bank, $L = \sum_{k=1,n} S_k$, is therefore a random variable the probability distribution of which depends on the marginal distributions of frequency n and severity S and their dependence structure.

The operational risk capital, C (regulatory capital or economic capital), for the processing error in the retail banking business of the bank is then defined as $C = L_\alpha - E(L)$, where L_α is the α-quantile of the probability distribution of L and $E(L)$ is the expected annual loss. In other words, the probability is α that the annual loss due to processing errors in retail banking operations is less than or equal to L_α. The operational risk capital C is meant to cover the unexpected annual loss up to the amount $UL = L_\alpha - E(L)$. With $\alpha\%$ typically at 99% or above, the operational risk capital is designed to absorb extreme annual loss with a very high level of confidence.

For operational risk measurement, banks classify loss events into a limited number of units of measure. A unit of measure is the disaggregated level at which a bank starts distinguishing, specifying and then estimating the frequency and severity distributions. Basel II requires that all internal loss data be clearly mapped into seven Level I operational risk event types ($e = 1, 2, \ldots, 7$) and eight

Level I business lines ($b = 1, 2, \ldots, 8$).[4] If a bank follows this categorization for risk measurement as well, it will have 56 units of measure. However, subject to satisfactory mapping of the internal loss events, banks are allowed to chose different classification and as such units of measure for operational risk measurement (Basel Committee on Banking Supervision (2006, paragraph 673, pp. 152–153)). Say a bank selects M units of measure. Then, to estimate risk capital at the corporate or top of the house (TOH) level, the bank needs the distribution of annual TOH loss, $L_{\mathrm{TOH}} = L_1 + \cdots + L_M$. The operational risk capital for the bank is estimated as $C_{\mathrm{TOH}} = L_{\alpha,\mathrm{TOH}} - E(L_{\mathrm{TOH}})$. If no diversification is permitted across the units of measure, then the bank's operational risk capital, according to the LDA, hits the maximum amount, $C_{\text{undiversified}} = C_1 + \cdots + C_M$. The diversification benefit, $C_{\text{undiversified}} - C_{\mathrm{TOH}}$, critically depends on the dependence structure of the M annual losses.

While the LDA is conceptually appealing and straightforward, there are numerous issues in implementing the method. These issues may be classified into four main areas: datasets, annual loss distributions, dependence modeling and simulation. To implement the LDA, a bank starts with selecting/specifying frequency and severity distributions for each unit of measure separately, estimates these distributions, and combines the estimated distributions to arrive at an annual loss distribution for each unit of measure. This process is complicated by the Basel requirement that the bank directly or indirectly uses information from all four elements of operational loss data, namely, internal loss data, external loss data, scenario/workshop data and business environment and internal control (BEIC) data. Further, even when a bank starts with known parametric distributions for frequency and severity, most often the form of the resultant annual loss distribution for a unit of measure is not known. The next challenge is to aggregate the unit of measure loss distributions to the TOH loss distribution through the modeling of a dependence structure among the units of measure. Not only are the (unit of measure) marginal loss distributions varied and are often specified/estimated piecewise, there is the issue of whether the dependence should be modeled at the frequency, severity, annual loss level or some other level of aggregation.

In the remainder of this paper, we provide more details on the main issues related to data, distribution and dependence. In light of the complex data, distribution and dependence issues, and the fact that the risk capital involves an extreme quantile of the TOH loss distribution, it is obvious that the computational issues related to simulation will be daunting as well. However, we do not discuss the simulation issues in this paper.

[4] The seven Basel II Level I event types are: internal fraud; external fraud; employment practices and workplace safety; clients, products & business practices; damage to physical assets; business disruption and system failures; and execution, delivery & process management. The eight Basel II business lines are: corporate finance; trading and sales; retail banking; commercial banking; payment and settlement; agency services; asset management and retail brokerage.

3 ISSUES ABOUT DATASETS

3.1 Internal data

> "Internal loss data is crucial for tying a bank's risk estimates to its actual loss experience."
>
> [Basel Committee on Banking Supervision (2006, Paragraph 670)].

3.1.1 Classification of internal loss data

Loss events sharing the same economics and the same probabilistic nature should, in principle, be classified into a single unit of measure. Compilation of internal data set into units of measure that are compatible with both Basel II mapping and external data classification poses many challenges. In order to use the experience of their own operational loss events to better manage the risk of such future events, a bank is better off designing a customized event classification system that better reflects its unique operating model, control structure and risk monitoring mechanism. This, however, may pose difficulty in mapping into the standardized classification system of Basel II, especially for banks that are not typical large money center banks. To avoid punitive capital charges, the bank also needs to model less than perfect correlation among the units of measure. This, of course, has to be justified without the help of any external evidence due to the customized nature of the units of measure.

Further, an operational loss event could simply be such that it has important elements of more than one unit of measure. Since the event still has to be classified into a single unit of measure, the inevitable use of judgment may affect the quality of the internal loss data.

3.1.2 Length of internal data sample

The internal loss dataset for a bank may not be long enough to allow reliable estimation of the parameters of the frequency and severity distributions from the internal dataset alone. The importance of scenario/workshop data and external data is then enhanced. This could be problematic since elicitation of both frequency and severity assessment from the scenario/workshop participants over the entire range of losses becomes a formidable, if not a questionable exercise. External data, of course, may not be quite representative of the bank's operational risk profile.

3.1.3 Frequency in internal loss data

It is entirely possible that there is either no loss event at all in the internal dataset for some units of measure or the frequency appears abnormally low. Such a situation becomes more likely when the internal dataset is limited in length and the units of measure are defined at a more disaggregated level. Consequently, estimating the frequency and severity distribution for these units of measure will have to rely heavily on scenario/workshop and external datasets and will be subject to their limitations, especially for frequency estimation.

3.1.4 Severity in internal loss data

One of the well-known data biases is that internal datasets are typically biased toward low-severity losses. Operational risk capital, on the other hand, is meant to absorb low frequency large (high severity) losses and is thus more sensitive to accurate estimation of the loss distribution in the upper tail. Using the internal dataset alone for the estimation of severity distribution is thus likely to produce a too low operational risk capital.

Further, the loss data collection process is typically of poor quality for small losses and is not cost effective. Hence banks often collect internal loss data and record it into their dataset only if the size of the loss exceeds a threshold amount. This leads to a data bias known as the (left) truncation bias since the true frequency of losses below this lower threshold is not zero although it seems that way in the internal loss dataset.

3.1.5 Delays in reporting

The reported timing of the loss events in the internal dataset often lags the timing of detection and actual occurrence. Thus a reporting delay vector needs to be estimated that in turn injects measurement error especially in the frequency estimates derived from the internal dataset.

3.1.6 Protracted/split events

In some cases, either the event itself or the losses from an operational risk event extend over several quarters or sometimes years. The actual reporting of such events in the dataset may have a bearing on the frequency and severity estimates from the internal dataset. If a US$1 billion total loss is reported as four loss events of US$250 million each, the frequency goes up while the severity goes down. How the operational risk capital is ultimately affected is unclear though since in general higher frequency drives up risk capital while lower severity depresses it.

3.1.7 Mergers and acquisitions

When a bank acquires another banking operation, the assimilation of the two pre-acquisition internal datasets can pose challenging issues. For example, their data collection thresholds and units of measure may vary. To complicate matters, the acquired bank may be from a different country, thus adding foreign exchange and differential inflation rate considerations. Even more challenging is to project the frequency and severity distributions of the combined operations going forward, taking into account possible cannibalization, synergy and efficiency implications. Differences in corporate culture and employment practices could further impact the operational risk profile in event types such: as internal fraud; clients, products and business practices; and employment practices and workplace safety.

3.2 External data

> "A bank's operational risk measurement system must use relevant external data (either public data and/or pooled industry data), especially when there is reason to believe that the bank is exposed to infrequent, yet potentially severe, losses."
>
> [Basel Committee on Banking Supervision (2006, Paragraph 674)].

There are three well-known providers of external loss data for banks, namely the Fitch Group, the SAS Institute and the Operational Riskdata eXchange Association (ORX). Fitch and SAS construct their databases from publicly available information (media reports, regulatory filings, legal judgments, etc,) about operational losses over US$1 million, and the data can be purchased from these vendors.[5] The ORX data, on the other hand, is comprised of internal loss data of the member banks joining the consortium and is available only to its consortium members that are mostly European. The ORX members submit data to a common standard and format developed by the ORX Definitions Working Group. In the ORX database, the reporting threshold is €20,000.

The major challenge in using external data is the adaptation of the external loss data to the operational risk context of a specific bank. The adaptation is problematic, because operational risk events are quite idiosyncratic in nature. Factors that drive the operational risk profile (eg, size, organizational and management culture, human resources, geography of operations, technological infrastructure, risk assessment and control procedures, etc,) vary widely across financial institutions. As such, making sense of external frequency and severity information for a bank's own use requires the careful filtering and processing of such information.

3.2.1 Relevance of external data points

The main point of contention here is whether past loss events at other institutions seem likely or even plausible for the user bank going forward. For example, a US$7 billion operational loss (experienced by Société Générale in 2008) may not be a plausible event at all if the user bank's trading business is quite small and/or the internal control mechanism is much stronger relative to that of Société Générale. Of course the user bank may not have any trading business at all. The filtering decision, however, becomes less clear when the bank has a smaller trading book than Société Générale, but it is still sizable and is perhaps more leveraged.

What this means is that the bank has to go through a labor-intensive process of filtering external data points that is partly query-driven but largely judgmental in nature. This constitutes an unavoidable source of error in frequency and severity

[5] In the past, the OpVantage division of Fitch used to compile and manage the OpVar database. After the acquisition of the IC2 database, the integrated database OpVantage First is now offered through the Algorithmics division of Fitch. The OpRisk Global database was provided by the company OpRisk Analytics which has since been acquired by the SAS Institute.

estimation using external data. More often than not, the influential and controversial data points involve high severity. Hence, if the bank is too aggressive in filtering out external loss data points based on relevance consideration, the operational risk capital could be seriously underestimated. In the same vein, too much conservatism could lead to a punishing capital requirement.

3.2.2 Quantity of relevant data points

To start with there may not be enough data points in the external dataset that can supplement the internal loss data. The exercise of relevance-based filtering can only make this problem worse. From a practical point of view, it can thus often become a quantity versus quality trade-off in using the external dataset. Relevance-based filtering may reduce bias by eliminating irrelevant data points, but increase variance due to potential filtering error and clearly fewer data points.

3.2.3 Nature of the information in an external database

Vendors like Fitch obtain loss event information from public sources. While detailed in nature, such information may not be complete enough since it is not provided by the financial institutions concerned. Further, there is the potential for misclassification and reclassification of the events by the vendor. Additionally, the vendor data classification may not jive with the units of measure adopted by a bank, in which case direct use of the vendor data may not be possible. Alternatively, the bank may be forced to realign its units of measure to be able to directly use the vendor data.

Since the ORX data is based on common standards and formats, classification related errors are less likely. As ORX can perform custom-made analyses of the dataset, there is also more flexibility in using the external data for both direct use and for validation purposes. On the other hand, event-specific descriptions are quite lacking in the ORX data and hence ascertaining the relevance of high severity external data points is hindered considerably.

Another key limitation of the ORX data is that it may not contain important operational loss events simply because the banks concerned in these events are not members of the ORX consortium. The US$7 billion rogue trading loss at Société Générale will not be available in the ORX data since the bank is not a member of the ORX consortium. Along the same line, one may argue that the dearth of US financial institutions in the ORX consortium may limit the usefulness of the ORX data in capturing the operational risk profile of the US banks. Interestingly, the larger member banks may enjoy a capital relief using the consortium data, such as the ORX data, since the loss experience of the smaller banks may dilute the operational risk profile of the larger member banks.

3.2.4 Reporting bias

A major problem with publicly available external loss data is that not all loss events reach the public domain. As the extent of under-reporting could be influenced by many factors, it tends to vary across the various operational risk event types (de Fontnouvelle *et al* (2003)). The most well-known pattern about the under-reporting phenomenon is that publicly available external data has a strong bias in favor of large and well-publicized losses. As a result, the under-reporting phenomenon is likely to bias the operational risk capital upward when publicly available external data is directly used.

To be more specific, there are three types of problems associated with the under-reporting phenomenon. First, knowing that all losses are not reaching the public domain anyway, public data vendors impose a known threshold in the compilation of their database. For Fitch data, this threshold is US$1 million. This problem is similar to the case of a truncated dataset at a known truncation point (here left truncated at US$1 million). In fact, since most banks do not collect and record internal loss data below an internally imposed threshold, typically around US$5,000 to US$10,000, internal datasets are also left-truncated. In the ORX data, the lower threshold of €20,000 is common across all member banks; but in publicly available external data such as the Fitch data, this company-specific lower threshold is unknown and random from a given bank's perspective.[6] Frachot and Roncalli (2002) and Baud *et al* (2002) describe how internal data can be compared with external data having different collection thresholds. For a variety of severity distributions (lognormal, lognormal-gamma, generalized Pareto and Burr), Mignola and Ugoccioni (2006) show that the effect of a known data collection threshold on the extreme quantiles is minimal for threshold values up to €100,000.

Second, while the compilation threshold/truncation point used by an external data vendor such as Fitch is known, the threshold/truncation level of loss size above which loss events reach the public domain is not known. This unknown reporting threshold for a given event type poses the most challenging problem with regards to the under-reporting phenomenon.[7] Intuitively, the higher the unknown threshold, then the greater the extent of the under-reporting phenomenon. That is, the reported losses will appear more skewed towards higher severities than the true losses are.[8]

Third, the unknown reporting threshold is likely to vary across event types and there is no obvious way of relating the under-reporting bias correction efforts to the characteristics of the event types.

3.2.5 Scale bias

A clear positive for using consortium external data, such as the ORX data, is that by construction such data minimizes the above mentioned reporting bias problems.

[6]Ignoring the conditional nature of the data leads to biased value-at-risk (VaR) estimates (Chernobai *et al* (2004)).

[7]For various approaches to addressing this under-reporting bias, see Frachot and Roncalli (2002), Baud *et al* (2002), de Fontnouvelle *et al* (2003), Guillen *et al* (2006) and Buch-Kromann *et al* (2006).

[8]Since expected losses will also be affected, one cannot be quite sure about the net effect of the reporting bias on capital. The nature of the severity distribution could have important bearings too.

However, in publicly available external data as well as in consortium data, there still remains the problem of proper scaling of the external data. The external loss data concerns banks of different sizes (size-related proxies include asset, revenue, transaction volume and number of employees) and geographic regions, and creates the fundamental problem of comparability of external loss data. Unless the size differential is properly accounted for, the frequency and severity distributions for a bank of a specific size could be biased and imprecise.

Using both internal data from different business units and publicly available data from other institutions, Na *et al* (2006) find a strong relationship between operational losses and gross revenue that is well-explained by a scaling power law. Applying the quantile regression technique to selected the ORX data, Cope and Labbi (2008) find that large losses scale differently from small losses for a number of business lines and event types, and loss severity is greater in Western Europe compared to North America. Shih *et al* (2000) and Dahen and Dionne (2010), among others, also explore the scaling issue.

3.3 Scenario/workshop data

> "A bank must use scenario analysis of expert opinion in conjunction with external data to evaluate its exposure to high-severity events."
> [Basel Committee on Banking Supervision (2006, Paragraph 675)].

Expert opinion data on loss frequency and severity distributions and on correlation of losses is commonly referred to as scenario data. As expert opinion data is typically gathered through workshops, this type of data is alternatively called workshop data. Internal and external loss data capture what operational loss events a bank or its peers experienced in the past. However, there may be events that seem possible for the bank going forward that neither the bank itself nor its peers have experienced in the past. Suffice to say, the 2007–2009 crisis that gripped the banking world attests to the need for thinking beyond the past experience and for conditioning probabilistic assessments on the fast moving business dynamics. The expert opinion data collected in relatively recent workshops can thus help fill a potential gap in historical loss data, internal and external.

While some banks use expert opinion for validating extreme loss and capital estimates (for various units of measure) based on internal and external loss data, others choose or are advised by their regulators to directly incorporate expert opinion data in deriving the extreme loss and capital estimates. Obviously direct use of the expert opinion data is more consequential and by far more challenging, and this is our focus here. For illustrative purpose, consider the expert opinion data of a bank for the loss distribution of a given unit of measure collected through a workshop of N internal experts. The key issues here concern designing the workshop to minimize the effect of behavioral/cognitive biases and obtaining information in a way that facilitates proper assimilation with other data (internal, external, business environment and internal control).[9]

3.3.1 What should the opinion be about

It is a common practice to estimate frequency and severity distributions separately when using internal and/or external data. Thus, from a data assimilation perspective, it might make sense to seek opinion about the frequency and severity distributions instead of the loss distribution. Further, seeking expert opinion directly about the loss distribution can be quite taxing on the experts since the loss distribution is seldom captured by well-known closed form distributions.

Note that what is typically referred to as frequency distribution is the distribution of the frequency of any loss, that is, the frequency of loss exceeding zero. However, seeking expert opinion on this can pose a significant practical problem if the frequency distribution estimated from the internal loss data is widely divergent from the expert opinion. Further, a key purpose of scenario data is to supplement internal loss data with expert assessment of the prospect of higher severity losses that are typically of lower frequency. Accordingly, if at all, expert opinion is usually sought about the frequency of higher severity losses, eg, the frequency of a loss greater than US$0.5 million, the frequency of a loss between US$500,000 and US$1 million, etc.

In the context of frequency and severity assessments, complexity arises due to the conditional nature of the severity distribution. What the severity distribution describes is the prospect for losses of different sizes conditional on the fact that a loss greater than zero (an operational loss event) has taken place. The probability of a loss (severity) equal to or greater than S_t, therefore, depends on the frequency of a loss equal to or greater than S_t as well as the frequency of an operational loss event (a loss greater than zero) taking place. Heuristically:[10]

$$\text{Prob}(S_k > S_t \mid S_k > 0) = \lambda_{1/t}/\lambda$$

In the above equation, λ is the expected frequency of an operational loss event ($S_k > 0$) in a year, that is, $E(n) = \lambda$, and $\lambda_{1/t}$ is the expected frequency of a loss equal to or greater than S_t in a year, with $\lambda_{1/t} < \lambda$. In operational risk parlance, we may expect to see a loss equal to or greater than S_t once every t years, given that λ events are expected in a year. If the expected (arithmetic average) length of intervals between losses is equal to or greater than S_t is t years, then $\lambda_{1/t} = 1/t$.

Note that given the expected frequency of operational loss in a year (λ), an assessment of the expected frequency ($\lambda_{1/t}$) of a loss equal to or greater than S_t is

[9] One of the most well-known works on behavioral bias is that of Kahneman and Tversky (1979). For recent research, see, for example, Kahneman and Frederick (2006) and prior works of these authors. For a generic list of cognitive biases, please see http://en.wikipedia.org/wiki/List_of_cognitive_biases.

[10] The frequency distribution is assumed to be Poisson and frequency and severity are assumed to be independent. This will be discussed further later in this paper.

essentially an opinion on the conditional probability of severity of a loss equal to or greater than S_t, that is, the integral of the severity distribution to the right of S_t. As such, a series of assessments by varying t is equivalent to describing the right tail of the severity distribution for severity corresponding to the lowest t and above.

Given the dependence of the severity distribution on the frequency assessments, a critical choice is to determine the lowest t; that is, the targeted right tail of the severity distribution for which expert opinion is sought. Given the emphasis on extreme loss prospects in workshop data, a natural choice for t is one year; that is evaluating the prospects for losses that are expected to occur once a year or less frequently. However, for a given unit of measure of the bank, the frequency of operational loss events in a year (λ) may be low and as such the $t = 1$ year severity threshold may be too low for expert assessment. On the other hand, varying the lowest t across various units of measure may be confusing to the workshop participants and also pose a problem in evaluating the reasonableness of expert assessments. The choice of the lowest t or targeted right tail of the severity distribution may also be influenced in an important way by severities observed in the internal and external data and the length (years) and number of loss data points available in these data sets.

Once the tail segment of the severity distribution is targeted, the next issue is whether expert opinion should be solicited about the key statistics or the quantiles of the severity distribution. For example, if key statistics are sought after, then workshop participants will be asked to provide their opinion about the expected losses given that the loss exceeds various levels such as US\$500,000, US\$1 million, etc. On the other hand, if quantile information is sought after, then the workshop participants could be prompted about S_t for various values of t or $\lambda_{1/t}$ starting with the lowest value of t selected, or about t or $\lambda_{1/t}$ with various levels of S_t specified. Another workshop design is to divide the severity domain into several buckets and then ask the workshop participants to assign the percentages of times they expect the severity of a loss to fall into different buckets.

The choice of the way the probabilistic assessments should be obtained is not clear. One main reason for this is that the directions and the net effect of the various behavioral biases associated with the way the probabilistic assessment is sought is unclear and not a lot of published research in the operational risk workshop context is available in this regard. A second reason is that not much is known publicly about the comparative statistical properties of the estimated parameters of various distributions (especially the fat tailed ones) fitted to the different types of workshop responses (conditional expected loss, S_t, t, $\lambda_{1/t}$). The ultimate implications of how the level and the stability of the estimated operational risk capital are affected are also unclear. It is possible that the choice of the way the assessments are sought could be dictated by whether the bank's objective is to minimize the bias or the variance of the operational risk capital or some other objective function. Yet another compounding factor is how the workshop data is assimilated with the internal and external data sets.

Another controversial but consequential issue is how far into the tail of the severity distribution the assessments should go. For example, should the assessments go up to $t = 50$ or $t = 100$ or even lower frequency events. Risk capital is about infrequent large losses, but human judgment may become cloudy and erratic in fathoming extremely infrequent events. Some argue that workshop participants are conditioned by their lifetime work experience and as such their estimates beyond $t = 50$ become questionable. However, for a given unit of measure at a bank, the expected annual frequency of operational loss events (λ) may be low and hence $t = 50$ may not correspond to a high quantile of the severity distribution, and in that case the workshop purpose of obtaining tail assessment is defeated. The opposite case of going too far into the tail may occur when the expected annual frequency of operational loss events (λ) is high. The choice of maximum t is also importantly linked to how the scenario data is assimilated with internal and external data. For example, if external data is to be used for fitting the severity distribution beyond $t = 25$ type of events, then $t = 25$ may be chosen as the maximum t in the workshop.

Workshop designers also debate whether the experts should be asked about the existence of a cap for the severity for a given unit of measure and if so an estimate of this maximum loss amount. Of course, severe reconciliation problems may arise in situations such as the maximum loss estimate of one participant being way below the median severity estimate of another participant.

3.3.2 What information about internal and external data should be provided

Providing information about other data may bias the participants toward conformity with such information. On the other hand, information about other data may help the experts to place their assessments into proper perspective and thus improve the quality of their assessments. Also, the experts may already be aware about such information, especially about the BEIC data. As such, some banks may decide to use the BEIC data only indirectly by informing the participants in detail about the BEIC data. The downside of this approach is that the marginal impact of workshop data on the bank's operational risk capital estimate cannot be disentangled from the influence of the BEIC data.

It seems reasonable that some limited information about internal and external data should be provided to the workshop participants. In providing such information, the workshop design must strike a balance so that the workshop responses are not deliberately driven toward lowering operational risk capital estimates.

3.3.3 How the workshop is conducted

This is a crucial aspect of the workshop design that is potentially subject to a number of behavioral biases and can have important bearing on the nature of the workshop data generated. To start with, how many experts is the right number of experts is unclear. With more participants, the effect of outlier responses becomes less influential and can improve the statistical properties of the fitted distributions. However, it becomes more difficult to have a meaningful debate among the workshop participants about the operational risk profile.

A second issue is whether the workshop should be designed to achieve convergence of participants toward a single response for each of the workshop queries, eg, S_t, $t = 5$, 10, 20, 50. One outcome of convergent estimates is that a statistical distribution may be fit exactly to the workshop results. A clear price to pay for convergent estimates is the loss of flexibility in fitting meaningful severity distributions. On the other hand, with widely divergent opinions, fitting a single severity distribution to the workshop responses may be quite daunting a task.

A third issue is whether a single round or multiple rounds of estimates should be attempted in a workshop. Obviously, if convergence is targeted, multiple rounds of assessments are needed. However, as the debate between rounds progresses, there is no guarantee that the participants will change their mind and even worse is the case where the later range of responses becomes more divergent or outlying. A related workshop design issue is whether the individual responses should be collected in a discrete or public manner. Revelation of responses may lead to more careful thoughts by the participants, but in the context of multiple rounds, revelation may unduly pressure some participants to change their opinion to conform to others and as such the opinions may become too convergent.

Among many other issues, one that is worth noting is the issue of whether the participants should be informed about the behavioral biases. Some digression on probability assessments is most often considered quite desirable. However, the merit of informing the participants about the potential biases they may have is debatable.

4 FREQUENCY AND SEVERITY DISTRIBUTIONS

> "...a bank must be able to demonstrate that its approach captures potentially severe 'tail' loss events."
>
> [Basel Committee on Banking Supervision (2006, Paragraph 667)].

In the popular bottom up approach to the LDA method, a bank needs to decide what frequency and severity distributions to use for each individual unit of measure. Usually the same form of frequency distribution, albeit with different parameter values, is used for all units of measure. However, the form of severity distribution normally varies significantly across the units of measure. While there are numerous modeling and estimation issues, especially with respect to the tail of the severity distribution, in what follows in this section we draw attention to some of the more generic ones.

4.1 Dependence within the unit of measure

Dependence here refers to the intra-year dependence between successive events belonging to the same unit of measure. Most banks use a maintained hypothesis that the frequency and severity distributions are independent for an operational risk event, and severities across various events within the year are independent as well.

4.2 Frequency distribution

The three closely related alternatives for the frequency distribution are the binomial, the Poisson and the negative binomial. As it turns out, the frequency distribution is not nearly as consequential for the operational risk capital, many practitioners prefer the parsimonious nature and the analytical convenience of the Poisson distribution.

An important practical issue is the choice of the dataset to estimate the frequency distribution. As mentioned by Aue and Kalkbrener (2006), data completeness is essential for frequency distribution and as such banks may be inclined to use internal loss data for frequency estimation. This also makes sense intuitively as the internal control and risk management processes of a bank may set its frequency distributions apart from those at other banks. On the other hand, if sufficient internal data points are not available for a given unit of measure, a bank needs to explore frequency estimation from external data, and in this context, consortium data is likely more useful than publicly available external data. Another alternative is to weight frequency distribution parameter estimates from internal, external and possibly scenario datasets. However, the bank needs to determine the weighting scheme and rationalize such a scheme before the regulators. One approach, other than ad hoc weighting, is to combine various datasets using the Bayesian approach (Shevchenko and Wüthrich (2006)) or the similarly motivated credibility theory (Frachot and Roncalli (2002); Bühlmann *et al* (2007)). For example, in a Poisson-gamma mixture, the gamma distribution could be the Bayesian prior for the stochastic intensity, λ; the prior distribution is estimated possibly from the external dataset or the workshop dataset.

4.3 Single severity distribution

The most challenging task in operational risk capital modeling concerns the construction of the severity distributions for the various units of measure at a bank. A starting point for modeling severity distribution is that a single parametric distribution describes the probabilistic behavior of severity over its entire range. The key benefit of such an approach is that it is relatively easier to estimate the parameters of such a distribution using internal and external (and possibly workshop) loss datasets simultaneously. The assumption here is that all loss datasets are drawn from the same distribution although, as pointed out by Frachot and Roncalli (2002), it is important to recognize the effect of any unknown reporting threshold(s) of external loss data. A traditional maximum likelihood estimation (MLE) method can then be used to estimate the parameters of the severity distribution including the unknown reporting threshold(s) of external loss data.

There are two key problems in using a single severity distribution over the entire range of severity. First, it is unlikely that the internal and external loss data are drawn from the same underlying severity distribution. Second, there is a general recognition (eg, Wei (2006)) that a single parametric distribution is inadequate to capture the probabilistic behavior of severity over its range. It is a widely held view among practitioners that the severity distribution for the extreme losses, that is those in the tail, behaves differently and this behavior is better captured by heavy-tail or fat-tail parametric distributions (Moscadeli (2003); de Fontnouvelle *et al* (2004)).

Although there is no universal definition of what is a fat- or heavy-tail distribution, one criterion based on maximal moment is that a distribution is a light tail one if finite moments exist for all orders, otherwise it is a fat-tail distribution. Accordingly, de Fontnouvelle *et al* (2004) catalog exponential, Weibull, gamma and lognormal distributions as light-tail, and log-gamma, Pareto, generalized Pareto (GPD), Burr and log-logistic distributions as fat-tail ones. However, such a classification is not unique. For example, Dutta and Perry (2007) classifies lognormal as a heavy-tail distribution. In the end, whether a distribution is a heavy tail one depends on the thickness of the tail (above a large threshold).

4.4 Piecewise severity distribution

A natural consequence of the recognition that the form of the severity distribution varies over different ranges of severity is that the severity distribution that applies to the entire range is a piecewise one that results from concatenating/splicing different forms of severity distributions at the threshold(s) separating the distinct severity ranges. With a single threshold, the severity range below (above) the threshold is commonly called the "body" ("tail") of the piecewise severity distribution. The circumstances for some units of measure may lead a bank to pursue a three-piece distribution, for example, "body", "torso" and "tail" ("body", "tail" and "extreme tail").

Among the practical implementation issues, determining the number of thresholds (pieces) and more importantly the level of these thresholds is a difficult one. An imprecise but easy solution is to impose exogenously decided threshold levels. However, in this case, a bank needs to estimate the quantile of the overall severity distribution that each of the thresholds represents; that is, a tail probability scheme is to be applied. Avoidance of arbitrary schemes requires frequency estimates above and below the thresholds, an exercise that becomes daunting with higher values of the upper thresholds, especially with the reporting bias of publicly available external data.[11] Also, to ensure smoothness of the piecewise severity distribution at the threshold(s), the density estimates from the distributions below and above need to be equated. It is, however, possible to model endogenous threshold(s) or joining points for the pieces of the piecewise severity distribution.[12] If the choice of the severity distributions for the pieces were relatively stable, then a bank could try to

[11] In the study by Wei (2006), the Poisson frequency parameter and the lognormal severity distribution parameters below the threshold of US$10 million are exogenously specified, based on prior banking studies. In essence, this translates to an arbitrary total frequency and hence arbitrary probability assignment to the tail (above US$10 million) of the piecewise severity distribution.

[12] Only by chance, the endogenously estimated joining point(s) would coincide with the external data reporting threshold(s).

estimate the joining points simultaneously with the parameters of the distributions. In practice, such invariance of the distribution types and their parameter values with respect to the joining points or threshold levels is not easy to find. This issue of exogenous versus endogenous thresholds or joining points appears to be a promising and useful area of further research.

With a given threshold strategy, the next important issue in estimating the piecewise distribution is to determine how the different datasets will be used to estimate the severity distributions of the various segments ("body" and "tail" or "body", "tail" and "extreme tail"). A justifiably popular view is that the internal loss data is the most suitable for estimating the severity distribution in the "body" to the extent there are sufficient number of internal loss data points in this range for the unit of measure concerned.[13]

With this accepted view, the next set of issues are whether external data and workshop data are to be used directly for estimating the "tail" (and perhaps "extreme tail") severity distribution, whether internal loss data should also be used and how should the datasets be used and/or combined. If external and workshop data are used for validation purposes, then only internal loss data is available for estimating the "tails". With the widely accepted view that exclusive reliance on internal loss data would bias the operational risk capital downward, it is more realistic for most banks to directly use at least one of the expert opinion and external datasets in constructing the "tail" severity distribution(s). One alternative is to construct the expert opinion dataset such as to fill up gaps or paucity of the external loss data points. A more burdensome second alternative is to ask experts to assign quantile percentages to external loss data points. A third alternative is to use workshop data to construct the middle piece ("tail") of a three-piece severity distribution and to use external data to construct the extreme piece ("extreme tail"). Yet another alternative is to use expert opinion to determine the single joining point in a two-piece distribution above which only the external data is used to construct the "tail" piece. Again, further research is called for to explore the implications of alternative uses of the workshop and external datasets for the level and stability of operational risk capital estimates and the statistical properties of the "tail" distributions and their parameter estimates.

4.5 Parametric or empirical

In the discussions above, we implicitly assumed that the piecewise severity distribution segments will be constructed using parametric distributions. While this is the norm in the industry, a bank may opt to use the empirical (non-parametric) distribution for at least some units of measure. The use of the empirical distribution leads to the use of a historical simulation method to simulate severity. This is an

[13] Although it sounds strange, it is not unimaginable that a bank either did not experience any, or had too few, low severity operational loss events in some units of measure, for example, internal fraud.

open issue as to how to evaluate the suitability of using empirical distribution versus parametric distribution in the context of fitting the piecewise severity distribution.

4.6 Estimation of the tail

The extreme value theory is the branch of statistics that focuses on the study of rare events which in the case of operational risk concerns the tail of the severity distribution. According to the Pickands–Balkema–de Haan theorem of extreme value theory, under some regularity conditions, as the threshold increases, the distribution of excess losses over the threshold approaches the GPD for many underlying severity distributions. The most popular estimation method for GPD is the peak over threshold (POT) approach that simply uses the sample severity points above the threshold for estimating the GPD. Thus, to reliably capture the tail behavior with GPD using the POT estimation method, the threshold needs to be large enough and there still needs to remain sufficient severity data points above this large enough threshold. Accordingly, even with publicly available external data such as Fitch, the threshold for GPD estimation may need to be larger than the vendor's compilation threshold of US$1 million.

One clear problem in this context is that there may not be sufficient data points to estimate the GPD reliably, especially with internal loss data and with consortium data that is not comprehensive in terms of years and bank coverage. This is of course similar to the well-known peso problem; by definition, rare phenomena occur seldom and are in short supply. In the case of POT estimation of GPD, a bank can lower the threshold to enhance sample size in the tail and thus the lower the variance of parameter estimates. However, this will very likely create a bias since GPD is a limiting distribution only for sufficiently large threshold levels. A related problem is that sizable gaps in the sample range of high-severity data points may lead to significant widening or shrinking of the "body" (in a two-piece severity distribution) which in turn can create instability in the severity distribution for the "body". Yet another issue is whether to retain or drop influential data points in the process of external data filtering. It is also worth mentioning that a bank would have to sacrifice the flexibility of varying the threshold for POT estimation purposes if it decided to go with endogenous joining points for the piecewise severity distribution. This creates another trade-off that needs to be evaluated in further research.

Lastly, the POT-based estimate of high quantiles may be slow to converge (Degen and Embrechts (2008); Degen *et al* (2007)) and therefore may produce unstable risk capital estimates.

4.7 Issues with estimation methods (internal and external loss data)

Since internal and external loss data are commonly subject to collection and reporting thresholds, dependence arises between the parameters of the frequency and severity distributions (Ergashev (2007)) and calls for joint estimation of their parameters. In practice, however, the Poisson distribution is widely used for frequency distribution and its parameter is estimated independently from the internal and/or external loss data. The focus then shifts to estimation of the severity distribution. While non-parametric or empirical distributions are a possibility, parametric severity distributions are decidedly more popular in practice. In this case, the parameters of alternative severity distributions are estimated from the internal and/or external loss data. Here, by far the most popular choice is the MLE method. Alternative estimation methods include the quantile estimation method, the distance based methods and the method of moments.[14]

One area of concern for the MLE is that the loss observations are weighted in accordance with their likelihood under the presumed distribution (eg, exponential distribution). Since internal loss data is typically characterized by low severity losses, the MLE parameters are overly influenced by these losses for reliable estimation of the high quantiles that dictate the operational risk capital. In the case of external loss data, on the other hand, the MLE parameter estimates may be unduly influenced by the relative preponderance of high-severity losses in some units of measure. This issue remains alive when the MLE is used to estimate the parameters of an extreme value distribution (eg, generalized Pareto) for the exceedance of tail or high-severity losses above a high threshold.

To enhance the focus on quantile estimates, one could use the quantile estimation method that matches a set of quantiles of the empirical loss data and the parametric severity distribution. Dutta and Perry (2007) use this method to estimate key parameters of the g-and-h distribution. While the quantile estimation method may be suitable for transform distributions such as the g-and-h distribution, Dutta and Perry suggest that the method may not be appropriate for all distributions. As an alternative, Ergashev (2007) proposes several distance-based methods that hold promise in reducing bias in estimating capital.

Instead of quantiles, the method of moments works with moments of the severity distribution. This rather simple method solves for the parameters of the presumed distribution by equating the theoretical or population moments (functions of the parameters) and their sample estimates.[15] In the operational risk context, there are two critical limitations to this method. First, the operational risk capital estimates may be unstable as the loss data sample changes, especially so with small samples. Second, reliability of the high quantile estimates is questionable.

Instead of moments, the generalized method of moments (GMM) focuses on a set of conditions the expected value of which involves the parameters. These conditions are known as the moment conditions. In the context of the parametric

[14]For issues about suitable estimation methods and their properties, please see Embrechts *et al* (1997), Embrechts *et al* (2002) and Wei (2006).

[15]Generally speaking, operational loss severity exhibits high kurtosis and right skewness and of course tail heaviness.

severity distribution, the GMM moment conditions could simply be the equation of empirical and theoretical moments as in the method of moments. Alternatively, one could use conditions that are satisfied by moments of various orders. In either case, the specialty of GMM is that the various moment conditions can be weighted. In the context of operational risk capital modeling, the choice of the moment conditions and the weighting scheme of GMM might hold some promise for reliable estimation of the high quantiles. However, more research using both simulated and empirical loss data is needed to better evaluate the prospects of the GMM in the context of severity distribution and capital estimation.

One final issue is the paradoxical nature of the robustness of the severity estimation method in the context of operational risk capital modeling. Robustness can be considered in terms of the bias and consistency, and variance or standard error of parameters and VaR or capital.[16] Since capital estimates can be very sensitive to the shape of the distribution in the tail implied by the estimated severity parameters, the undesirable consequence is that the operational risk capital estimates can be quite unstable (Mignola and Ugoccioni (2006)) depending on the sample data points and the estimation method used for severity distribution. The robustness paradox in this context (Chernobai and Rachev (2006)) is that the low-frequency high-severity data points that would traditionally be considered as outliers compared to the bulk of the loss data are also the most relevant for estimating the very high quantiles that determine the risk capital. As a result, while the commonly used methods of trimming and Winsorizing to achieve robustness may reduce the chances of overestimating the risk capital, they may also lead to underestimation of risk capital. These methods have low break-down points and as such a small degree of data contamination (too low or too high) can sway the severity distribution parameter and consequently capital estimates. Instead of trimming, one could consider the use of influence functions to detect influential data points (Chernobai and Rachev (2006)) to screen loss data points. These methods have high break-down points and parameter estimators can tolerate more contamination.

Clearly, more research is called for in improving the robustness of loss severity parameter estimates and the consequent risk capital estimates considering of course the paradoxical nature of robustness here. One possible direction is the use of Bayesian prior. For the lognormal-gamma model of severity distribution, Ergashev (2009) reports a substantial reduction in statistical uncertainty around capital estimates using the Markov chain Monte Carlo method and Bayesian prior on model parameters.[17]

[16]Embrechts *et al* (2003) indicate the minimum number of observations needed for the lognormal and Pareto (extreme value theory) to have accuracy in terms of bias and standard error of VaR at high confidence levels. With high minimum number of observations, robustness requires lower threshold for the POT method and/or lesser trimming. However, extreme value theory becomes questionable if the threshold is low.

[17]Recent works on the Bayesian approach to operational risk modeling include Peters and Sisson (2006), Shevchenko and Wüthrich (2006) and Alexander (2000).

4.8 Estimation issues with workshop data

Suppose that the bank has conducted probabilistic assessments that are directly usable for estimating the severity distribution and the bank has decided to directly use the workshop data. There can be a host of issues associated with such a pursuit.

If the workshop participants were asked to allocate percentages (summing to 100%) to various severity buckets covering the entire severity range, then essentially each participant is providing a discrete severity distribution. Assume that all the participants have the same underlying distribution (type and parameter values) in mind, but they can only guess the parameters with some error. Then, the workshop data can be viewed as similar to data on alternative transport choices or automobile purchases in various cities. Thus, as in the case of quantile choice samples, the MLE method can be applied to fit various parametric distributions, one at a time. Note, however, that the number of workshop participants is typically not that large and as such the small sample limitations of MLE estimates may arise. But more importantly, if the bucket allocations of some participants are too far apart, then the MLE estimation may stall due to a lack of convergence. Also, initial values of the parameters may influence the parameter estimates and of course there is no assurance of a global maximum in optimization.

An opposite problem arises when the participant opinions are too similar. In this case, any participant who holds a sufficiently different opinion becomes an outlier and in a small sample the outlying opinion can sway the parameter estimates considerably if convergent estimates are at all feasible. This raises the research issue of how to detect influential data points in a workshop data context and how to trim the outliers.

An extreme case of workshop data is where the workshop design leads to unanimous opinion. In this case, the bank can fit a two parameter distribution exactly using alternate pairs of buckets. The issue is then how to select or combine the parameter values obtained from various pairs of buckets. For assimilation with internal and external data, it is important that the process leads to economically sensible parameter estimates.

Now consider the workshop design where the participants are asked about the severity levels associated with a number of, say five, conditional quantiles of the severity distribution. For example, the query could be about the severity levels (S_t) corresponding to $\lambda_{1/t} = 1/t$, $t = 10, 20, 50, 100, 1,000$. Unlike the design of assigning conditional probabilities to buckets surrounding given severity levels (conditional quantiles), here the design solicits the S_t given the conditional probabilities (assuming a common λ for all participants). Quite distinctly now the participant opinions are about statistics (quantiles) that describe the severity distribution, and as such the workshop data points cannot be treated as sample severity data points (as in internal and external loss data) or as relative frequencies (as in the workshop exercise of bucketing). An interesting and potentially promising interpretation of the workshop data points in the current case is that they are like alternative estimates of moments of the underlying severity distribution. Accordingly, a bank might wish to entertain estimation methods such as the generalized method of moments. The GMM approach allows weighting schemes based on the variance-covariance structure of expert opinion errors across the S_t and/or across the participants. However, not much published research is currently available in this regard.

4.9 Selection of severity distributions

Much is known about the statistical distributions that are well-suited for operational risk modeling, but selecting the most suitable distribution remains as intriguing as ever. It is well-known in the operational risk literature that the distribution type matters a lot for the estimated level of operational risk capital and its stability.[18] As such model (distribution type) selection is an important practical issue in operational risk capital modeling.

If some distributions are nested within another, then some likelihood-based criteria (eg, likelihood ratio test, Akaike information criterion, Schwarz Bayesian criterion) may be applied. Wei (2006, p. 22) applies to non-nested models, an idea from Klugman *et al* (2004) that evaluates the difference in log-likelihood versus the number of parameters. Wei notes that at the 5% significance level, the log-likelihood needs to increase by at least 1.92 for an extra parameter and by at least 3.00 for two additional parameters. It is not, however, clear how we could compare two non-nested models with the same number of parameters.

An alternative model selection method is based on the distance between the true or empirical density and the candidate parametric densities. Gustafsson and Thuringb (2007) consider three such distance estimators based on absolute distance, squared distance and weighted squared distance. One problem of applying such estimation methods in practice is that not enough data points are available in the tail for reliable empirical or non-parametric density estimation in this range where it matters most for operational risk capital. Also, non-parametric density estimators may not be reliable in extrapolating beyond the range of estimation.

In practice, the current state of model selection is more of an art than a science. It is not unusual for banks to use eyeballing Q–Q plots; that is, plotting of empirical quantiles to quantiles based on alternative parametric distributions. Additionally, banks evaluate the economic reasonableness of the projected capital estimates based on alternative distributions.

Recently Cohen-Cole and Prono (2007) have proposed a potentially appealing concept of model averaging. Instead of selecting "a model", perhaps a bank should consider using different models each weighted by its support from the various loss

[18] Wei (2006, pp. 6–7) notes that the tail index from POT varies substantially for various units of measure and across studies and thus indicates that the best tail distribution varies too. Typically, the lognormal and Weibull distribution perform well in fitting a large range of severity data, but their fit is generally poor in the "tail". On the other hand, distributions such as GPD perform better in the tail but they are weak in the "body".

datasets. Interestingly, this approach has further promise in combining models that tend to minimize the variance of estimated operational risk capital.

5 DEPENDENCE MODELING

> "Risk measures for different operational risk estimates must be added for purposes of calculating the regulatory minimum capital requirement. However, the bank may be permitted to use internally determined correlations ... provided ... its systems for determining correlations are sound ...."
>
> [Basel Committee on Banking Supervision (2006, Paragraph 669(d))].

In a typical bottom up LDA application, the marginal loss distribution L_m for the M individual units of measure are first estimated, as we have discussed previously. It is the modeling of the dependence structure that connects the marginal distributions at the unit of measure level to the aggregate loss distribution at the top of the house level. If operational losses across the bank's units of measure are perfectly dependent, then a bank's operational risk capital is the sum of the operational risk capitals estimated individually for the units of measure. This usually leads to punitively high operational risk capital for a bank. If, on the other hand, dependence of losses across the units of measure is imperfect, then normally there is diversification benefit as small losses in some units of measure are expected to keep a lid on the top of the house operational loss. The operational risk capital relief that results from imperfect dependence can be quite substantial.

5.1 Dependence of what

Relatively speaking, easiest to handle is the dependence of the frequencies across the units of measure and the most challenging is the modeling of dependence between the severities.[19] Frachot *et al* (2004) show that if the severities between two units of measure are dependent, then the severity draws within the units of measure cannot be independent; independence of the severity draws for a unit of measure is of course a maintained assumption in most LDA applications. Intuitively also, it makes more sense to expect dependence of the frequencies across the units of measure. For example, if frequency is related to the size of the bank's operations or market volatility, we may expect dependence between the frequencies of different event types or business lines. Importantly, for frequency dependence, the loss events of different units of measure need to occur only within the same year. For severity dependence, however, proximity in their timing is needed since severities are losses conditional on specific events taking place. Further, it becomes quite challenging to simulate dependent severities across various units of measure. In

[19] Although dependence between the frequency of the unit of measure and the severity of another unit of measure is a possibility, in practice this is rarely considered if at all.

terms of joint distribution, as a matter of practice, the frequencies of different units of measure are assumed to follow the same type of distribution, most commonly the Poisson distribution, albeit with different values for the parameter λ. With Poisson marginal distributions, a multivariate Poisson distribution can be created by adding the dependence structure of the frequencies.[20] By comparison, (marginal) severity distribution type usually varies across the units of measure and is likely to be piecewise distributions in each case. As such, a multivariate distribution of a known type is quite unlikely.

While modeling severity dependence is quite challenging, it is the possibility of large losses, not necessarily very many in multiple units of measure, that matters most for the bank operational risk capital estimate. Practitioners often find that the impact of frequency dependence on bank operational risk capital estimate is marginal at best when severities are simulated in an independent fashion.[21] As such, some banks find a middle ground by modeling dependence at the unit of measure annual loss level. Clearly this is potentially more consequential for capital estimates than merely considering frequency dependence, and simulating dependent losses is somewhat less complex compared to simulating dependent severities.[22]

5.2 How to add a dependence structure

Even the simplistic dependence structures do not generally yield known types of TOH loss distributions. To introduce a dependence structure to the units of measure losses through parametric specification of a known type of joint distribution, is thus not feasible. A popular solution is the use of copulas, meaning a link or a join, to add a dependence structure to a set of marginal loss distributions that are not necessarily of the same form. The main idea of a copula is to describe the dependence structure of two or more random variables using the dependence structure of their monotonically increasing mappings.[23] This is possible due to the invariance property of strictly increasing mapping. Say $h_1(S_1)$ and $h_2(S_2)$ are monotonically increasing functions of the losses for the units of measure 1 and 2, and suppose the function C describes the dependence structure of S_1 and S_2. Then, if the function C_h describes

[20] See Yahav and Shmueli (2008) for a computationally efficient method for simulating multivariate Poisson data that allows flexible correlation structure and unequal Poisson rates.

[21] When severities are simulated in a dependent manner, frequency dependence may become consequential.

[22] Frachot *et al* (2004), however, suggest that the loss correlation is usually less than 10% even with high-frequency correlation, when severities are assumed to be uncorrelated. Using the ORX data, Cope and Antonini (2008) report quarterly loss correlation to be generally less than 20% across event types, business lines and cells of event type-business line.

[23] An excellent technical discussion of copulas can be found in (Malevergne and Sornette 2006, Chapter 3). Bouye *et al* (2000) describe a host of finance applications, Hurd *et al* (2007) show the use of copulas in modeling bivariate foreign exchange distributions and Dorey (2006) offers a non-technical discussion on copulas. Valle *et al* (2008) illustrate the use of copulas in operational risk under the LDA method.

the dependence structure of $h_1(S_1)$ and $h_2(S_2)$, then $C = C_h$. Functions like C and C_h that describe the dependence structure of random variables are called copula functions.

The mileage derived from copulas is that known forms of dependence structure can be added to widely varying types of marginal distribution including those of no known form. In the latter case, a large number of simulations are first performed (often using known frequency and severity distributions); then the collection of the simulated draws are used as an empirical distribution to invert the CDFs simulated from the known copula.

While the Gaussian copula is popular in many applications, it tends to underestimate risk capital since the dependence tapers off in the tails. A popular copula that addresses this potential limitation is the t-copula that uses the dependence structure of a multivariate t-distribution. In a t-copula, the common degrees of freedom parameter controls the extent of tail dependence.[24] With a low value of the degrees of freedom parameter, the tail dependence under the t-copula is much more severe than that under the Gaussian copula. The research issue here is the determination of the degrees of freedom parameter and its calibration to empirical estimates of such dependence.[25]

5.3 Statistical measure of dependence

It is commonplace in the operational risk literature and elsewhere to describe dependence using correlation. Since operational loss dependence may not be linear in nature, the popular Pearson's linear correlation coefficient may not properly capture the dependence of losses across the units of measure. A bank needs to consider additionally rank-based correlation measures such as Spearman's Rho and Kendall's Tau. Once such non-parametric measures are estimated, they can then be converted to Pearson linear coefficients for simulation purposes.

To say the least, empirical estimation of loss correlation across the units of measure is perhaps one of the most challenging implementation issues in operational risk capital estimation. Both internal loss data and consortium data for other banks typically cover only a few years, thus making estimates of annual loss correlations to be quite unreliable. Publicly available external data is perhaps relatively better in this regard, also allowing correlation estimates (of severities) above various thresholds for the tail. However, how to adjust for the reporting bias of public data in estimating correlations is not quite known.

[24] See Demarta and McNeil (2004) for a discussion of the t-copula and the related copulas.

[25] Based on the ORX data, Cope and Antonini (2008) find little evidence of tail dependence among various units of measure. This suggests the use of a Gaussian copula or a t-copula with higher degrees of freedom. Interestingly, with heavy tails and due to the fact the quantiles are not coherent measures, a lower level of dependence among the units of measure does not ensure a lower operational risk capital charge compared to the case of independence.

When it comes to simulating correlated losses, a vexing issue with the popular copula approach is that the correlation in the simulated losses for the units of measure rarely matches that which is used in the copula to generate the multivariate CDFs. Thus, it is worth researching how to calibrate the correlation in the copula in order to achieve a targeted correlation in the simulate losses for the units of measure.

6 SUMMARY AND CONCLUSIONS

This paper provides an overview of the various practical issues in modeling operational risk capital. To facilitate orderly exposition, we divided the issues in operational risk capital modeling discussed in this paper into three broad areas, pertaining to datasets, frequency and severity distributions and dependence structure. Given the firm-specific nature of operational risk and the lack of high-severity experience for most units of measure, banks would normally be inclined towards relying on internal loss data for the estimation of frequency distribution and the lower tail or body of a piecewise severity distribution. Future research is needed with regards to reliable estimation of these distributions in the presence of a shallow loss data history, reporting delays and measurement errors in the data, protracted loss events possibly spanning multiple units of measure and changing business dynamics including acquisitions and corporate restructurings.

The severity distribution beyond the body, generically called the “tail”, is perhaps the most consequential and challenging component of operational risk measurement for a unit of measure. Numerous challenges remain in the actual estimation of these distributions using external and/or scenario data and then selecting a suitable distribution. Practical implementation requires answers to questions such as where should the tail start; should the tail be further segmented to handle the heavy tail behavior at high quantiles; should the thresholds be determined endogenously; if scenario data is used directly for estimation rather than for model validation then how should it be integrated with external loss data and how should the behavioral biases of scenario data and the special nature of this data be addressed; how could a bank detect and retain possibly influential external loss and scenario data points while screening out the irrelevant ones; how to choose between or possibly undertake a selective merger of publicly available external loss data and privately subscribed consortium data; and what are the suitable estimation methods given the data choices.

Except in some exceptional circumstances, the probability distribution of the TOH annual loss will not be of a known form. The copula approach to modeling the dependence structure holds promise in this regard. However, much more research is called for in the choice of copula in the context of bank operational risk capital estimation and the calibration of the copula to the empirical estimates of operational loss correlations.

Based on published research and the author’s own experience, there does not appear to be dominant modeling choices or solutions for many of these issues. Further, given the complex nature of the issues, researchers and practitioners often address them in a narrowly focused and in an ad hoc manner. While this has advanced understanding of the stand-alone specific issues, the impact of the separately made modeling choices on the properties of the TOH capital estimate remains opaque. Additionally, the importance of the various modeling issues and properties of the capital estimate may vary across banks and regulators. It would thus be useful to develop an approach that addresses the detailed modeling choices with sufficient rigor and yet connects their choice to the properties of the capital estimates with as much transparency as possible, in the meantime allowing bank or regulator specific prioritization of the modeling choices.

One possible line of thinking on this comprehensive approach is a multi-stage sequential optimization. In the first stage, the bank may undertake a series of local optimizations with respect to the detailed modeling choices concerning a unit of measure and determine the ranks or scores of competing choices based on bank and regulator specific priorities. In the second stage of optimization, each set of competing local modeling choices for a unit of measure will be evaluated holding the modeling choices at their local optima in other areas of modeling for the same unit of measure. At the end of the second stage for each unit of measure, a bank would have generated two scores for all the competing modeling choices for the various modeling areas of the unit of measure, the local scores and the unit of measure scores. In the third stage, these two sets of scores can then be combined into a composite score, for example by simply adding the scores, to determine the unit of measure optima in the various modeling areas of the unit of measure. In the fourth stage, using the unit of measure optima for all the units of measure, a bank would score competing modeling choices in the specific modeling areas of dependence modeling in ways similar to the first stage. These scores might be called the dependence-local scores. This will be followed by the fifth stage scores, named say dependence scores, of the dependence modeling choices in terms of their effect on the attributes of the TOH capital estimate while holding the modeling choices at their local optima in other areas of dependence modeling. Similar to the third stage, the sixth stage and final stage would involve combining the dependence-local scores and the dependence scores to arrive at the optima for the dependence modeling areas.

In the above multi-stage sequential optimization, the bank or regulator specific priorities could be reflected in the weights of the attributes in the objective functions at various stages. The optimization scheme suggested here is merely a sketch to systematically gather together what often feels like a myriad of disparate and seemingly irreconcilable issues in operational risk capital estimation by practicing banks. It is nonetheless hoped that future research will ensue along these lines to improve rationalization and transparency of the modeling choices in operational risk capital estimation.

This paper is not exhaustive by any means in terms of the modeling issues and there are many other important issues that were not discussed here in detail. For example, practitioners often struggle with convergence and stability problems in operational risk capital simulations. Given that operational risk capital concerns a quantile typically well beyond known data points, projection into this unchartered territory can be quite problematic. It is not unusual in simulation to see alternative seed values and the number of runs affecting the operational risk capital estimate. How to reduce simulation error variance and establishing the convergence properties of estimators is thus a potentially interesting area of research as well.[26] Among other issues that require future research we can include the modeling of insurance and offsets, and operational risk benchmarking, back testing and stress tests.

REFERENCES

Alexander, C. (2000). Bayesian methods for measuring operational risk, derivatives use. *Trading and Regulation* **6**(2), 166–186.

Aue, F., and Kalkbrener, M. (2006). LDA at work: deutsche bank's approach to quantifying operational risk, *The Journal of Operational Risk* **1**(4), 49–93.

Basel Committee on Banking Supervision (2006). International convergence of capital measurement and capital standards: a revised framework. Bank of International Settlements.

Baud, N., Frachot, A., and Roncalli, T. (2002). Internal data, external data and consortium data for operational risk measurement: how to pool data properly? Working Paper, Groupe de Recherche Opérationelle, Crédit Lyonnais.

Bouye, E., Durrieman, V., Nikeghbali, A., Riboulet, G., and Roncalli, T. (2000). Copulas for finance: a reading guide and some applications. Working Paper, Groupe de Recherche Opérationelle, Crédit Lyonnais.

Buch-Kromann, T., Englund, M., Gustafsson, J., Nielse, J. P., and Thuring, F. (2006). Nonparametric estimation of operational risk losses adjusted for underreporting. Working Paper, University of Copenhagen.

Bühlmann, H., Shevchenko, P. V., and Wüthrich, M. V. (2007). A "toy" model for operational risk quantification using credibility theory. *The Journal of Operational Risk* **2**(1), 3–19.

Chernobai, A., Menn, C., Truck, S., and Rachev, S. T. (2004). A note on the estimation of the frequency and severity distribution of operational losses. *Applied Probability Trust* (December), 1–16.

Chernobai, A., and Rachev, S. T. (2006). Applying Robust methods to operational risk modeling. Working Paper, University of California at Santa Barbara.

Cohen-Cole, E., and Prono, T. (2007). Loss distribution, external data and model averaging. Working Paper, Federal Reserve Bank of Boston.

Cope, E., and Antonini, G. (2008). Observed correlations and dependency among operational risk losses in the ORX consortium database. *The Journal of Operational Risk* **3**(4), 47–76.

Cope, E., and Labbi, A. (2008). Operational risk scaling by exposure indicators: evidence from the ORX database. *The Journal of Operational Risk* **3**(4), 25–46.

Cummins, J. D., Christopher, L. M., and Wei, R. (2006). Market value impact of operational loss events for US banks and insurers. *Journal of Banking and Finance* **30**(10), 2605–2634.

Cummins, J. D., and Embrecht, P. (2006). Introduction: special section on operational risk. *Journal of Banking and Finance* **30**(10), 2599–2604.

[26] See Pritsker (1997) for accuracy and computational time of VaR methodologies.

Dahen, H., and Dionne, G. (2010). Scaling models for the severity and frequency of external operational loss data. *Journal of Banking and Finance* **34**(7), 1484–1496.

de Fontnouvelle, P., DeJesus-Rueff, V., Jordan, J., and Rosengren, E. (2003). Using loss data to quantify operational risk. Working Paper, Federal Reserve Bank of Boston.

de Fontnouvelle, P., Rosengren, E., and Jordan, J. (2004). Implications of alternative operational modeling techniques. Working Paper, Federal Reserve Bank of Boston.

Degen, M., and Embrechts, P. (2008). EVT-based estimation of risk capital and convergence of high quantiles. *Advanced Applied Probability* **40**, 696–715.

Degen, M., Embrechts, P., and Lambrigger, D. D. (2007). The quantitative modeling of operational risk – between g-and-h and EVT. *AST Bulletin* **37**(2), 265–291.

Demarta, S., and McNeil, A. J. (2004). The t copula and related copulas. Working Paper, Federal Institute of Technology, Zurich.

Dorey, M. (2006). Why aren't copulas far more popular? *Professional Investor* (February), 18–21.

Dutta, K., and Perry, J. (2007). A tail of tails: an empirical analysis of loss distribution models for estimating operational risk capital. Working Paper, Federal Reserve Bank of Boston.

Embrechts, P., Furrer, H., and Kaufmann, R. (2003). Quantifying regulatory capital for operational risk. *Derivatives Use, Trading and Regulation* **9**(3), 217–33.

Embrechts, P., Kaufmann, R., and Samorodnitsky, G. (2002). Ruin theory revisited: stochastic models for operational risk. Working Paper, ETH-Zurich and Cornell University.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, New York.

Ergashev, B. (2007). Should risk managers rely on maximum likelihood estimation method while quantifying operational risk? Working Paper, Federal Reserve Bank of Richmond.

Ergashev, B. (2009). Estimating the lognormal-gamma model for operational risk using the MCMC method. Working Paper, Federal Reserve Bank of Richmond.

Frachot, A., Georges, P., and Roncali, T. (2001). Loss distribution approach for operational risk. Working Paper, Group de Recherche Opérationnelle, Crédit Lyonnais, France.

Frachot, A., Mouduoulaud, O., and Roncali, T. (2003). Loss distribution approach in practice. Working Paper, Group de Recherche Opérationnelle, Crédit Lyonnais, France.

Frachot, A., and Roncalli, T. (2002). Mixing internal and external data for managing operational risk. Working Paper, Groupe de Recherche Opérationnelle, Crédit Lyonnais.

Frachot, A., Roncalli, T., and Salomon, E. (2004). The correlation problem in operational risk. Working Paper, Groupe de Recherche Opérationnelle, Crédit Lyonnais.

Guillen, M., Gustafsson, J., Nielsen, J. P., and Pritchard, P. (2006). Using external data in operational risk. *The Geneva Papers* **32**, 178–189.

Gustafsson, J., and Thuringb, F. (2007). Best distribution for operational risk: a computational evaluation of different loss distribution assumptions. Working Paper, University of Copenhagen.

Hurd, M., Salmon, M., and Schleicher, C. (2007). Using copulas to construct bivariate foreign exchange distributions with an application to the sterling exchange rate index. Working Paper 334, Bank of England.

Kahneman, D., and Frederick, S. (2006). Frames and brains: elicitation and control of response tendencies. *Trends in Cognitive Sciences* **11**(2).

Kahneman, D., and Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica* **47**(2), 263–291.

Klugman, S. A., Panjer H. H., and Willmot, G. E. (2004). *Loss Models: From Data to Decisions*. Wiley, Hoboken, New Jersey.

Malevergne, Y., and Sornette, D. (2006). *Extreme Financial Risks: From Dependence to Risk Management*. Springer, The Netherlands.

Mignola, G., and Ugoccioni, R. (2006). Effect of a data collection threshold in the loss distribution approach. *The Journal of Operational Risk* **4**(1), 35–47.

Moscadeli, M. (2003). The modeling of operational risk: the experience from the analysis of the data collected by the risk management group of the basel committee. Working Paper, Bank of Italy.

Na, H. S., Van Den Berg, J., Miranda, L. C., and Leipoldt, M. (2006). An econometric model to scale operational losses. *The Journal of Operational Risk* **1**(2), 11–31.

Netter, J. M., and Poulsen, A. B. (2003). Operational risk in financial service providers and the proposed basel capital accord: an overview. *Advances in Financial Economics* **8**, 147–171.

Perry, J., and de Fontnouvelle, P. (2005). Measuring reputational risk: the market reaction to operational loss announcements. Working Paper, Federal Reserve Bank of Boston, Boston, MA.

Peters, G. W., and Sisson, S. A. (2006). Bayesian inference, Monte Carlo sampling and operational risk. *The Journal of Operational Risk* **1**(3), 27–50.

Pritsker, M. (1997). Evaluating value-at-risk methodologies: accuracy versus computational time. *Journal of Financial Services Research* **12**, 201–242.

Shevchenko, P. V., and Wüthrich, M. V. (2006). The structural modelling of operational risk via Bayesian inference: combining loss data with expert opinions. *The Journal of Operational Risk* **1**(3), 3–26.

Shih, J, Samad-Khan, A., and Medapa, P. (2000). Is the size of an operational loss related to firm size? *Operational Risk Magazine* **2**(1).

Tripp, M. H., Bradley, H. L., Devitt, R., Orros, G. C., Overton, G. L., Pryor, L. M., and Shaw, R. A. (2004). Quantifying operational risk in general insurance companies. *British Actuarial Journal* **10**(5), 919–1012.

Valle, L. D., Fantazzini, D., and Giudici, P. (2008). Copulae and operational risks. *International Journal of Risk Assessment and Management* **9**(3), 238–257.

Wei, R. (2006). Quantification of operational losses using firm-specific information and external database. *The Journal of Operational Risk* **4**(1), 3–34.

Yahav, I., and Shmueli, G. (2008). An Elegant method for generating multivariate poisson random variables. Working Paper, University of Maryland.

Yu, W. (2005). New capital accord Basel II: using LDA approach for measuring operational risk? Working Paper, Vrije Universiteit, Amsterdam.