Applying fuzzy scenarios for the measurement of operational risk

Isis Bonet\textsuperscript{a,*}, Alejandro Peña\textsuperscript{b}, Christian Lochmuller\textsuperscript{a}, Héctor Alejandro Patiño\textsuperscript{a}, Francisco Chiclana\textsuperscript{c,d}, Mario Góngora\textsuperscript{c}

\textsuperscript{a}Computational Modelling and Simulation Research Group, EIA University, Envigado, Colombia
\textsuperscript{b}Escuela de Administración, Universidad EAFIT, Medellín, Colombia
\textsuperscript{c}Institute of Artificial Intelligence, School of Computer Science and Informatics, De Montfort University, Leicester, UK
\textsuperscript{d}Andalusian Research Institute on Data Science and Computational Intelligence, University of Granada, Spain.

Abstract

Operational risk measurement assesses the probability to suffer financial losses in an organisation. The assessment of this risk is based primarily on the organisation’s internal data. However, other factors, such as external data and scenarios are also key elements in the assessment process. Scenarios enrich the data of operational risk events by simulating situations that still have not occurred and therefore are not part of the internal databases of an organisation but which might occur in the future or have already happened to other companies. Internal data scenarios often represent extreme risk events that increase the operational Value at Risk (OpVaR) and also the average loss. In general, OpVaR and the loss distribution are an important part of risk measurement and management. In this paper, a fuzzy method is proposed to add risk scenarios as a valuable data source to the data for operational risk measurement. We compare adding fuzzy scenarios with the possibility of adding non fuzzy or crisp scenarios. The results show that by adding fuzzy scenarios the tail of the aggregated loss distribution increases but that the effect on the expected average loss and on the OpVaR is lesser in its extent.

Keywords: Scenario analysis, Value at Risk, Integration of different data sources, Fuzzy Scenarios, Operational Risk, Loss Distribution Approach

1. Introduction

The Basel Committee on Banking Supervision (BCBS), which develops standards for international banking regulations, defined operational risk (OpRisk) in the second Basel accord, known as Basel II, 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” \cite{1}. All organisations are exposed to operational risk, and its management was first regulated for the financial sector in Basel II \cite{1}. Throughout history, there have been several cases with significant losses due to operational risk, with some of them having gained global news coverage and attention. The most emblematic cases in recent times include: Barings Bank (1995) with $1$ billion of losses; Long-Term Capital Management (1998) with $4$ billion of losses;

*Corresponding author

Email addresses: isis.bonet@eia.edu.co (Isis Bonet), japena@eafit.edu.co (Alejandro Peña), christian.lochmuller@eia.edu.co (Christian Lochmuller), hector.patino@eia.edu.co (Héctor Alejandro Patiño), chiclana@dmu.ac.uk (Francisco Chiclana), mgongora@dmu.ac.uk (Mario Góngora)

Cases like the mentioned above have increased interest on this area of research. Available research approaches include different aspects of measuring operational risk to clarify the magnitude of this risk. However, unlike in market risk, for example, one characteristic of operational risk is that there is a lack of data in corporate databases with regard to operational risk events. Although operational risk events can generate significant losses (high severity), these events usually do not occur very frequently (low probability of occurrence). BCBS allows a specific method for measuring operational risk to be selected by each bank according to its risk profile, which will depend on a number of factors including: the size of the institution; the nature and complexity of its activities; and its maturity level in risk management. Accordingly, BBBS defined, within the Advanced Measurement Approach method (AMA), different sources of data for estimating the expected and non-expected losses and the value at risk (VaR) that refer to operational risk. These data sources include: internal data of a financial entity; external data; environmental and control factors; and scenarios. The latter is generally used for modelling forward looking loss situations, taking into account possible exposures to operational risk.

The internal data of a financial institution is the most important element in risk management since this data reflects its business. However, because companies often do not have enough internal data, additional external data sources play an important role when measuring operational risk. One type of external data sources are operational risk databases of other financial institutions or operational risk databases that represent an entire industry. Nevertheless, these databases availability is limited because most companies are not willing to share this data as it is considered a valuable asset and a “business secret”. This is why it is useful to generate scenarios based on possible risk events that can occur quickly which may lead to significant losses for the company.

The incorporation of scenarios is challenging when determining the magnitude of operational risk. However, applying operational risk scenarios can make an important different contribution since possible losses, supported by domain experts’ judgement and experience, can be included. These possible losses can also include significant losses that have not yet occurred in the company but are likely to occur in the future or have occurred in other organisations [4]. The application of scenarios in the context of the AMA framework consists of different steps: the generation of scenarios; scenario assessment; data quality assurance; the selection and combination of the values for potential losses in risk matrices; and the estimation of the regulatory capital based on a certain quantile of the aggregated loss distribution [5]. The first three elements lie within the main responsibility of experts selected by the institution. Here, we will focus on the last two steps, which can be summarised in two main phases: representing and incorporating the scenarios to the values of the aggregated loss distribution of the company.

Hence, this paper focuses on scenario analysis and in particular, on how to define operational risk scenarios in a financial institution for measuring operational risk. Also, taking into account that the definition of scenarios is subjective, a novel method is proposed for designing ‘fuzzy scenarios’ with regard to the two dimensions of operational risk: frequency and severity. The proposed approach is to be implemented within the context of scenario based AMA (sbAMA) for the measurement and management of operational risk. This implies the incorporation of the human factor and the fuzziness of the information that can be used to design the scenarios in financial institutions. The process starts with the generation of data by using the loss distribution approach (LDA) and by modelling scenarios. For the latter ‘worst-in-M-years’ situations will be generated. The result refers to the tail of the loss distribution, where we compare different ways of mixing scenarios when calculating the VaR for operational risk (OpVaR).
2. Literature review

As per the BCBS definition, OpRisk materialises when processes and business units do not perform well at the different qualitative elements involved. These are often elements that are related to the decision making of employees, suppliers or customers of a financial institution. These elements are not static, they evolve over time and, therefore, it is not easy to incorporate them into the traditional models that are applied to measure OpVaR.

The opVaR represents the minimum capital reserve required to cover potential operational losses within a certain period of time, typically one year, and a certain confidence level. This value can be estimated based on the sum of expected losses (EL) and unexpected losses (UL) of a loss distribution, which can be constructed by recording all losses of an organisation, caused by operational risk events and that have been generated within a certain period of time. These risk events occur in a particular business line and are associated with a particular risk event type [5–7].

In general, internal measurement models, which consider the specific risk profile of a business, are often applied by financial institutions for calculating opVaR. OpRisk originates in a business line and it is associated to one or several risk event types. Basel II defines the following eight business lines for financial institutions (banks): 1) corporate finance; 2) trading and sales; 3) retail banking; 4) commercial banking; 5) payment and settlement; 6) agency services; 7) asset management; and 8) retail brokerage. The same standard also defines seven risk event types: 1) internal fraud; 2) external fraud; 3) employment practices and workplace safety; 4) clients, products and business practices; 5) damage to physical assets; 6) business disruption and system failures; and 7) execution, delivery and process management. To represents the conceptual schema for storing, measuring, controlling and managing operational risk in financial institutions, a matrix of dimension $8 \times 7$ is constructed based on the business lines and risk event types listed above.

Basel II defines several methods for calculating opVaR. Each of these methods has certain benefits and costs associated to it [8]. The Advanced Measurement Approach (AMA) is the most flexible and sophisticated approach since it allows a financial institution to develop an internal risk quantification framework, which in turn allows adapting the techniques to determine the specific risk for the situation at each institution.

The first version of Basel II proposed three methodologies for modelling AMA: the Internal Measurement Approach, the Score Card Approach, and the Loss Distribution Approach (LDA) [9]. Nevertheless, the latest version of this accord is more flexible and does not specify a particular mathematical model that banks have to use in an AMA context. Indeed, it is stated that financial institutions “will be free to use models based, for example, on variance-covariance matrices, historical simulations, or Monte Carlo simulations” [1]. This offers financial institutions the possibility to select or create their own methods for calculating the amount of capital that is required by the supervising body. For the aim of this paper, the LDA is used since it is the most common approach within the AMA.

Basel II also specifies that any operational risk measurement system “must include the use of internal data, relevant external data, scenario analysis and factors reflecting the business environment and internal control systems” [1]. Thus, in computing OpRisk, challenges may arise when combining the different kinds of data (internal data, relevant external data, scenario analysis) into one single risk measure (opVaR) [5]. Related to the quantification of OpRisk, several difficulties have been identified. The first one is about the simulation of internal data. This area is one of the most researched areas, with some methods, such as Monte Carlo simulation, being proved to produce good results when representing the distribution of the data and when computing the
operational value at risk at a specific quantile of the distribution \[ 4, 5 \]. The second difficulty is related to the question of how to combine or mix internal and external data to determine the opVaR \[ 10 \].

In addition, scenario analysis provides a valuable data source for opRisk measurement. Scenario analysis has its own challenges: how to represent and model scenarios, and how to integrate them with internal data in order to estimate the \( 99.9^{th} \) percentile confidence interval that is used for operational risk modelling and measurement.

### 2.1. Including scenarios when calculating the VaR

Based on a count of publications indexed in research databases such as EBSCO, Science Direct and Scopus, it is obvious that scenario analysis based AMA (sbAMA) has been less researched than internal and external data based AMA. However, scenario analysis is not new in the context of operational risk management. In general, scenarios are used for modelling risk events that are severe, in terms of their impact (loss), but which are characterised by a very low probability of occurrence. Scenarios help evaluating the exposure to severe losses (tail events) and, therefore, they can be a key element in determining the VaR \[ 11 \]. Thus, the use of expert judgement regarding high severity events, uncommon in the historical data of an institution but which may threaten the continuity of the business, can be a powerful tool for predicting future losses \[ 9 \]. The ISO 31010 risk management standard defines the analysis of scenarios as a useful method or tool in the process of risk identification and risk analysis and evaluation. Annex B10 of ISO 31010:2009 states that scenario analysis refers to the development of descriptive models that define how the future might turn out. Hence, scenario analysis can be used to identify risks by considering possible future developments and exploring their implications. In general, as a form of sensitivity analysis, scenarios can be built in terms of ‘best case’, ‘worst case’ and ‘most likely’ scenarios in order to analyse the potential consequences and their probabilities for each possible future \[ 12 \]. Although scenario analysis is often focused on understanding extreme but plausible events, each business will have their unique specific method, whose success will crucially depend on the adequate modelling and implementation of the opinion and judgement of experts who understand the forces and drivers of change and also the business objectives \[ 13 \].

The sbAMA method shares with LDA the idea of combining the two dimensions of risk events, frequency and severity, to calculate the distribution of aggregate losses used to determine the VaR \[ 14 \].

Since scenarios are hypothetical events of an institution or, in general terms, the inherent risks of the industry, they can be useful to make projections towards the future and for adjustments with regard to the frequency and severity of corresponding historical events. Indeed, scenarios are used to take into account the latest news or future changes, not reflected in historical losses, in the risk profile of the entity and the business environment. Regarding opRisk measurement it has been recommended to focus only on the formulation and evaluation of the ‘worst scenarios’ in order to avoid a bias by reducing the estimated risk and the required capital through a disproportionate number of scenarios that are generated with respect to the body of the severity distribution \[ 15 \]. It can be concluded that incorporating analysis of scenarios in opRisk measurement provides a benefit in the form of more accurate assessment of capital. Even companies that are not required to calculate VaR use scenarios to develop a better understanding of their exposure to high risks.

The definition of a scenario, as a very first step in scenario analysis, is usually carried out by one or a group of experts, in this last case through consensus. Some authors have worked on different ways to combine the opinion of several experts for the same scenario to arrive at a consensus. For example, the use of different weights for different experts, based on their expertise with regard
to a specific scenario, and their assigned probability to each scenario is one of the methodologies proposed in literature [16]. The present work assumes that each scenario is already defined by consensus of a group of experts [17–19], so that the input to our model is just the consensus for each scenario.

Based on the above definition of scenario, it is clear that one disadvantage relates to their subjectivity construction based on experts’ opinions and the representation of future events often without historical record of an organisation. Another difficulty related to the few data points generated by scenarios, which makes difficult to model them as a distribution. Therefore, the question of how to represent and integrate scenarios with internal data is important to be addressed.

How to represent scenarios is a challenge, because in general, there are only few scenarios and often their definition is vague. Furthermore, they need to be modelled in the same ‘universe’ of internal data, so that scenarios can be integrated or mixed with internal data. Usually a scenario is described by two risk event variables: frequency of the event, i.e. number of years required for a risk event to happen once (on average); severity of the risk event if it occurs. For example, a scenario for a financial institution may represent a risk event which occurs once in 10 years (frequency) with an associated loss of $10 million (severity), while another scenario for a financial institution may represent a risk event that occurs ten times a year with a loss of $1 million each time it occurs. Severity is usually represented in two ways: as a range defined solely by a lower and an upper bound values; as a range defined a lower, upper and mean values. Whatever the representation of the severity is, in a certain scenario, it is important to be consistent and to apply the same mechanism for modelling the frequency with which the risk events occur. The frequency and severity of the scenarios can be standardised to a number of years, $N$, which depends on how these scenarios are to be added: if the scenarios will be added directly to the internal data, then $N$ is the number of years that cover the internal data; however, if they are added to the model after processing the internal data, $N$ is the number of years that will be generated to estimate the VaR.

In addition, several methods to incorporate scenarios with internal data in measuring OpRisk are described in the literature there are. One of such methods aims to fit a distribution to the internal data augmented with the scenarios, i.e. the internal data with the addition of a few additional risk events as per the scenarios to implement [20], that must refer to the same time horizon. This way of adding the scenarios can cause a change in the shape of the fitted distribution had this been obtained using only the internal data. Variations of this method have also been developed. In addition, there are methods based on Bayes theory [21–23].

The construction of scenarios can be done based on interviews, surveys, or workshops/focus groups of domain experts, who can be recruited inside and outside the organisation. Different experts may construct different risk scenarios with the same context and same information whether they have similar or different levels of knowledge and experience [24–28]. For example, one expert may be overconfident or enthusiastic with regard to the effectiveness and financial return of the implementation of fraud prevention measures, compared to other experts. Therefore, the construction of risk scenarios is pervaded with uncertainty, which in most cases is expressed or described in an uncertain qualitative way based on using natural language or linguistic values rather than in a quantitative way based on precise numbers. The use of linguistic assessment for both the frequency and severity variables of scenarios is addressed in this paper by using a fuzzy set based approach for modelling and representing such variables as fuzzy linguistic variables [29–31].
2.2. Scenario analysis using fuzzy logic

When measuring OpRisk, the estimated risk events are a) uncertain and b) there is neither precision with respect to the estimation of their probability of occurrence (frequency) nor with respect to their severity in the case of their occurrence. That is to say, there is not always defined boundaries and precise values, since the construction of scenarios with respect to the frequency and the impact of an operational risk is highly qualitative and also biased, depending on the cognitive processes, applied by the actors that are involved in the generation of the scenarios.

Several authors have used fuzzy sets in different ways and with different purposes in risk management. However, its application is more frequent in the general assessment of operational risk based on internal data, with the majority based on the fuzzy representation of the frequency and severity variables as fuzzy linguistic variables. Examples of these include the definition of fuzzy rules to create a Fuzzy Logic Inference System [32], and the use of a fuzzy risk matrix to model the assessment of the level of risk of specific frequency and severity linguistic variables [33]. A fuzzy model based on questionnaires for capturing experts’ opinions has also been proposed [34], where an analytical hierarchy process (AHP) is used to determine if, and how much, one of the factors represented in the questionnaire is more important than the others. The authors evaluate rules to obtain a fuzzy output number from which a precise number, that they call ‘blurred frequency modifier’, is obtained by defuzzification. The authors conclude that this fuzzy concept allows the integration of the human factor in the generic frequency of failures, obtaining a more realistic frequency. Fuzzy cognitive maps (FCMs) have also been used for the construction of scenarios for photo-voltaic solar cells [35]. Although, there is evidence that fuzzy logic has been used in operational risk management, there are only few studies where fuzzy logic is used to integrate scenarios with internal or external data. This is also true, for the use of fuzzy logic when representing expert opinion and combining these opinions to reach a consensus on scenarios. The research of Durfee and Tselykh [16] applies fuzzy numbers to represent scenarios and to sum up the different opinions of experts for each of the scenarios, weighting the opinion of the expert with the $\alpha -$cut. The aggregation of expert opinion is based on a fuzzy weighted average as used in Luukka et al. [36]. In both works, the authors use triangular fuzzy loss estimates to represent the scenarios, based on the judgement and valuation of the domain experts. Each expert evaluates severity and frequency for each scenario by using three valuations: optimistic, realistic and pessimistic. In this paper, fuzzy logic is not used for the combination of expert opinions but for describing the assumed agreed scenarios, so that these are easily integrated with internal data and to affect mainly the tail of the operational loss distribution.

3. Data and Methodology

3.1. Data

For the experiments we use “lossdat”, a hypothetical example database publicly available through the OpVaR package of the R programming language [37]. This database contains four data sets, where each element represents a risk event with its corresponding loss. The data is described in a three-column matrix: the loss, the date when it occurred and a period, which in this case is in terms of quarters. These sets comprise data of 10 years, from 2007 to 2017. Since the Basel proposal and the most common definition of the opVaR refer to a one year period, the period for the data sets was changed to one year. For the data sets, which can be considered as data registered for different business lines of a financial institution, an independent VaR calculation is calculates, which means that four data sets are analysed.
Table 1 shows a summary of the data, describing each set: the number of events per year on average (Events per year), the minimum (Min Loss) and maximum loss (Max Loss), and also the average loss (Avg Loss) for each data set.

Table 1: Description of data sets.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Events per year</th>
<th>Min Loss</th>
<th>Max Loss</th>
<th>Avg Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>196.5</td>
<td>5</td>
<td>6382</td>
<td>1016.951</td>
</tr>
<tr>
<td>2</td>
<td>202.5</td>
<td>3</td>
<td>6213</td>
<td>1139.488</td>
</tr>
<tr>
<td>3</td>
<td>199.5</td>
<td>48</td>
<td>12092</td>
<td>1051.507</td>
</tr>
<tr>
<td>4</td>
<td>194.1</td>
<td>201</td>
<td>6215</td>
<td>969.1051</td>
</tr>
</tbody>
</table>

3.2. Simulation of aggregated losses using the Monte Carlo method

There are various computational methods to calculate the distribution of the aggregated losses (LDA). The most widely used are: Monte Carlo, Panjer recursion and Fourier inversion techniques. In the literature, the Monte Carlo method prevails because it is easy to implement [5].

The LDA approach focuses on the question, how often a risk event occurs over a period of time (frequency) and if it occurs, what is its impact in terms of the amount of money that the organisation loses (severity). Thus, it is based on the following two variables:

- **Frequency** – $F$: Probability distribution that describes the number of times a risk event is repeated in a defined time interval. Annual frequency of loss ($N$) is computed as the aggregation of the severity values over a one-year horizon.

- **Severity** – $S$: Severity refers to the amount of a loss, due to a risk event that occurs during the period of evaluation. It is represented as a probability distribution, $S = \{s_1, s_2, \ldots, s_N\}$, that describes the amount of the losses of OpRisk for a business line or risk event type.

The total loss ($Z$) is obtained as the sum of losses in the different business lines or risk event types:

$$Z = \sum_{i=1}^{NL} SL_i,$$

where, $SL_i$ (see (2) below) is the aggregated severity of business line $i$, and $NL$ is the number of business lines (eight as per Basel II – see Section 2, although some banks operate with less business lines based on their business model). If the loss events are reported at the level of risk types, then $SL$ can be expressed as the sum of the aggregated severity for all risk event types in each business line.

$$SL = \sum_{i=1}^{N} s_i$$

Value at Risk (VaR) is used by financial institutions to calculate their capital charge for a one year holding period (next year), and is defined as the 99.9th percentile, one-tailed confidence interval, as per the below expression:

$$VaR_q = \inf \{z \in \mathbb{R}: \text{Prob}[Z > z] \leq 1 - q\}.$$
where \( q \) is the quantile of the loss distribution, usually the value 99.9% mentioned before, although sometimes a range of quantiles is also used. As mentioned before, this paper is based on the LDA model and uses the Monte Carlo method for the simulation.

To estimate VaR it is necessary to construct the experimental distributions of each variable \((F, S)\) of the input matrix. It can be noted that the estimation of the VaR is carried out according to the selection of the user. Thus, the VaR can be determined for an entire company; a business line; a risk event type; or for the intersection of a business line and a risk event type such as, for example, internal fraud (risk event type) in retail banking (business line). Notice that the frequencies that are registered in the database cannot be fit to a distribution, since these are quarters frequencies and the model requires annual frequencies. Thus, the distribution of the annual frequency are randomly generated using a Poisson distribution.

The number of annual frequencies corresponds to the number of years. To generate the \( N \) numbers of annual frequencies, the average of the annual frequencies contained in the database is calculated. With the mean value, the \( N \) numbers are generated from a Poisson distribution \([38]\). For the generation of the severity values we use an empirical distribution based on the original data.

The Monte Carlo method consists in iterating this process \( k \) times. In each iteration we obtain a \( Z \) value (annual aggregate loss), and as a result the aggregate loss distribution \( \{Z_1, \ldots, Z_k\} \). The estimated VaR from the data that compose the LDA is obtained using (3) with \( q = 99.9th \) percentile of the aggregate loss distribution.

### 3.3. Representing scenarios using fuzzy sets

Fuzzy logic allows for modelling uncertainty and imprecision and can address problems with no sharp boundaries and/or imprecise values \([39]\). Fuzzy Numbers (FN) allow to represent imprecise values in a quantitative and more realistic way than just with precise real numbers. FN\s can be conceptualised as subsets of the real numbers characterised by normal convex membership functions: \( \mu_A : \mathbb{R} \rightarrow [0, 1] \), such that \( \mu_A(x) = 1 \) for some \( x \) and \( \mu_A(\lambda x_1 + (1 - \lambda) x_2) \geq \lambda \mu_A(x_1) + (1 - \lambda) \mu_A(x_2) \), where \( \mu_A(x) \) represents the degree up to which the real number \( x \) satisfies the property represented by the fuzzy number \( A \). The main types of fuzzy numbers in the literature are triangular, trapezoidal and bell shaped numbers \([40]\). A triangular fuzzy representation of the concept ‘about \( x_2 \)’ is shown in Fig. 1 with membership function given in (4) that it is represented herein as \((x_1, x_2, x_3)\).

\[
\mu_A(x) = \begin{cases} 
0, & x \leq x_1 \\
\frac{x-x_1}{x_2-x_1}, & x_1 < x \leq x_2 \\
\frac{x_3-x}{x_3-x_2}, & x_2 < x \leq x_3 \\
0, & x > x_3 
\end{cases} \tag{4}
\]

Before applying fuzzy logic to the scenario analysis, it has to decided which definition of scenarios experts will provide. Durfee and Tselykh \([16]\) represent each scenario with three real values corresponding to an optimistic, realistic and pessimistic estimation for its frequency and severity variables. Since estimation from experts tend to be provided linguistically rather than numerically, this paper develops a fuzzy approach for modelling, combine and integrate scenarios with internal data.
3.4. Integrating scenarios to obtain aggregated losses

Assuming the previously stated three values representation of scenarios, the integration process starts with the mixing of scenarios and internal data. First, it is important to clarify that the scenarios, for example for the severity values of the risk events, are generated within a unit of measurement. A unit of measurement is a business line, a risk event type, the intersection of a business line and a risk event type, or the entire company. Crisp and/or fuzzy membership functions can be used to represent the aforementioned three values.

3.4.1. Integration of crisp scenarios

The objective of using crisp scenarios in this paper is to allow the comparison of results with the use of fuzzy scenarios. By crisp scenarios we refer to scenarios whose three values representation is based on the use of crisp numerical values, which according to [20] is as follows:

- Severity is represented with a range \([a, b]\). The following notation is actually used \([a, m, b]\): \(a\) is the optimistic estimation value, \(m\) is the realistic estimation value, and \(b\) the pessimistic estimation value.

- Frequency is denoted by \(1/t\), where \(t\) is the number of years, in which an event occurs once.

Notice that scenarios cannot be simply added to the internal data of the institution unless internal data use the same scale for frequency. Thus, as per [20], a frequency normalisation of scenarios, with internal data number of years time frequency \(T\), is required. In other words, if an event \(i\) occurs \(n\) times in a number of years \(t_i\), then its internal data normalised frequency will be

\[
\bar{F}_i = \frac{nT}{t_i}
\]  

(5)

In addition to the above, severity ranges of scenarios may overlap. Frequency values of scenarios with disjoint severity ranges remain unchanged, which is not the case for scenarios with overlap severity ranges, and cumulative common frequencies (CCF) to account for the overlap are computed as follows [20]:

Figure 1: Triangular fuzzy number ‘about \(x_2\)’
1. Assume that the set of scenarios ordered in increasing order of their severity range lower bounds be \{SC_1, SC_2, \ldots, SC_r\}, i.e. \(a_1 \leq a_2 \leq \cdots \leq a_r\).

2. The percentage of overlap of the severity ranges of ordered scenarios are stored in the following matrix \(R = (r_{ij})\):

\[
r_{ij} = \begin{cases} 
\frac{l_i + l_j}{b - a} - 1 & \text{if } j \leq i \text{ and } l_i + l_j \geq a + b \\
0 & \text{otherwise.} 
\end{cases}
\]

where \(l_i = b_i - a_i\), \(l_j = b_j - a_j\), \(a = \min\{a_i, a_j\}\) and \(b = \max\{b_i, b_j\}\).

3. The cumulative common frequency (CCF) for scenarios are computed as follows

\[
CCF_i = \sum_j r_{ij} \cdot \bar{F}_j.
\]

The CCF for the scenarios can be added to the historical internal data. The aggregation can be done by adding them directly to the data, looking for frequency ranges like those of the scenarios and finding the frequency direction by analysing the historical (internal) data and the scenarios. This way, only the difference is added. Another form of aggregation consists in constructing intervals of frequencies to generate the random numbers for the Monte Carlo simulation, which can be added in the already created intervals. Any of these two methods can be very ‘invasive’ and increase the aggregate distribution in a very abrupt manner, significantly changing the VaR. This crisp way was proposed in [20] and it is used in the present paper to compare the results against the use of the proposed fuzzy representation, as described in the following section.

3.4.2. Integrating scenarios by using fuzzy logic

The fuzzy approach proposed in this paper model scenarios using triangular fuzzy numbers rather than crisp numbers or crisp intervals. Following the method [20], described above, we propose and develop a novel approach to model overlap fuzzy severity values of scenarios, for which the union of fuzzy sets operation is used.

Recall that given two triangular FNs \((x_1, x_2, x_3)\) and \((y_1, y_2, y_3)\) corresponding to the severity of scenarios \(SC_1\) and \(SC_2\), respectively, their union will be the fuzzy set with membership function

\[
\mu_{SC_1 \cup SC_2}(x) = \max \{\mu_{SC_1}(x), \mu_{SC_2}(x)\}.
\]

The application of the union to more than one triangular FN is illustrated in Fig. 2. Figure 2(a) represents an four triangular fuzzy numbers, three of which overlap \((SC_1, SC_2, SC_3)\), while Fig. 2(b) illustrates their union fuzzy set.

Considering the nature of the problem to be solved, it is not convenient to perform a union of all the scenarios. When integrating the scenarios with the internal data, a summation of severity values can occur, generating a very large interval, which is not a good representation of the different frequencies of the scenarios. This is why, each set of the overlapping intervals will be merged into one, but those that do not relate to each other will remain separate. This way, the union will produce an outcome, not in terms of a fuzzy set, but as a set of fuzzy sets. In other words, each section of the membership function defined in (8) where \(\mu(X) = 0\) means that a disjoint FN exists, and causes a division of fuzzy sets. Figure 3 shows this approach to dealing with overlapping and disjoint severity FNs fuzzy set, where the union of the first three overlapping triangular FNs is generated separately from the fuzzy set corresponding to the last disjoint FN.
Figure 2: Example of the Union of four triangular FNs

Figure 3: Example of a Union of Scenarios
The integration of the results of a union of the scenarios depends on the frequency of each risk event defined by the scenario. In this case the adjustment of the VaR is performed at the end of a Monte Carlo simulation after obtaining the loss distribution by applying a Monte Carlo simulation with internal data. The aggregated losses are grouped by intervals and the scenarios are associated with the interval in which they fall, and the centre of the corresponding fuzzy set is multiplied by the normalised frequency and added to comprise one year, as this is normally the period that is applied for operational risk measurement.

As the scenarios represent extreme risk events with a very high severity, it is possible that there are some scenarios with values that are higher than the maximum of the upper bounds of the internal data severity values. In this case, an interval corresponding to such scenario is added.

4. Results and Discussion

In this section we present a comparison of the results of the ‘classic’ computation of VaR by using only internal data, the VaR determined as an outcome of the combination of internal data and crisp scenarios as proposed in [20], and the VaR obtained using the proposed fuzzy modelling of scenarios. For this comparison, a single scenario is used, although different ranges of severity values are implemented to show the increase of the VaR in each of the three forms of measuring operational risk. Although the objective of this paper consists in the fuzzy combination of different scenarios, some independent scenarios are also used. Since a scenario represents an extreme event with high severity, the event of the maximum loss in the internal data is taken as a reference. The ranges for the scenarios are selected with respect to the highest values. Tests were carried out with scenarios where the range is included in the internal data to guarantee that in these cases a scenario can be interpreted just as additional internal data.

As mentioned in Section 3.1, the database includes four internal data sets. It is worthwhile mentioning that the VaR of a company normally is required for a period of one year; that is why the original periodicity of the data sets of the database was adjusted to one year. Figure 4 shows the representation of the aggregated loss data for each data set and its respective OpVaR based on the internal data only, which was calculated using the Monte Carlo method with 20000 iterations.

Table 2 shows the scenarios used for the comparison of the proposed method. The scenarios are described by their periodicity in years (T), their frequency in that period (Freq), and their minimum (lower bound) and maximum (upper bound) severity (Severity_Min and Severity_Max).

The first scenario occurs every two years, the second once a year and the last once every 10 years.

Table 2: Example of scenarios with their frequency (Freq) per period of time (T) and severity range

<table>
<thead>
<tr>
<th>Nro</th>
<th>T</th>
<th>Freq</th>
<th>Severity_Min</th>
<th>Severity_Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>15000</td>
<td>20000</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>22000</td>
<td>30000</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>1</td>
<td>25000</td>
<td>35000</td>
</tr>
</tbody>
</table>

Table 2 has three scenarios with losses outside of the internal range of losses. So, these can be considered extreme losses with different frequencies, as described above.

The expectation is that the scenarios extend the tail of the aggregate loss distribution, and consequently increase the OpVaR. However, the average losses stay the same. While, if we add
scenarios with a loss range that is not too extreme because some of the corresponding losses are in the range of the internal data, the expectation is that the average losses will change, without increasing the tail.

The scenarios in the table were tested as a set. As can be observed from the numbers in the table, the second and third scenario overlap, while the first scenario is disjoint to the second and third. In the analysis of these scenarios, first it is necessary to perform a decomposition of the union of the scenarios before carrying out a defuzzification. The union based overlapping method proposed in Section 3.4.2 leads to two different fuzzy sets, which are depicted in Fig. 5. The frequency normalisation per one year period means that Freq/T is used as maximum membership of the severity in such period, which in turn implies that the corresponding FNs are cut at that height, so their aggregated value will be truncated as shown by the coloured shaded areas of their union.

The values for the scenarios can lead to different severity values, depending on the kind of defuzzification implemented. In this paper the centroid defuzzification is used. The Monte Carlo simulation with defuzzification is is used when it is necessary to generate a loss in the scenario intervals.

Figure 6 shows the aggregated losses based on the scenarios of Table 2. The annual aggregated loss distribution based on the internal data only is plotted in black; the annual aggregated loss distribution based on internal data with scenarios using the classical method is plotted in blue; while the annual aggregated loss distribution with internal data with the fuzzy scenarios proposed in this paper is plotted in red. In addition, the average loss is indicated by corresponding coloured dashed vertical lines and the OpVaR at the 99.9th percentile with corresponding coloured solid vertical lines.
Figure 5: Combination of fuzzy scenarios, considering their frequency

Figure 6: Comparison of OpVaR without scenarios, with non-fuzzy scenarios and with fuzzy scenarios
It can be observed that, as expected, the OpVaR based on the proposed fuzzy scenarios is higher than the OpVaR based solely on with internal data but a bit more conservative than the OpVaR with non-fuzzy scenarios.

Regarding the tails of the distributions, the addition of scenarios in both cases (fuzzy; non-fuzzy) causes an increase. The most interesting aspect regards the average of losses. Although the average loss changes (increases) when applying the proposed fuzzy method, the shift of the expected loss is much lower than the one obtained with the classical approach. It is intuitive that the average of the aggregate losses (expected loss) is affected when we add scenarios. We might also expect, when representing the most likely losses and extreme events through the scenarios, that the average will only move a little compared to the average obtained using only internal loss data. However, it is evident that this will depend on the generated scenarios and the frequency that is assigned to them. In the specific case analysed in this experiment, with the given frequencies per year for the first two scenarios, it is expected that if these events occur at least once a year, the annual loss will change.

![Figure 7: Comparison of OpVaR and mean loss with non-fuzzy scenarios and with fuzzy scenarios](image)

To complement the previous experiment, additional results were obtained with ten different scenarios, as illustrated in Figure 7. These scenarios were selected with lower bounds ranging from 15000 to 25000 with upper bounds 10,000 higher. The objective was to measure the influence of the scenarios when added to internal data using the classical and the proposed fuzzy approach proposed herein. In both cases, an increase of the losses of the scenarios leads to an increase of the OpVaR. However, the rate of increase of the average loss is lower when applying the fuzzy method. Consequently, this experiment confirmed that the non-fuzzy method delivers a worse picture than the fuzzy method with regard to the expected losses caused by operational risk events. However, both methods and their comparison are useful for an institution because they help to better analyse and understand the range of extreme losses, which helps to determine whether these are in line with the defined risk appetite of the institution or threatening the business continuity of an organisation.

5. Conclusions

This paper highlighted the contribution of scenarios as a data source for operational risk measurement. Indeed, scenarios refer to possible futures that represent the magnitude of possible losses due to operational risk. Consequently, scenarios add a forward looking perspective to risk measurement. Scenarios as a data source complement historical data about operational risk events and their losses that are typically part of the databases of an organisation.
Since scenarios are constructed by experts and most of the time they are difficult to assess quantitatively, it is reasonable to study the use of fuzzy logic approaches to modelling the vagueness inherent to this process as well as to the qualitative character of the information they represent. Indeed, this is an advantage of the proposed model. As such, this paper presented a model to add scenarios by using a fuzzy representation of risk events. This process of adding scenarios as fuzzy sets is based on the union of fuzzy sets, the frequency alpha cut of fuzzy scenarios, and the use of the centroid defuzzification method.

In terms of the advantages of the proposed model, the most significant one is the representation of scenarios with fuzzy sets for a smooth integration with historical data. Considering that the scenarios are a representation of possible future risks, generally catastrophic, these can make a drastic difference when assessing risk.

The experimental results showed that, as expected, adding scenarios that often represent extreme risk events increases the tail of the aggregate loss distribution curve. The fuzzy approached proposed in the paper was compared with the classical method of adding non-fuzzy scenarios, and it was observed that the OpVaR estimate is more conservative in the case of applying fuzzy scenarios. It was also shown that the rate of increase of the average loss by adding fuzzy scenarios is lower than the classical method of adding non-fuzzy scenarios. The proposed method enriches the alternatives for estimating the OpVaR in companies. Hence, operational risk can be analysed and measured based on a range of OpVaR values, beginning with losses that are only reflected through the internal data (historical data) of an organisation, which can subsequently extended by adding non-fuzzy or crisp scenarios to account for the future and for more extreme risk events. Eventually, we can add fuzzy scenarios to produce more conservative values, in terms of the average loss and OpVaR, which means that the organisation will be required to hold less capital.

As future work, this model of fuzzy scenarios could be applied in the risk assessment of transition risks, which arises from the transition to a low-carbon economy. To size climate-related financial risks, organizations require plausible ranges of scenarios to assess the potential impacts of transition risk drivers on their exposures [41, 42]. Considering that climate-related risk event scenarios are uncertain, a fuzzy representation for them is a promising approach when integrating them in a quantitative risk evaluation.

References

URL https://www.bis.org/publ/bcbs128.htm


URL https://ideas.repec.org/p/red/sed016/352.html


URL https://cran.r-project.org/web/packages/OpVaR/index.html


URL https://www.bis.org/bcbs/publ/d517.pdf