Universidade de Brasília
Faculdade de Economia, Administração,
Contabilidade e Gestão de Políticas Públicas
Departamento de Economia

Copulas for Operational Risk: a case study

Vinícius de Oliveira Watanabe

Brasília, DF
2022

Vinícius de Oliveira Watanabe

Copulas for Operational Risk: a case study

Projeto de Pesquisa da Monografia a ser apresentada como requisito parcial à obtenção do título de Bacharel em Ciências Econômicas pela Universidade de Brasília.

Orientador: Daniel Oliveira Cajueiro

Co-orientador: Herbert Kimura

Brasília, DF

2023

Vinícius de Oliveira Watanabe

Measuring Operational Risk with Copulas: An Application to Banking

Projeto de Pesquisa da Monografia a ser apresentada como requisito parcial à obtenção do título de Bacharel em Ciências Econômicas pela Universidade de Brasília.

Dr. Daniel Oliveira Cajueiro,

Departamento de Economia – Universidade de Brasília

Dr. Herbert Kimura,

Departamento de Administração – Universidade de Brasília

03 de agosto de 2022

RESUMO

O conceito que norteia a definição de Risco Operacional foi estabelecido pelo Comitê de Supervisão Bancária de Basileia. Esta definição entende o Risco Operacional como “possibilidade de ocorrência de perdas resultantes de falha, deficiência ou inadequação de processos internos, pessoas e sistemas, ou de eventos externos"(BCBS, 2004). Existe três abordagens para a modelagem do Risco Operacional: *Basic Indicator Approach*, *Standardized Approach* e *Advanced Measurement Approach*. Dentre estas opções, a última se tornou a mais utilizada e pesquisada ao longo das últimas duas decadas. Loss Distribution Approach (LDA) é um dos metodos que são pertencentes a última abordagem citada, modelando o risco operacional por meio das distribuições marginais da severidade e da frequência. Entretanto, esta metodologia assume independencia entre as duas variáveis, gerando um resultado superestimado. O uso de *Copulas* serve como uma forma de resolver este problema, e como mostrou este estudo, gerou uma economia que ficou entre 20 a 50% quando comparado ao LDA.

Palavras-chave: Risco Operacional. Copulas. Gaussian-Copula. Distribuições Marginais. Finanças.

ABSTRACT

The concept that guides the definition of Operational Risk was established by the Basel Committee on Banking Supervision. This definition understands Operational Risk as "the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events"(BCBS, 2004). There are three approaches to modeling Operational Risk: Basic Indicator Approach, Standardized Approach, and Advanced Measurement Approach. Among these options, the latter has become the most widely used and researched over the past two decades. The Loss Distribution Approach (LDA) is one of the methods belonging to the last mentioned approach, modeling Operational Risk through the marginal distributions of severity and frequency. However, this methodology assumes independence between the two variables, resulting in an overestimated outcome. The use of Copulas serves as a way to solve this problem, and as shown in this study, as it generated savings that ranged between 20 to 50% when compared to LDA.
Keywords: Operational Risk. Copulas. Gaussian-Copula. Marginal Distributions. Finance.

Lista de Tabelas

Sumário

1 Introduction

Operational Risk is a term that gained more popularity in 2000s. This type of risk is very broad since it considers every risk that's not classified as market or credit risk. Among many definitions, the most famous is the definition that was settled by Basel Committee (BCBS, 2004) which states that Operational Risk is "the risk of direct or indirect loss resulting from inadequate or failed internal processes, people and systems, or from external events.".

There are three approaches for calculation of operational risk. The most famous one is Advanced Measurement Approach (AMA), which gives more freedom to managers to use their own models. One of the most used approaches under AMA is Loss Distribution Approach (LDA). This method, in it's standards applications, presents a problem: it considers that frequency and severity are independent. LDA tends to overestimate capital charge by this inefficient assumption. Copulas can model structure dependence between two or more variables, helping managers to save money by reducing capital charge estimates.

There are two big Copulas Families: Archimedean and Elliptical. Elliptical Copulas usually are the most used in Operational Risk modeling. Archimedean can't model negative covariance, which can lead to problems since some events presents negative covariance between frequency and severity. Overall, Elliptical Copulas are usually the choice for this task. For that reason, Gaussian Copulas and T-Copulas were briefly explained later in this work.

In section 2, a brief literature review is made, explaining basic knowledge about Operational Risk and methods to estimate capital charge. In section 3, the Loss Distribution Approach is described, and in subsection 3.1 regulatory capital measurement and two measures: VaR and CVaR are detailed. Section 4 presents a brief definition of Copulas setting the basic knowledge to understand how it works. In section 5, an empirical analysis is made, first by defining the marginal distributions for frequency and severity of each event. Two main tests were done: a standard LDA and Copulas-based, comparing it's results to check which performed better. As literature pointed, our test showed that Copulas performed better for the estimation of capital charge, VaR and Expected Shortfall. Then a simple test was made: to check how copulas would behave when it's parameters were changed (covariance matrix) and how close it's mean would be to the standard LDA (when both models assumed covariance = 0). The last test was to check how sensible was the

Copulas model to the marginal distributions. It was shown that the marginal distribution made a big impact over the copulas results.

2 Literature Review

This section provides background information for subsequent sections of this work by explaining the concept of operational risk and discussing recent literature on methods for measuring it. Specifically, we will present the official definition of operational risk as defined by BCBS(BCBS, 2004) and a more specified definition by Jarrow (2008).

2.1 Operational Risk

After Basel Committee in 1988, in 2002 the Basel II Accord stated a significant view on risk: capital charge also, implicitly, covered operational risk (LU, 2011). Operational Risk was defined by the Basel Committee as "the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events" (BCBS, 2004). Losses pointed by Basel II are heterogeneous and classified in a matrix 7x8, that is, seven event types and eight business lines.

The Basel Committee recommends three approaches to assess capital requirements for operational risk, and does not imposed any restriction on the methods that can be used. The Basic Indicator Approach (BIA) and the Standardized Approach (SA) follow a "Top-down" perspective, whereas the Advanced Measurement Approach (AMA) follows a "bottom-up" perspective. As Linkov et al. (2014) pointed, traditional risk assessment is bottom-up, it starts with collected data and end with risk estimates. Top-down is simply a decision analysis, it starts with objectives and goals and ends with decision making. The Basic Indicator Approach (BIA) and Standardized Approach (SA or TSA) are fundamentally distinct from the Advanced Measurement Approach (AMA). The AMA relies on internal data and measurement methodologies developed by the organization itself, whereas the BIA and SA are pre-determined by regulatory authorities (SUNDMACHER, 2007). In contrast to the AMA, which allows organizations to tailor their approach to their specific operations and risk profiles, the BIA and SA are more standardized and do not provide the same level of flexibility, thus being a more common choice.

In this context, some financial institutions have explored AMA (AZAR; DOLATABAD, 2019). Within this approach, Loss Distribution Approach (LDA) is one of the most studied and cited methods over recent literature. For example, recent literature has examined the application of LDA in various contexts, including Frachot, Georges e Roncalli (2001), Shevchenko (2010), Zhou, Durfee e Fabozzi (2016). These studies highlight the benefits of using AMA and, specifically, LDA in operational risk context.

LDA is a method that allows the estimation of the final loss distribution by combining a frequency distribution of events of operational risk and a severity distribution of monetary losses of individual operational events. Since many LDA models use historical data of operational losses, without relying on an a priori hypothesis to connect operational risk events and losses, they can be classified as a data driven model (AZAR; DOLATABAD, 2019)

The LDA approach can be challenging, especially with regards to the integration with the day-to-day risk management, as there can be issues on data availability and adequacy (FRACHOT; GEORGES; RONCALLI, 2001). Nevertheless, with sufficient data, the statistical model can approximate the loss distribution with enough accuracy, making results useful from a managerial point of view.

Jarrow (2008) separates operational risk in two types: one that's incurred by the firm's operating system; the second one that incorporates the agency cost (BREALEY et al., 2006), corresponding as a loss due to incentives. That being said, individuals decision making changes over time, making the measure of operational risk harder as the elements are not static.

Although Basel II does not imply any strong restriction over LDA, it does specifies the necessity to include the "use of internal data, relevant external data, scenario analysis and factors reflecting the business environment and internal control systems" (BCBS, 2004). That implies challenges that arise when combining those types of data as shown by Bonet et al. (2021).

As Lu (2011) points, Moscadelli (2004) work served as a initial measurement of severity and frequency from Operational Risk Loss Data Collection Exercise (LDCE) launched in June 2002. The later indicates that the results shows a low performance of conventional severity models. What happens is a fit to the central observations, failing into incorporating the extreme percentiles. Such issue is even worse when Zhou, Durfee e Fabozzi (2016) shows that advanced approaches appoints operational risk exposure as defined at a very high quantile.

BCBS (2009) shows that correlations are used in the modeling process mainly by the use of: copulas, representing 43%; 36% of the banks use Gaussian copulas; 17% use a correlation matrix and 31% use methods other than those above. Copulas method got popular because it offers an alternative for the usual Basel method which considers the loss variables as perfect depended. Therefore, models can measure dependence between variables and work with a more robust method which takes more into account besides the usual multivariate normal distribution.

3 Operational Risk measurement

Actuarial model is a widespread statistical model that can be used (FANTAZZINI; VALLE; GIUDICI, 2008). The probability distribution is described as:

$$F_i(S_i) = F_i(n_i) \cdot F_i(X_{ij})$$

$F_i(S_i)$ is equal to the probability function of the expected loss for risk i; $F_i(n_i)$ is the probability of event/frequency for the risk i; $F_i(X_{ij})$ is the severity for risk i. Fantazzini, Valle e Giudici (2008) points two conditions for the model above:

- the losses are independent and identically distributed (i.i.d.) variables;
- the distribution of frequency and severity are independent.

Klugman, Panjer e Willmot (2012), Moscadelli (2004), and Panjer (2006) points that the most common continuous distributions for severity are Lognormal, Exponencial, Weibull and Gamma. The most common discrete distributions to represent frequency are Poisson and Negative Binomial.

Note that, since Loss Distribution Approach is part of the Advanced Measurement Approach (AMA), it's use is based on internal data, although it can present external data.

3.1 Regulatory Capital measurement

A great part of the sector uses the somatory method to calculate the regulatory capital. There's two kinds of vision of what should the regulatory capital cover

1. Expected and Unexpected losses;
2. Only unexpected losses;

For each business unity there's a definition of what is expected or unexpected loss. Therefore, the definition of Value at Risk (VaR) is necessary since it is one of the most used measures in the financial system. It represents the volatility of what is being measured and, in Operational Risk cases, is the 99,9% percentile of X's probability distribution. The definition can be written as:

Definition 1 (Value at Risk) VaR is the α -quantile of the loss for the i-th risk, α being the significance level:

$$VaR(S_i;\alpha) : Pr(S_i \geq VaR) \leq \alpha.$$

Then VaR represents the maximum loss of a risky intersection i, for a confidence level of $1-\alpha$.

Basel Comittee states that banks must hold capital equal to their unexpected losses with AMA, if they made provisions for the expected loss.

The amount of capital that Basel Comittee states that a bank must cover is equal to the expected losses with AMA, in case there's provision for the expected losses.

While the expected loss can be understood as the loss to the expected value ($E(X)$), unexpected loss is the quantile for the level $\alpha - E(X)$. That is, the losses that stand between $E(X)$ and VaR. After the risk measure for all intersections is computed, usually a sum of all individual measures is done to compute which is the total VaR of the business. Note that, as Fantazzini, Valle e Giudici (2008) points, in this case the model assumes perfect dependence among the losses S_i. Thus, the paper of Copula to describe the dependence structure between the losses is intended to reduce required capital for global VaR.

Definition 2 (Expected shortfall) The Expected Shortfall (ES), also known as Conditional Value at Risk (CVaR) can be an alternative risk measure to the VaR method. As said on (FANTAZZINI; VALLE; GIUDICI, 2008), the ES at the confidence level α is defined as the expected loss for intersection i, given the loss has exceeded the VaR with probability level α:

$$ES(Si;\alpha) \equiv E[Si|Si \geq VaR(Si;\alpha)]$$

Therefore it can be said that when we have a confidence level of 1- α, in a determined time horizon, the value of the average losses is equivalent to the losses that exceeded VaR in 1-α confidence level.

This method, has more sensibility to the tail events. Different from the VaR, the ES captures tail risk. Once we had computed the VaR, is interesting and helpful to also check the ES, as we can cover more precisely all risks, and have a more robust measure of risk.

ES measures the expected value of the losses that occur beyond the VaR level, given that the VaR has been breached. ES provides a more comprehensive measure of risk than VaR, as it captures not only the magnitude of potential losses but also the frequency with which they occur.

4 Copulas

Copulas are gaining more popularity over the years (GIACOMETTI et al., 2008), one of the main reasons is the tendency to reduce total VaR (Value at Risk). It was shown to be a very efficient and important tool in the finance area, especially when the subject is operational risks. This function, sometimes called the dependence function, is used, basically to describe the inter-correlation between different variables.

Furthermore, Copula is a function that model the dependence structure between the variables of a certain vector (NELSEN, 2007). This function was used in Operational Risk context, as seen in Lu (2011) and Fantazzini, Valle e Giudici (2008), in a way that the vector contains the losses of every risk event in the data and when applied to the variable's marginal distributions their multivariate distribution is defined - note that is doesn't mean that the marginal distributions are necessarily equal. That kind of application allows a flexible way to model the dependence structure.

Nelsen (2007) discuss an introduction to copulas. In that sense, for a brief overview, the following subsection pretends to condense the main knowledge for the reader to understand further applications.

4.1 Copulas-based methods

Definition 3 (Copula) *N-dimensional copulas can be defined as a function C with 3 properties as pointed out by Nelsen (2007):*

- Dom C is the unit interval [0,1];
- C' is grounded and N increasing;
- For all $u_i \in [0, 1]$ $C(1, ..., 1, u_i, 1, ..., 1)$

As Fantazzini, Valle e Giudici (2008) point out, these conditions provides the lower bound on the distribution function, and also ensures that the marginal distributions that we have are uniform. The role of copulas as dependence functions is described by Sklar's theorem that Bouyé et al. (2000) and Fantazzini, Valle e Giudici (2008) points in the copula's definition.

Theorem 1 (Sklar's theorem) Let **H** be a N-dimensional distribution function with the margins $F_1, ..., F_N$. In that sense, there is a N-copula C represented as:

$$H(x_1, ..., x_n, ..., x_N) = C(F_1(x_1), ..., F_n(x_n), ..., F_N(x_N))$$

This theorem allow a way to analyze the dependence structure of multivariate distributions. The importance of this theorem is that it allows the modeling between two variables through Copulas. Beyond that, it also implies that is possible to interconnect two or more univariate distributions with a copula and get a multivariate distribution. Nelsen (2007) also points a Corollary that implies that given *any* two marginal distributions and *any* copula it is possible to build a joint distribution.

For instance, Sklar's theorem applied to Operational Risk case indicates that a joint distribution H could be described as:

$$H(S_1, ..., S_R) = C(F(S_1), ..., F(S_R))$$

The equation above is the copula of the cumulative distribution functions of the marginals.

Corollary 1 Let $F^{(-1)}$ and $G^{(-1)}$ be the generalized inverses of the marginal distributions. Then, for every $(u, v) \in I$ we have:

$$C(u, v) = H(F^{(-1)}(u), G^{(-1)}(v),)$$

That corollary that allows the method of building a Copula.

Bouyé et al. (2000) points that in many financial applications the main problem is to find a convenient distribution. That way, Bouyé et al. (2000) restate Frees e Valdez (1998) affirmation that is not obvious to identify the copula. With Copulas the processes of modeling is divided in two non-trivial steps: identification of marginal distributions; and defining the appropriate copula.

The two biggest "families" in Copulas are: Elliptical and Archimedean. The later presents limitations to only model positive dependence, as pointed by (FANTAZZINI; VALLE; GIUDICI, 2008) it also presents strict restrictions on bi-variate dependence parameters. Therefore, Elliptical copulas are the most common Copulas type for modeling Operational Risk. *Normal-copula* and *T-Copula* can both be derived from the Corollary

1 procedure[1]. Both copulas densities can be used to fit the operational risk data with maximum likelihood.

Chen, Fan e Patton (2004)'s approach is a mixed parametric approach. It is based on the Maximum Likelihood estimates and Method of Moments. This approach can be described in four steps:

1. Transform the data into uniform variables by the use of a parametric distribution function or an empirical distribution;

2. Let $\sum^{\wedge}$ denotes the correlation matrix for the Gaussian Copula.

$$\overset{\wedge}{\sum} = \frac{1}{T}\sum_{t=1}^{T}\zeta_t'\zeta_t$$

3. Estimate the v by maximizing the log-likelihood function of the T-Student Copula density:

$$v^{\wedge} = argmax\sum_{t=1}^{T} log(c^{student}(u^{\wedge}_{1,t}, ..., u^{\wedge}_{N,t}; \overset{\wedge}{\sum_{Ga}}, v)$$

Note that $\sum^{\wedge}_{Ga} = \sum^{\wedge}$

4. $\sum^{\wedge}_{T-copula}$ can be obtained by the use of $\sum^{\wedge}$. Let $\zeta_{vt} = (t_v^{-1}(u^{\wedge}_{1t}, ..., t_v^{-1}(u^{\wedge}_{Nt}))$ then we can get $\sum^{\wedge}_{T-copula}$ as:

$$\overset{\wedge}{\sum}_{student} = \frac{1}{T}\sum_{t=1}^{T}\zeta_{vt}'\zeta_{vt}$$

These steps can be further explored in the article by Chen, Fan e Patton (2004), however, other methods will not be detailed here as they are computationally complicated and don't serve well for this work case.

5 Empirical Analysis

The empirical analysis was applied over a anonymous bank data that presents a time horizon between January 1st, 2016 and December 31st, 2020. Thus, Month observations are 60 in total. Also, the data presents three types of events which are anonymous and defined by a code (6,9 and 10). Data was transformed and the loss was multiplied by a random

[1]See Nelsen (2007) for more details.

number so it can preserve the bank's data. The first step for a standard Loss Distribution Approach and for a Copulas-based is to find the marginal distributions. Therefore, the following subsection presents how this selection was made.

5.1 Marginal Distribution Selection

As mentioned before, Poisson and Binomial distributions are the most suitable distributions for frequency (LU, 2011). Fantazzini, Valle e Giudici (2008) showed how Poisson with a 72 observations tend to be more consistent then Negative Binomial for low observation's data, which presented a high MSE and Variation Coefficients, along with 40% of the cases presenting negative estimates for θ. Then, the chosen distribution for modeling frequency was Poisson. The mentioned distribution assumes that losses happen randomly through time. Lachowicz (2016) recall the Poisson distribution as the probability of n losses in time T, which was considered to be months in this work, as:

$$Pr = exp(-\lambda T)\frac{(\lambda T)^n}{n!}$$

λ parameter above was estimated by the average number of losses per month, analogue to what Hull (2012) shows. This estimation was made by the use of Python programming language. λ found for frequency of the given data is around 22.37 for event 6; 10.21 for event 9; and 40.03 for event 10. These parameters were stored and are disposed in Table 2 Random variables were then drawn from these λ with a size of 10^5.

Tabela 1: Poisson Frequency

Poisson	***6***	***9***	***10***
λ	22.37	10.21	40.03

Severity was analyzed by event risk type. The common distributions, as pointed earlier in this work, are Gamma and Log-normal, with some rarer cases as Pareto. Log-normal distribution was the chosen one for the three types of event. The three distributions were fitted and compared with the package *Fitter* from Python. Since data is scarce in operational risk, it is difficult to choose which distribution to follow. For that reason, the criteria used was Kolmogorov-Smirnov statistic, which has been used for goodnes-of-fit testing for decades (DREW; GLEN; LEEMIS, 2000). In a simple way, it is used to identify if the sample we have came from the theoretical distribution.

The fit was done for the severity of each event type, since is a special case, Gamma and Log-normal distributions were compared. In the fitting step, the Kolmogorov-Smirnov statistic is lower in the Log-normal distribution, eventually, Log-normal distribution was the one chosen to model severity distribution in all event types presented in the data.

To continue the analysis it is required to store the parameter values of $'s'$, $'scale'$ and $'locale'$ from the fit. Table 2 shows the parameters results from each event.

Tabela 2: Parameters from Fitter

Parameter\Event	***6***	***9***	***10***
's'	0.003516957817371905	0.004435205787612206	0.35775167328096374
'scale'	463.3171169512617	-458.2313003545787	-0.07510232629328495
'locale'	-456.46632098862665	466.4658846290197	3.96811427516901

The Log-normal distribution probability density function (pdf) can be written as:

$$p(x;\sigma,loc,scale) = \frac{1}{x\sigma\sqrt{2\pi}}exp\left[-\frac{1}{2}\left(\frac{logx}{\sigma}\right)^2\right]$$

Where the $x = (y - loc)/scale$ as pointed out by (LACHOWICZ, 2016).

Since both frequency and severity for each event type were determined and explained, the next step subsection will be an estimation of a standard Loss Distribution Approach (LDA), considering that both Frequency and Severity are independent. By this exercise, we can compare the capital charge output from a standard LDA method and a Copulas-based one.

5.2 Standard Loss Distribution Approach

Since a more detailed discussion over Loss Distribution Approach (LDA) was made before in this work, the subsection only presents the main steps to the capital charge estimation.

Only four packages from Python were used for this exercise: Scipy; Matplotlib; Numpy; and Pandas. In accord to what is common in literature, the significance level α used was 1% leading to a confidence level of 99%. Steps taken are the ones suggested by Hull (2012) and inspired by (LACHOWICZ, 2016) which consist of:

1. Generate a sample from the *frequency* distribution as a method to determine loss events' numbers.

2. Generate n samples from the loss *severity* to determine the loss for each loss event i, (L_1, L_2, L_3)

Tabela 3: Standard LDA results

Standard LDA	***6***	***9***	***10***
mean	3501.5619869601587	30928.386293341242	347.37536190694686
VaR	46173.2910641196	558760.7255271096	4828.464494853837
capital charge	42671.72907715944	527832.3392337683	4481.0891329468905
expected shortfall	63477.78734866325	1445597.5760668376	10235.603889775757

3. And then sum all L_i and determine the total loss experienced.

The number of simulations done were 10^5. A random number was drawn from a uniform distribution, if this number was less then the value from the Poisson Cumulative Distribution Function (CDF) with the given Lambda for the x losses' number then a zero loss was assumed. If the number was greater then the value from the CDF then a random variable was drawn from the Log-normal distribution (for the given event). This process was made for each event type and the *mean*, *VaR with 99% confidence level*, *capital charge*, and *expected shortfall* were calculated for each event. Results can be seen in Table 3.

The capital charge can be obtained from the unexpected losses. So, the capital charge can be calculated as the difference between the 99% percentile of the Modelled Loss Distribution and the Excpected Loss. So the capital charge for the given data can be measured as the distance between Value-at-Risk (VaR) and the Mean. Expected Shortfall is the average loss after VaR.

5.3 Copulas-based

Most important difference between this approach and the standard is that the latter assume independence between loss frequency and loss severity. Therefore, Copulas can be used to model the structure dependence.

In this subsection I'm still assuming that frequency follows a Poisson distribution, and that severity follows a log-normal distribution. Since variables can present a positive dependence, Archimedean copulas were not considered in this sample. More over, the Copula used was the Normal Copula, also known as Gaussian Copula. It is a common choice for modeling severity and frequency in Operational Risk (FANTAZZINI; VALLE; GIUDICI, 2008).

Gaussian Copulas presents only one parameter: covariance matrix. Therefore, the covariance matrix for each risk was computed and stored. These three covariance matrix are presented in Table 4, Table 5 and Table 6.

After fitting the data to the Gaussian Copula, the next step is to generate random

Tabela 4: Covariance Matrix for Event 6

Covariance	Matrix
1.000000	-0.073725
-0.073725	1.000000

Tabela 5: Covariance Matrix for Event 9

Covariance	Matrix
1.000000	-0.13596431
-0.13596431	1.000000

variables from the dependence structure combined with the marginals distributions. Results are displayed in Table 7.

When comparing VaR from the Standard LDA and from GaussianCopula approach it shows an approximate reduction of 23% for event 6, 30% for event 9 and 46% for event 10. Capital charge had a approximate reduction of 24% for event 6, 31% for event 9 and 47% for event 10. This reduction is close to what was observed in the literature[1]. Another metric that we can use to compare Copulas efficiency is the Expected Shortfall. Besides the event 6, which presented a expected shortfall 5% higher, all events presented a significant reduction of expected shortfall. In event 9 the reduction was approximately of 40% and event 10 presented a reduction of 107%.

One point that is worth observing is that the correlation wasn't way too distant from zero, which can indicate that there is no perfect correlation. Also, covariance matrix for the given case was negative, which is not necessarily the overall case, but caution must be taken if a Archimedean Copula is the choice for modeling dependence structure.

5.4 Zero Covariance Between Severity and Frequency

Although zero covariance doesn't mean independence between two variables, it is a good proxy to check if a simulation from a Gaussian Copula can approximate to the standard LDA. If covariance lying in the covariance matrix between frequency and severity is set to 0, in theory, the result should approximate to the result from the standard LDA simulation. This test was applied to all three events, which gave us the parameter that can be seen in Table 8 for the Gaussian Copula.

The results shows a small distance between the mean observed in standard LDA and the zero covariance copula. Distance for the mean in event 6 was -2.22%, in event 9 it was 4.61%, and in event 10 it was 0.35%. Since the data was generated randomly, it was

[1]Fantazzini, Valle e Giudici (2008) found savings between 30 and 50%.

Tabela 6: Covariance Matrix for Event 10

Covariance	Matrix
1.000000	-0.024262
-0.024262	1.000000

Tabela 7: Results from GaussianCopula

GaussianCopula	6	9	10
mean	3187.810003780882	26350.815630778772	235.60673017850783
VaR	35281.172037978286	390787.19403882226	2591.639475960621
capital charge	32093.362034197402	364436.37840804347	2356.032745782113
expected shortfall	66931.89692421933	1029017.4249754763	4946.666379168993

expected that the distance between the mean observed in each simulation was not equal to zero.

5.5 Marginal Distribution's Importance

As mentioned before, the most common distributions for Frequency are Poisson and Negative Binomial. For severity it is Lognormal, Gamma, Weibull and, in some cases, Pareto. What was tested is "how much the results change if we pick different marginal distributions?". In the first LDA simulation made in this work, we assumed that Frequency had a Poisson marginal distribution. For this test we are going to suppose that Frequency follows a negative binomial, which the probability mass function can be described as:

$$Pr(X = k) = \left(\frac{k + r - 1}{k}\right)(1 - p)^k p^r$$

This distribution presents two parameters: *r* and *p*. These parameters can be described in terms of the mean μ and variance σ^2:

$$p = \frac{\mu}{\sigma^2}$$

$$r = \frac{\mu^2}{\sigma^2 - \mu}$$

From the data, the μ and σ^2 from every event was computed to calculate *r* and *p* for each Frequency. These results can be seen in Table 9.

These parameters were then used to generate random samples for each event's frequency. Later on, a Gaussian Copula was applied over this new distribution and the Log-normal that describes Severity. The results were compared to the outputs from the

Tabela 8: 0 Covariance

Covariance	Matrix
1	0
0	1

Tabela 9: Negative Binomial Parameters

Negative Binomial	**6**	**9**	**10**
p	0.0211	0.0100	0.0007
r	0.2965	0.1286	0.4054

GaussianCopula that used Poisson as frequency's marginal distribution. For event 6 the capital charge was 36.41% higher, for event 9 capital charge was 37.57% higher, for event 10 it was 131.17% higher. As observed, the marginal distribution chosen for frequency made a big difference.

6 Final Considerations

This work's main goal was to apply both standard LDA and Copulas-based modeling to a given bank data. Since Operational Risk data is scarce, models' implementations are difficult to happen given that most data stands behind banks' systems. What was shown is that Gaussian Copulas presented lower capital charge, which translates to savings for companies.

Gaussian Copulas presents one parameter: covariance matrix. This parameter, in some cases, presented a negative covariance between frequency and severity. Although it may not be the overall case, managers should be careful when using Archimedean Copulas since it doesn't accept negative covariance.

We did fitting tests over severity to see which marginal distribution would be the best. Since the decision isn't clear, given that Operational Risk usually presents scarce data, this decision was made by Kolmogorov-Smirnov Statistics. The best fit, by this metric, was Log-normal. For frequency, there are two common distributions: Poisson and Negative Binomial. It was shown by literature that Negative Binomial doesn't work well with small samples. Thus, that's why Poisson was the chosen one. Even-though, later on the empirical analysis, Negative Binomial was applied to the model, so we can see how sensible is the Copulas model to the marginal distributions. As expected, the capital charge increased over 30% for every event when compared to the Poisson's marginal distribution. Once again managers should take very careful steps when selecting the best distribution for modeling it's case.

Referências

AZAR, A.; DOLATABAD, K. M. A method for modelling operational risk with fuzzy cognitive maps and bayesian belief networks. **Expert systems with applications**, Elsevier, v. 115, p. 607–617, 2019.

BCBS. **International Convergence of Capital Measurement and Capital Standards: A Revised Framework**. Basel Committee on Banking Supervision, 2004. (Basel Committee Publications). Disponível em: <https://books.google.com.br/books?id=HO16RdIa-9wC>.

BCBS. **Observed range of practice in key elements of Advanced Measurement Approaches (AMA)**. [S.l.]: Bank for International Settlements Basel, 2009.

BONET, I.; PEÑA, A.; LOCHMULLER, C.; PATIÑO, H. A.; CHICLANA, F.; GONGORA, M. Applying fuzzy scenarios for the measurement of operational risk. **Applied Soft Computing**, Elsevier, v. 112, p. 107785, 2021.

BOUYÉ, E.; DURRLEMAN, V.; NIKEGHBALI, A.; RIBOULET, G.; RONCALLI, T. Copulas for finance-a reading guide and some applications. **Available at SSRN 1032533**, 2000.

BREALEY, R. A.; MYERS, S. C.; ALLEN, F.; KRISHNAN, V. S. **Corporate finance**. [S.l.]: McGraw-Hill/Irwin Boston et al., 2006. v. 8.

CHEN, X.; FAN, Y.; PATTON, A. J. Simple tests for models of dependence between multiple financial time series, with applications to us equity returns and exchange rates. **London Economics Financial Markets Group Working Paper**, n. 483, 2004.

DREW, J. H.; GLEN, A. G.; LEEMIS, L. M. Computing the cumulative distribution function of the kolmogorov–smirnov statistic. **Computational statistics & data analysis**, Elsevier, v. 34, n. 1, p. 1–15, 2000.

FANTAZZINI, D.; VALLE, L. D.; GIUDICI, P. Copulae and operational risks. **International Journal of Risk assessment and management**, Inderscience Publishers, v. 9, n. 3, p. 238–257, 2008.

FRACHOT, A.; GEORGES, P.; RONCALLI, T. Loss distribution approach for operational risk. **Available at SSRN 1032523**, 2001.

FREES, E. W.; VALDEZ, E. A. Understanding relationships using copulas. **North American actuarial journal**, Taylor & Francis, v. 2, n. 1, p. 1–25, 1998.

GIACOMETTI, R.; RACHEV, S.; CHERNOBAI, A.; BERTOCCHI, M. Aggregation issues in operational risk. **Journal of Operational Risk**, Citeseer, v. 3, n. 3, 2008.

HULL, J. **Risk management and financial institutions,+ Web Site**. [S.l.]: John Wiley & Sons, 2012. v. 733.

JARROW, R. A. Operational risk. **Journal of Banking & Finance**, Elsevier, v. 32, n. 5, p. 870–879, 2008.

KLUGMAN, S. A.; PANJER, H. H.; WILLMOT, G. E. **Loss models: from data to decisions**. [S.l.]: John Wiley & Sons, 2012. v. 715.

LACHOWICZ, P. **Computation of the Loss Distribution not only for Operational Risk Managers**. Quant at Risk, 2016. Disponível em: <https://quantatrisk.com/2016/06/05/loss-distribution-operational-risk-severity-frequency-python/>.

LINKOV, I.; ANKLAM, E.; COLLIER, Z. A.; DIMASE, D.; RENN, O. Risk-based standards: integrating top–down and bottom–up approaches. **Environment Systems and Decisions**, Springer, v. 34, p. 134–137, 2014.

LU, Z. Modeling the yearly value-at-risk for operational risk in chinese commercial banks. **Mathematics and Computers in Simulation**, Elsevier, v. 82, n. 4, p. 604–616, 2011.

MOSCADELLI, M. The modelling of operational risk: experience with the analysis of the data collected by the basel committee. **Available at SSRN 557214**, 2004.

NELSEN, R. B. **An introduction to copulas**. [S.l.]: Springer science & business media, 2007.

PANJER, H. H. **Operational risk: modeling analytics**. [S.l.]: John Wiley & Sons, 2006.

SHEVCHENKO, P. V. Implementing loss distribution approach for operational risk. **Applied Stochastic Models in Business and Industry**, Wiley Online Library, v. 26, n. 3, p. 277–307, 2010.

SUNDMACHER, M. The basic indicator approach and the standardised approach to operational risk: an example-and case study-based analysis. **Available at SSRN 988282**, 2007.

ZHOU, X.; DURFEE, A. V.; FABOZZI, F. J. On stability of operational risk estimates by lda: From causes to approaches. **Journal of Banking & Finance**, Elsevier, v. 68, p. 266–278, 2016.