



Analysis of an aggregate loss model in a Markov renewal regime



Pepa Ramírez-Cobo^{a,*}, Emilio Carrizosa^b, Rosa E. Lillo^{c,d}

^a Departamento de Estadística e Investigación Operativa, Universidad de Cádiz Spain

^b IMUS, Instituto de Matemáticas de la Universidad de Sevilla Spain

^c Departamento de Estadística, Universidad Carlos III de Madrid Spain

^d uc3m-Santander Big Data Institute, Universidad Carlos III de Madrid Spain

ARTICLE INFO

Article history:

Received 3 September 2020

Revised 30 November 2020

Accepted 6 December 2020

Available online 22 December 2020

Keywords:

Loss modeling

Dependent loss times

Markov renewal theory

Overdispersion

Batch Markovian arrival process

PH distribution

Double-Pareto Lognormal distribution

MLE estimation

Operational risk

Value-at-Risk

ABSTRACT

In this article we consider an aggregate loss model with dependent losses. The loss occurrence process is governed by a two-state Markovian arrival process (MAP_2), a Markov renewal process that allows for (1) correlated inter-loss times, (2) non-exponentially distributed inter-loss times and, (3) overdispersed loss counts. Some quantities of interest to measure persistence in the loss occurrence process are obtained. Given a real OpRisk database, the aggregate loss model is estimated by fitting separately the inter-loss times and severities. The MAP_2 is estimated via direct maximization of the likelihood function, and severities are modeled by the heavy-tailed, double-Pareto Lognormal distribution. In comparison with the fit provided by the Poisson process, the results point out that taking into account the dependence and overdispersion in the inter-loss times distribution leads to higher capital charges.

© 2020 Elsevier Inc. All rights reserved.

1. Introduction

The modeling of losses in actuarial science has occupied a prominent role in the last decades. In particular, the estimation of aggregate loss models constitutes a challenge for the insurance industry and financial entities due to the need for a correct assessment of the premium rates and safety margins. The standard definition of an aggregate loss model is as follows. Let X_1, X_2, \dots be independent and identically distributed random variables representing the sizes of individual claims (or severities), and let N count for the total number of losses in a fixed period (frequency). We shall assume that the severities are independent of the occurrence process. Then, the total loss for the time period can be written as the compound model

$$Z = \sum_{i=1}^N X_i. \quad (1)$$

* Corresponding author.

E-mail address: pepa.ramirez@uca.es (P. Ramírez-Cobo).

The usual approach for defining the aggregate loss, which shall be considered in this paper, is to separately model the frequency and the severity of the losses. Then, the distribution of Z is obtained from that of N and the sequence X_1, X_2, \dots . For a thorough description of general aggregate models, the reader is referred to [55] or [40].

One of the most significant applications of loss aggregate distributions is the modeling of the operational risk (OpRisk). The OpRisk constitutes, together with market and credit, the financial risk and results in different types of losses as those generated by natural disasters, frauds, terrorist attacks or security and legal risks. Typically, in the OpRisk modeling, the total loss is defined as in (1), an expression which can be generalized to consider different risk cells or business lines.

Dependence has been a key issue in the recent risk research. The classical risk model has been extended in a number of ways to obtain more general models assuming dependence between claim sizes or inter-claim times, see for example [24]. Dependence has been also considered in multiple ways for OpRisk aggregate loss models. For example, dependencies among frequencies and/or severities can exist within or between cells [22,47,77]. Also, the type of loss may be a factor to generate dependence patterns, as for instance, in [64]. Other authors consider models where the correlation is introduced at the aggregate loss level, as in [15,53]. Copulas constitute a widely used approach for modeling dependence between risk cells, see [7,26,46], to name a few among multiple works. Mixture models have also been considered in the OpRisk literature to introduce dependence among frequencies and severities, as in [65]. Finally, point processes are also one of the tools employed for constructing dependent occurrence processes, as shown in [17,33,58].

As pointed out in [18], an alternative approach for studying the frequency of losses in a OpRisk context is to look into the statistical properties of the inter-loss times. As previously commented, the Poisson process has been widely used for modeling the occurrence of losses. However, it might be the case that the exponential distribution does not properly fit real datasets, and in such cases a more general continuous distribution supported on the positive real line is needed [18]. In addition, although few studies have been devoted to examine the dependence on time of the losses, others point towards the need of considering serial correlations [16,22]. Finally, the equidispersion property of the Poisson process may be inadequate in practice: for example, in the Cruz study fraud loss data [23] the empirical variance doubles the mean. This overdispersion characteristic of the OpRisk has been also documented by [3,30].

The contribution of this paper is three-fold. First, we consider an aggregate loss model where the losses occurrences are modeled by a Markov renewal process that allows for non-exponential and correlated inter-loss times as well as for overdispersed loss counts. Second, some measures of persistence related to the loss occurrence process which may be of interest for financial decision makers are derived. In particular, closed-form expressions for the transition probabilities between short and long inter-loss periods as for the mass probability function of long/short inter-loss times spells are calculated. Third, the aggregate model and quantities of interest as persistence coefficients and capital charges are estimated on a base of a real OpRisk dataset.

Markov renewal process and renewal processes have been a traditional tool for insurance risk modeling, see for example [20,31,45,69]. In this work we consider the Markovian arrival process (MAP), a type of Markov renewal process which is in turn a subtype of Batch Markovian arrival process (BMAP). The general BMAP [51] defines a versatile class of point process containing both renewal and non-renewal processes, as the Phase-type renewal processes and Markov modulated Poisson processes, respectively. In a BMAP, the times between events (losses, arrivals...) are dependent and phase-type (PH) distributed, and as will be described later, the mean and variance of the counting process do not coincide. All these properties have made the BMAP suitable for modeling a variety of real life contexts as queuing theory, reliability, teletraffic, and climatology, see [14,38,42,62,66,78], just to cite a few. In particular, the MAP has been also widely considered in the insurance literature, see [1,4,19,21,32,52,79]. In [64], a variant of the MAP, the Marked Markovian arrival process, which differentiates between multiple categories of losses, is considered.

A number of papers have considered statistical inference for different classes of BMAPs. For example, [5,11,28,66,74] and [78] investigate moments matching approaches. Other authors adopt a Bayesian viewpoint, as [29] or [60], where exact Gibbs samplers are proposed to obtain the posterior distribution of the model parameters. Finally, the EM algorithm is the method suggested for instance in [8,39] and more recently, in [54]. In this paper, given a real OpRisk dataset, we carry out estimation of the aggregate loss model by fitting separately the frequency and severities. In regards the frequency, we apply direct maximization of the likelihood for a sequence of consecutive inter-loss times in a MAP_2 , an approach that follows [10].

As previously commented, heavy tailed distributions constitute the common tool for modeling severities. In our example, the double-Pareto Lognormal distribution [63] provides reasonable estimates for both the body and tail of the empirical distribution. Once the occurrence process and the severities are fitted, we are able to estimate the loss aggregate distribution and both persistence and risk related measures as the distribution of short (long) inter-loss times spells, or the Value-at-Risk and Expected Shortfall. It should be noted that in spite of the large amount of aggregate loss models defined in the literature, papers dealing with their statistical estimation are very scarce. As examples, see [25], [75] and [2]. In the first, a simple distribution (Poisson or Negative Binomial) is fitted to the frequency data [75]; consider Bayesian estimation for a loss model under different combinations of distributions for the pair frequency/severity. Finally, in [2], a different model (the Coxian distribution) is estimated to inter-loss data. However, in all the previous works independent losses were assumed.

This paper is organized as follows. Sections 2.1 and 2.2 review the definition and properties of the MAP_2 , in particular those related to the inter-loss time distribution and the loss counting process. Section 2.3 describes the estimation algorithm for fitting the MAP_2 to a sequence of inter-loss times. In Section 3 we analyze in detail several aspects of the MAP_2 that remained unexplored, up to our knowledge. In particular, Section 3.1 considers the issue of the overdispersion, while

Section 3.2 is devoted to present novel measures of persistence in the loss occurrence process. While Sections 2 and 3 are focused on the loss occurrence process, Sections 4 and 5 deal with the aggregate loss model. Section 4 reviews basic theoretical aspects of loss aggregate models in the context of OpRisk and Section 5 represents the applied contribution of the paper. The inference of the MAP_2 allows for estimating the considered aggregate loss distribution, given a real OpRisk data set. Setting a specific type of heavy-tailed distribution for the severities (double Pareto Lognormal distribution, see [63]), then a Monte Carlo algorithm, for which the convergence is assessed, is used to simulate samples from the aggregate loss model. The differences obtained under a MAP_2 and, the commonly assumed Poisson process, in terms of the loss distribution and risk measures are object of discussion. Finally, Section 6 considers conclusions and future work.

2. A Markov Renewal loss occurrence process

The Markovian arrival process (MAP) constitutes a class of Markov renewal process that generalize in matrix form the Poisson process by allowing for dependent and non-exponentially distributed inter-loss times. In this paper we will concentrate on the stationary two-state MAP, noted MAP_2 , for two reasons. First, it has proven to be a versatile process in a number of different contexts [36,60,62,68]. Second, unlike higher order MAPs, it can be represented by a canonical, unique parametrization in terms of a small number of parameters, which are convenient properties from a statistical inference viewpoint. This section summarizes the main properties of the MAP_2 , with special emphasis on the inter-loss time distribution and the counting process.

2.1. Definition of the MAP_2

The MAP_2 is a doubly stochastic process defined by a Markov process with state space $S = \{1, 2\}$ and a counting process that changes according to the transitions of the Markov process. For a detailed description of the MAP_2 , see [43,44,51,59].

In the MAP_2 , the initial state $i_0 \in S$ is sampled from an initial probability vector $\theta = (\theta, 1 - \theta)$. At the end of a sojourn time which follows an exponential distribution with rate λ_i , two different type of transitions may occur. First, with probability $0 \leq p_{ij1} \leq 1$ a unique loss occurs and the MAP_2 goes to a state $j \in S$. On the other hand, with probability $0 \leq p_{ij0} \leq 1$, no loss occurs and the MAP_2 goes to a state j which is necessarily different from state i .

A stationary MAP_2 can be represented by $\{\lambda, P_0, P_1\}$, where $\lambda = (\lambda_1, \lambda_2)$, and P_0 and P_1 are 2×2 probability matrices with elements corresponding to transition probabilities p_{ij0} ($i \neq j$) and p_{ij1} , respectively. Instead of transition probability matrices, any MAP_2 can also be characterized in terms of the rate matrices $\{D_0, D_1\}$,

$$D_0 = \begin{pmatrix} -\lambda_1 & \lambda_1 p_{120} \\ \lambda_2 p_{210} & -\lambda_2 \end{pmatrix}, \quad D_1 = \begin{pmatrix} \lambda_1 p_{111} & \lambda_1 (1 - p_{120} - p_{111}) \\ \lambda_2 p_{211} & \lambda_2 (1 - p_{210} - p_{211}) \end{pmatrix}. \tag{2}$$

The matrix D_0 is assumed to be stable, and as a consequence, it is nonsingular and the sojourn times are finite with probability 1. The definition of D_0 and D_1 implies that $D = D_0 + D_1$ is the infinitesimal generator of the underlying Markov process, with stationary probability vector $\pi = (\pi, 1 - \pi)$, computed as $\pi D = \mathbf{0}$. When $D_0 = -\lambda$ and $D_1 = \lambda$ the MAP_2 is reduced to the Poisson process with rate λ .

The MAP_2 is a Markov renewal process. Indeed, if Y_n denotes the state of the MAP_2 at the time of the n th loss, and let T_n denote the time elapsing such loss and the previous one, then $\{Y_{n-1}, T_n\}_{n=1}^\infty$ is a Markov renewal process. In this case, the associated 2×2 semi-Markov kernel, whose (i, j) -th element is defined as

$$Q(i, j, t) = P(Y_{n+1} = j, T_{n+1} \leq t \mid Y_n = i),$$

is given by

$$Q(t) = (I - e^{D_0 t})(I - P_0)^{-1} P_1, \tag{3}$$

see [13].

From Markov renewal theory, $\{Y_n\}_{n=1}^\infty$ is a Markov chain whose transition matrix P^* is given by

$$P^* = (I - P_0)^{-1} P_1, \tag{4}$$

with stationary distribution ϕ as

$$\phi = (\pi D_1 \mathbf{e})^{-1} \pi D_1, \tag{5}$$

see [59]. It is a straightforward computation to prove that (4) can be also rewritten as $P^* = (-D_0)^{-1} D_1$.

The expression (2) for the MAP_2 in terms of 6 parameters is known to be overparameterized, [59]. However, [5] provide a unique, canonical representation for the MAP_2 in terms of just four parameters. Such canonical representation is the one we are using in this paper. Let $-1 \leq \gamma < 1$ be one of the two eigenvalues of the transition matrix P^* (since P^* is stochastic, then necessarily the other eigenvalue is equal to 1), [5]. If $\gamma > 0$, then the canonical form of the MAP_2 is given by

$$D_0 = \begin{pmatrix} x & y \\ 0 & u \end{pmatrix}, \quad D_1 = \begin{pmatrix} -x - y & 0 \\ v & -u - v \end{pmatrix}. \tag{6}$$

On the other hand, for those MAP_2 s such that $\gamma \leq 0$, then their canonical form is

$$D_0 = \begin{pmatrix} x & y \\ 0 & u \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0 & -x - y \\ -u - v & v \end{pmatrix}. \tag{7}$$

In the previous representations (6) and (7), x and u are less than 0 while the rest of elements are larger or equal to 0.

2.2. The inter-loss times distribution and the loss counting process

Special attention deserves the analysis of the T , the random variable representing the time between two sequential losses of a MAP_2 (or the inter-loss time distribution) in its stationary version. It is known that T is phase-type (PH) distributed with representation $\{\phi, D_0\}$ [13], and therefore the cumulative distribution function is

$$F_T(t) = 1 - \phi e^{D_0 t} \mathbf{e}, \quad \text{for } t \geq 0. \tag{8}$$

It should be pointed out here that the phase-type distribution generalizes both the exponential distribution as well as the Coxian distribution (used by [2] in the context of aggregate loss distributions).

The moments of T can be computed as

$$m_n = E(T^n) = n! \phi (-D_0)^{-n} \mathbf{e}, \tag{9}$$

where $\phi = (\phi, 1 - \phi)$ is the probability distribution satisfying $\phi P^* = \phi$, and \mathbf{e} is a vector with elements equal to 1. In particular, the loss rate is computed as

$$\lambda^* = 1/m_1 = \pi D_1 \mathbf{e}. \tag{10}$$

If the focus is not on the marginal properties of inter-loss times but on their joint structure then, the likelihood function for a MAP_2 trace of inter-loss times is given by

$$f(t_1, t_2, \dots, t_n | D_0, D_1) = \phi e^{D_0 t_1} D_1 e^{D_0 t_2} D_1 \dots e^{D_0 t_n} D_1 \mathbf{e}, \tag{11}$$

see for example [8,67].

As commented in Section 1, the MAP_2 allows for dependent inter-loss times. In particular, the consecutive times in the stationary version of the MAP_2 are correlated according to the correlation coefficient given by

$$\rho = \gamma \frac{\frac{m_2}{2} - m_1^2}{m_2 - m_1^2}, \tag{12}$$

where, as commented in the previous section, γ denotes the eigenvalue of the transition matrix P^* that satisfies $-1 \leq \gamma < 1$.

Next we shall briefly describe the counting process. Let $N_t(\tau)$ represent the number of losses in the interval $[t, t + \tau]$. Then, the number of losses observed in the interval $(0, \tau]$ will be given by $N_0(\tau)$. For $n \in \mathbb{N}$ and $\tau \geq 0$, let $P(n, \tau)$ denote the 2×2 matrix whose (i, j) th element is

$$P_{ij}(n, \tau) = P(N_0(\tau) = n, J(\tau) = j | N_0(0) = 0, J(0) = i),$$

for $1 \leq i, j \leq 2$. From the previous definition it is clear that

$$P(N_0(\tau) = n | N_0(0) = 0) = \pi P(n, \tau) \mathbf{e}. \tag{13}$$

The values of the matrices $P(n, \tau)$ cannot be computed in closed-form. Their numerical computation is based on the *uniformization method* addressed in [50] and summarized for the case of the MAP_2 in [66].

2.3. Statistical estimation of the MAP_2

A number of papers has considered estimation for different classes of $BMAP$ s, the method of moments and the EM algorithm being the most studied approaches, see [5,8,11,54,66,78]. Given a real trace of inter-loss times from the OpRisk context, we explore in this paper the estimation of the MAP_2 via a direct maximization of the likelihood function in the same spirit as in [10]. In particular, given a sequence of inter-loss times $\mathbf{t} = (t_1, t_2, \dots, t_n)$, we formulate the ML estimation problem from the likelihood expression (11) as

$$(P1) \left\{ \begin{array}{l} \max \quad \phi e^{D_0 t_1} D_1 e^{D_0 t_2} D_1 \dots e^{D_0 t_n} D_1 \mathbf{e} \\ \text{s.t.} \quad D_0 = \begin{pmatrix} x & y \\ 0 & u \end{pmatrix}, \\ \quad \quad D_1 = \begin{pmatrix} -x - y & 0 \\ v & -u - v \end{pmatrix}, \\ \quad \quad x, u, x + y, u + v \leq 0, \\ \quad \quad y, v \geq 0, \\ \quad \quad \phi (-D_0)^{-1} D_1 = \phi. \end{array} \right.$$

Note that (P1) is stated in terms of the first canonical form of the MAP_2 , see (6). Regarding the second canonical form, we formulate (P2) as (P1), substituting the expressions for D_0 and D_1 by (7). In practice, given $\mathbf{t} = (t_1, t_2, \dots, t_n)$, both problems (P1) and (P2) are solved and then the solution that maximizes the likelihood is kept.

Some computational aspects are detailed next. In order to solve (P1) (or (P2)), we used the MATLAB[©] routine `fmincon` (with default options), where the starting solution was selected as follows. It is known that the MAP_2 is completely characterized by the first three inter-loss times moments m_1, m_2, m_3 as in (9), and by the correlation coefficient ρ as in (12), see [5]. Then, a moments matching problem can be defined as

$$(P0) \begin{cases} \min & \delta(x, y, u, v) \\ \text{s.t.} & y, v \geq 0, \\ & x, u \leq 0, \\ & x + y \leq 0, \\ & u + v \leq 0, \end{cases}$$

where $\delta(x, y, u, v)$ is the objective function that measures the distance between the theoretical and empirical moments ($\bar{m}_1, \bar{m}_2, \bar{m}_3$ and $\bar{\rho}$):

$$\delta(x, y, u, v) = \{\rho(x, y, u, v) - \bar{\rho}\}^2 + \left(\frac{m_1(x, y, u, v) - \bar{m}_1}{\bar{m}_1}\right)^2 + \left(\frac{m_2(x, y, u, v) - \bar{m}_2}{\bar{m}_2}\right)^2 + \left(\frac{m_3(x, y, u, v) - \bar{m}_3}{\bar{m}_3}\right)^2.$$

To solve the multimodal Problem (P0), the routine `fmincon` was also used. A multistart was then executed with 100 randomly chosen starting points and found to yield satisfactory results. In practice, it turned out that the solution to the moments matching method is a good starting point for the ML problem. Indeed, other initial values could have been chosen, for example random starting MAP_2 s. In order to analyze the effect of the different starting points on the performance of the direct maximization likelihood approach, a total of one thousand random MAP_2 s were estimated as the solution of problem (P1), where the starting values were (1) randomly generated versus (2) the moments matching estimates. It was found that in the 61.53% of times, the objective function (the likelihood function) was larger using a moments method estimate. Finally, we discuss about the computational times characterizing the inference approach. Solving problem (P0) is fast in practice since it does not depend on samples sizes (the input is not the sequence of inter-loss times but the four empirical moments $\bar{m}_1, \bar{m}_2, \bar{m}_3$ and $\bar{\rho}$). For a prototype code written in MATLAB[©] the running time was equal to 6.3 seconds, when run on Intel Core i5 at 3.2 GHz and 16 GB of DDR3 RAM. Solving (P1) is slower than solving (P0), since it does depend on samples sizes. Using the solution of (P0) as the starting solution and a sample of size equal to $n = 100$, the running time was 18.2 s. If instead, a sample of $n = 500$ inter-loss times is considered, the running time increased to 30.7 s. The undertaken approach is in any case faster than other approaches. If the Bayesian method in [60] is implemented, then for a sample of 500 inter-loss times the running time is about 19 s for a total of 1000 iterations. Taking into account that convergence is achieved after 50000 iterations, the Bayesian method turns out significantly slower. Recently, [78] compared the EM algorithm with a moments matching method in the context of Batch Markovian arrival processes, resulting the last one considerably faster.

The estimation of the MAP_2 is not the main goal of this paper, and therefore we refer the reader to [10], where the technical details regarding the previous approach as well as several numerical illustrations are discussed more in depth.

3. Overdispersion and persistence

3.1. Overdispersion under the MAP_2

As commented in Section 1, there are works pointing towards the overdispersion of the loss counts in the OpRisk context, which implies that the mean of the cumulated losses is considerably smaller than its variance, see [3,23,30]. Unlike the Poisson process, the MAP_2 is able to model such overdispersion phenomenon, as we illustrate next.

Throughout this work, $N(\tau)$ shall denote the stationary counting process in an interval of length τ , that is $N(\tau) = \lim_{t \rightarrow \infty} N_t(\tau)$ where $N_t(\tau)$ counts for the number of losses in the interval $[t, t + \tau]$. Then, the mean number of losses is defined in terms of the loss rate (10) as

$$E[N(\tau)] = \lambda^* \tau, \tag{14}$$

and the variance of that count is given by

$$V[N(\tau)] = (1 + 2\lambda^*)E[N(\tau)] - 2\pi D_1(\mathbf{e}\boldsymbol{\pi} + D)^{-1}D_1\mathbf{e}\boldsymbol{\tau} - 2\pi D_1(I - e^{D\boldsymbol{\tau}})(\mathbf{e}\boldsymbol{\pi} + D)^{-2}D_1\mathbf{e}, \tag{15}$$

see [48] or [28].

It is also known that the counting process in a general m -state MAP has dependent increments. In particular, for the MAP_2 the covariance of the counting process between intervals of length τ is given by

$$C(\tau) = \boldsymbol{\pi}(M_1(\tau))^2\mathbf{e} - (\boldsymbol{\pi}M_1(\tau)\mathbf{e})(\boldsymbol{\pi}e^{D\boldsymbol{\tau}}M_1(\tau)\mathbf{e}),$$

where $M_1(\tau)$ represents the first moment matrix of the counts during an interval $(0, \tau]$. Asymptotically, [48] proved that

$$M_1(\tau) = E[N(\tau)]\mathbf{e}\boldsymbol{\pi} + (\mathbf{e}\boldsymbol{\pi} - D)^{-1}D_1\mathbf{e}\boldsymbol{\pi} + \mathbf{e}(\boldsymbol{\pi}D_1(\mathbf{e}\boldsymbol{\pi} - D)^{-1}) - 2\frac{E[N(\tau)]}{\tau}\mathbf{e}\boldsymbol{\pi}.$$

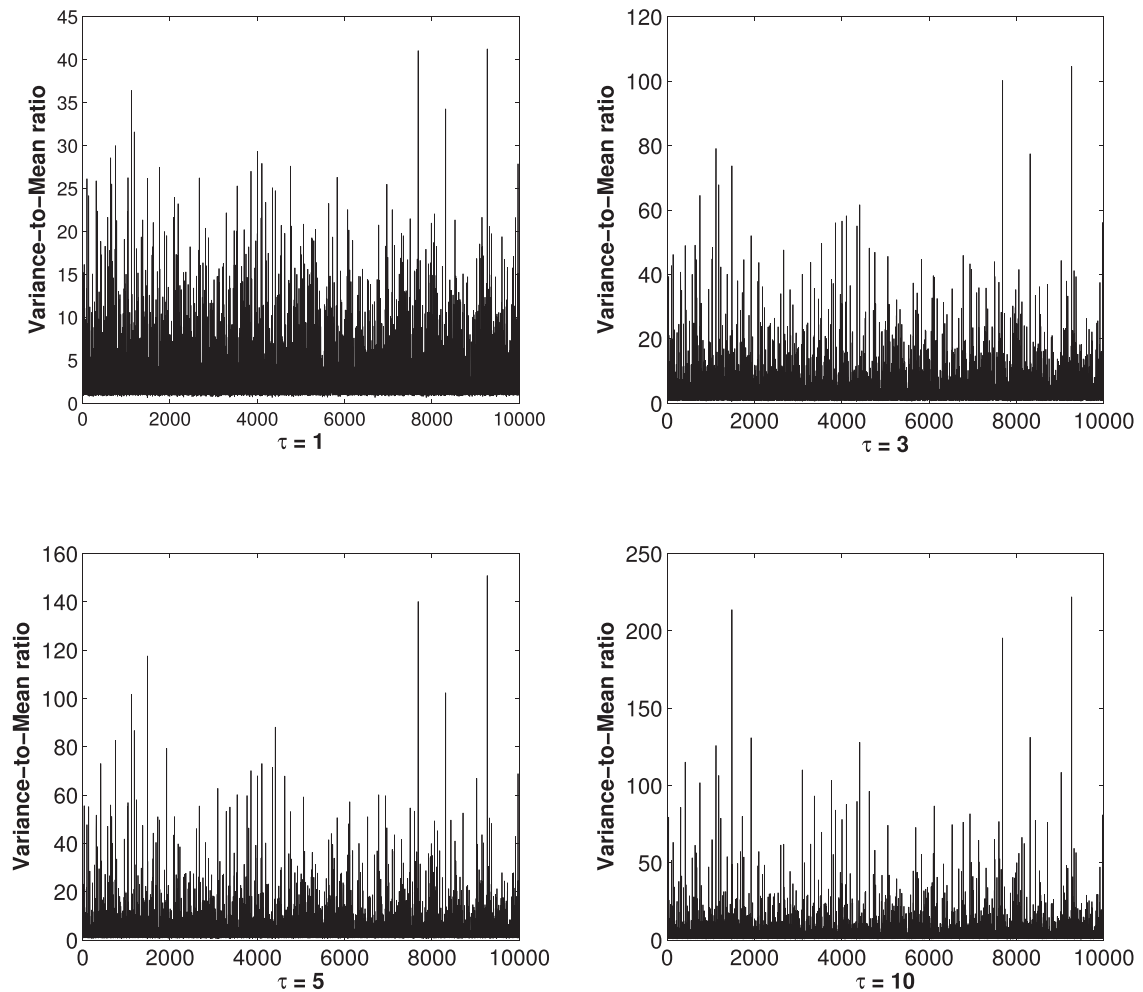


Fig. 1. Variance-to-Mean ratios for a total of 10000 simulated MAP_2 s for different values of τ .

A measure of the potential overdispersion of the loss counting process is the Variance-to-Mean ratio [41], defined as

$$VtM(N(\tau)) = \frac{V[N(\tau)]}{E[N(\tau)]}. \tag{16}$$

[49] recently define the Index of Dispersion of Counts (IDC) in the same way as in (16). In order to empirically study the range of values of (16), a total of 10000 random set of parameters $\{x, y, u, v\}$ as in (6),(7) characterizing a MAP_2 were simulated. Then, the Variance-to-Mean ratios as in (16) were calculated for an assortment of τ values ($\tau \in \{1, 3, 5, 10\}$). Fig. 1 depicts the obtained results. From the figure it is clear that there are multiple configurations of parameters $\{x, y, u, v\}$ under which $V[N(\tau)]$ is considerably higher than $E[N(\tau)]$, a phenomenon that intensifies as τ increases. This is better observed in Fig. 2, which shows the evolution of the Variance-to-Mean ratio with the value of τ , for four different representations of MAP_2 s. It is interesting to note that in the simulation, only a 6.74% of MAP_2 s led to a Variance-to-Mean ratio less than 1. These preliminary results point out the suitability of the MAP_2 for modeling overdispersed data. Further theoretical research needs to be carried out in this line but since it is not the scope of the paper, we leave it as an open task.

3.2. Persistence measures

Next we shall study some persistence properties concerning the loss counting process of the MAP_2 . There are works in the literature which consider the modeling of persistence since this is an inherent phenomenon in many financial and insurance contexts, see for example [73,76] and the references therein. In general, in the OpRisk (and any financial) context it might be of interest to infer the probability that the next loss occurs in a *short* (or *long*) time given that the previous loss was *recent* (or *far away*). Also, decision makers could benefit from the knowledge of estimates for consecutive long periods without losses (long inter-loss times spells) or in contrast, short inter-loss times spells which might result in a higher capital charge. We next define these event and provide formulae for their calculation.

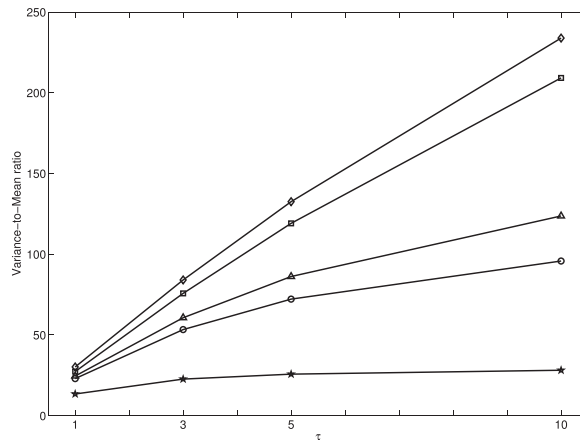


Fig. 2. Variance-to-Mean ratios at time instants $\tau = 1, 3, 5, 10$, for 4 randomly selected MAP_2 representations.

Let s denote a fixed time. Recall from Section 2 that T_n is the time elapsing from the occurrence of the $(n - 1)$ th loss until the next (n) -th loss. Define the following persistence measures as:

$$p_{01}(s) = P(T_{n+1} > s \mid T_n < s),$$

$$p_{11}(s) = P(T_{n+1} > s \mid T_n > s).$$

As previously commented, financial decision makers could be interested in estimating $1 - p_{01}(s)$ for *small* values of s , or equivalently, the probability of the occurrence of a loss in a *short* period of time given that the previous loss occurred *recently*. Also, it would be advantageous to infer $p_{11}(s)$ for *large* values of s , that is, the probability of a *long* time up to the next loss given that the previous loss occurred *long ago*. In practice, s can be chosen as a small or high percentile of a given historical trace of consecutive inter-loss times. Next result provides the expressions for the previous persistence measures.

Proposition 1. Let a MAP_2 be represented by $\{D_0, D_1\}$. Let P^* denote the transition matrix between states where losses occur and ϕ its stationary distribution. Then,

$$p_{01}(s) = \frac{\phi(I - e^{D_0 s})P^*e^{D_0 s}P^*e}{1 - \phi e^{D_0 s}e}, \tag{17}$$

$$p_{11}(s) = \frac{\phi e^{D_0 s}P^*e^{D_0 s}P^*e}{\phi e^{D_0 s}e}, \tag{18}$$

for $s > 0$.

See Appendix A for a proof.

Now, assume that the interest is to predict the probability of a sequence of *short (long)* inter-loss times spells. As before, consider a fixed time s . Then, define two discrete random variables S and L (standing for *Short* and *Long*, respectively) as

$$S = n \underline{\underline{d}} (T_1 < s) \cap (T_2 < s) \cap, \dots, \cap (T_n < s) \cap (T_{n+1} > s), \tag{19}$$

$$L = n \underline{\underline{d}} (T_1 > s) \cap (T_2 > s) \cap, \dots, \cap (T_n > s) \cap (T_{n+1} < s). \tag{20}$$

Defined as in (19),(20), then for a *small* value of s , S counts for the number of consecutive short inter-loss spells. On the other hand, if s is *large* then, L represents the number of consecutive long inter-loss spells. The following result expresses the probabilities of (19) and (20) in matrix form.

Proposition 2. Let a MAP_2 be represented by $\{D_0, D_1\}$. Let P^* denote the transition matrix between states where losses occur and ϕ its stationary distribution. Then, for a fixed value $s > 0$

$$P(S = n) = \begin{cases} \phi[(I - e^{D_0 s})P^*]^n e^{D_0 s}P^*e, & \text{if } n \geq 1 \\ \phi e^{D_0 s}e, & \text{if } n = 0. \end{cases} \tag{21}$$

$$P(L = n) = \begin{cases} \phi(e^{D_0 s}P^*)^n (I - e^{D_0 s})P^*e, & \text{if } n \geq 1 \\ 1 - \phi e^{D_0 s}e, & \text{if } n = 0. \end{cases} \tag{22}$$

See Appendix B for a proof.
 A remark is made at this point.

Remark 1. Propositions 1 and 2 are also true for higher-order MAPs. This is because in the matrix expressions for the different persistence measures only D_0 is involved (and not D_1 , or D_k for $k \geq 2$ for higher-order MAPs). Also, the definitions of P^* and ϕ are common for MAPs of all orders, see [13]. In the case of higher-order MAPs the proofs shall be identical but changing the counter of the summation from $i, j, k \in \{1, 2\}$ by $i, j, k \in \{1, 2, \dots, m\}$, being m the order of the process.

To conclude this section, we present several numerical results in order to illustrate some of the patterns that (17)-(18) and (21)-(22) may follow. This issue is of paramount importance to determine the versatility of the MAP_2 for modeling real-life situations. In this respect, the presented results are very preliminar and further research should be carried out in future work.

In order to illustrate some of the patterns that (17),(18) and (21),(22) are able to model, four different MAP_2 representations were chosen:

R1.

$$D_0 = \begin{pmatrix} -1.1272 & 0.0055 \\ 0 & -42.4417 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 1.1217 & 0 \\ 0.2173 & 42.2244 \end{pmatrix}.$$

R2.

$$D_0 = \begin{pmatrix} -1.4373 & 0.0498 \\ 0 & -14.2706 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 1.3875 & 0 \\ 0.5283 & 13.7423 \end{pmatrix}.$$

R3.

$$D_0 = \begin{pmatrix} -0.6830 & 0.0026 \\ 0 & -34.6904 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0 & 0.6804 \\ 34.5586 & 0.1318 \end{pmatrix}.$$

R4.

$$D_0 = \begin{pmatrix} -0.9751 & 0.5933 \\ 0 & -46.6547 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0 & 0.3818 \\ 10.7740 & 35.8806 \end{pmatrix}.$$

The previous selection of parameters was made among a total of one million simulated MAP_2 s so that the four sets of parameters lead to different forms for expressions (17), (18), (21) and (22).

For each representation, $1 - p_{01}(s)$ and $p_{11}(s)$ are computed for values of s equal to the low percentiles (5-th,6-th,...,49-th and 50-th) in the first case, and high percentiles (50th,51th,...,94th and 95th) in the second case. The results are shown in Figs. 3 and 4. Different models are obtained: in all the considered cases, the values of $1 - p_{01}(s)$ increase with s although the increase pattern differs between representations. On the contrary, there are cases where $p_{11}(s)$ is not monotonic.

Figs. 5 and 6 depict the probabilities of the random variables that count for the number of short and long inter-loss times. In particular, the value of s has been chosen as the 20th and 80th percentiles of the inter-loss times distributions, respectively for S and L . In the four cases, the probabilities mass functions are found to be unimodal at 0 and rapidly decrease towards 0, a pattern that was found in the total of the one million simulated MAP_2 s.

4. Modeling the OpRisk

According to the European Union Solvency II Directive for insurers [35], the OpRisk is defined as “the risk of a change in value caused by the fact that actual losses, incurred for inadequate or failed internal processes, people and systems, or from external events, differ from the expected losses”. Because of its high expected impact to the banking industry, since the 9/11 attacks and tradings at Société Générale, AIB and National Australia Bank, researchers have been studying approaches for properly modeling the OpRisk. The Basel II framework forces financial institutions to set aside a capital charge against unexpected losses derived from OpRisk. Three different methodologies have been proposed in the literature for OpRisk quantification: the Basic Indicator Approach, the Alternative Standardized Approach and, the most sophisticated one (in terms of statistical modeling), the Advanced Measurement Approach (AMA), where the Loss Distribution Approach (LDA) is included in. In the LDA the estimation of the annual loss distribution Z as in (1) is achieved by modeling separately the distributions of the number of losses per year (N) and the severities (X_i), under the assumption of independence between them. It is common that the Poisson process or the Negative Binomial distribution are chosen to model the frequency, while some type of heavy-tailed distribution is selected to fit the severities, [6,18]. For a thorough treatment of the OpRisk modeling, we refer the reader to [6,56,57,70,72].

The cumulative distribution function of Z is calculated as

$$F_Z(x) = P(Z \leq x) = \sum_{n=0}^{\infty} P(N = n)F_n^*(x),$$

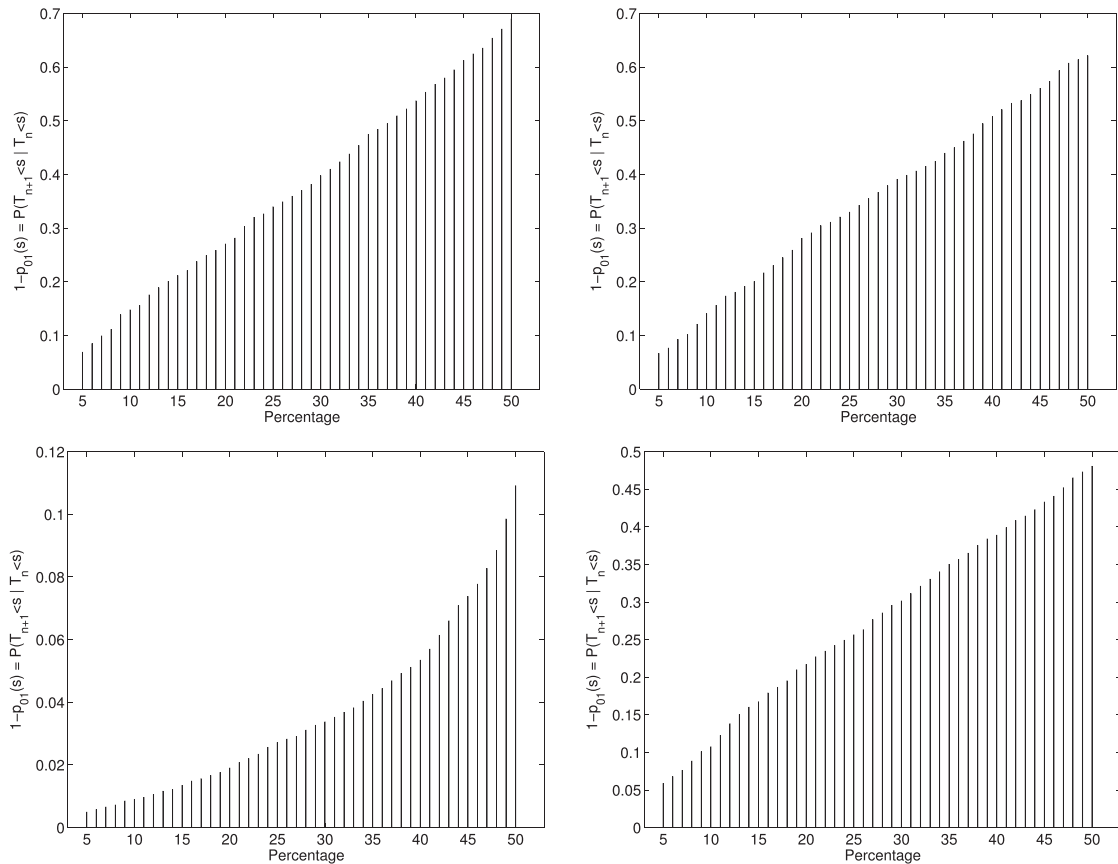


Fig. 3. $P(T_{n+1} < s | T_n < s)$ for the four MAP_2 representations: R1 (top left panel), R2 (top right panel), R3 (bottom left panel) and R4 (bottom right panel). The value of s is determined by the percentiles of percentages given by the X-axis.

where $F_n^*(x) = P(X_1 + \dots + X_n \leq x)$ is the n -th convolution of the severities' distribution. Although a closed-form expression for the loss aggregate distribution is usually hard to be obtained, its first and second moments can be computed in terms of those of the severity and frequency distributions as

$$E(Z) = E(N)E(X_1),$$

$$V(Z) = E(N)V(X_1) + V(N)E(X_1)^2.$$

Closed-form expressions for higher order moments can be found, for example in [71]. Under (1), the capital charge is obtained as the Value-at-Risk (VaR),

$$VaR_p(Z) = F_Z^{-1}(p), \tag{23}$$

where F_Z^{-1} is the inverse cumulative distribution function of the total annual loss, and p denotes the risk tolerance which is usually chosen close to 1. Another measure of interest in the OpRisk modeling is the expected shortfall (equivalently, conditional VaR), which unlike the VaR, satisfies the property of coherence. It is computed as

$$E(Z | Z \geq VaR_p(Z)). \tag{24}$$

If the distribution of Z has an infinite mean, then clearly the expected shortfall does not exist.

In order to approximate (1) a battery of numerical methods, either based on simulations, inversion of characteristic functions, or recursions, have been proposed in the literature. For an exhaustive review of such methods, see [34,56,71,72], or Ch. 9 in [40].

5. Numerical application

In this section we illustrate an application of the MAP_2 in the context of the OpRisk. In particular, given a real OpRisk database, described in Section 5.1, the MAP_2 is fitted to the times between consecutive losses in Section 5.2. Descriptors of interest associated to the distributions of the inter-loss times and the number of yearly losses, as well as estimations for the

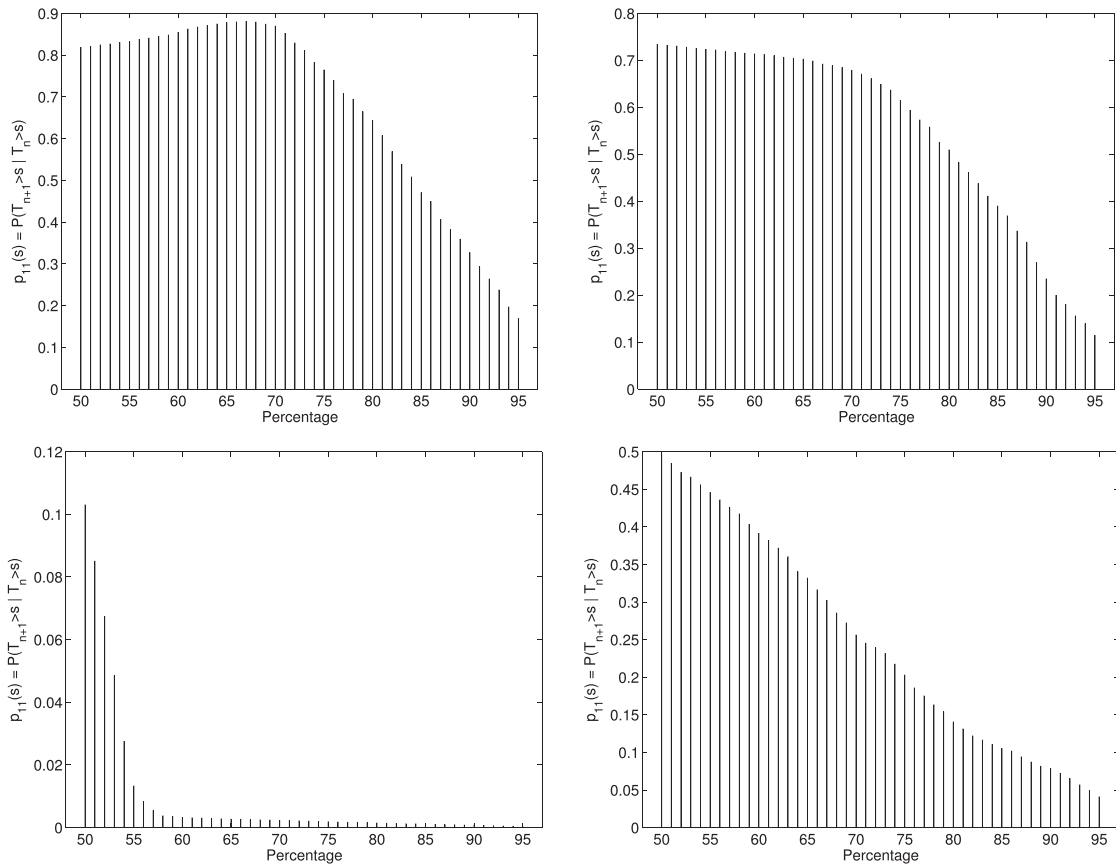


Fig. 4. $P(T_{n+1} > s | T_n > s)$ for the four MAP_2 representations: R1 (top left panel), R2 (top right panel), R3 (bottom left panel) and R4 (bottom right panel). The value of s is determined by the percentiles of percentages given by the X-axis.

persistence measures described in Section 3 are obtained. In Section 5.3 the severities' distribution are modeled through the heavy-tailed, double Pareto Lognormal distribution. As will be seen, the combination of the MAP_2 for the frequency with the heavy-tailed model for the loss severity distribution, will enable us to give an estimate of the loss aggregate model and the capital charge in Section 5.4.

5.1. Data description

The real OpRisk data set considered in this paper consists of 225 records of loss occurrences and associated severities in a Spanish financial institution. The sample, which corresponds to a unique risk cell (retail banking Basel II business line), was obtained from 4929 consecutive days (13.5 years) in the period between 30/12/1993 and 29/06/2007. In a total of 225 out of the 4929 days, a single loss occurred which was recorded together with the corresponding severity. Some descriptors of the dataset are as follows. The average, median, variation coefficient and maximum value of the inter-loss times are 22.0045, 8, 2.827 and 667 days. Also, the empirical correlation coefficient is non-negligible and given by 0.3546. For the severities, the average, median, variation coefficient and maximum are 151710, 34318, 2.0973 and 2289060. Note that, from the empirical values both samples are right-skewed with a tail longer than that of an exponential distribution. This can also be deduced from Fig. 7, which depicts the empirical quantiles in comparison to those of the fitted (via MLE) exponential distributions. Note how the larger empirical quantiles are far from the fitted ones.

5.2. The estimated frequency distribution

In this section we describe how the database inter-loss times are modeled by the MAP_2 . The estimated (joint) distribution of the times, the modeling of the counting process $N(\tau)$ and of the persistence measures described in Section 3 will be analyzed.

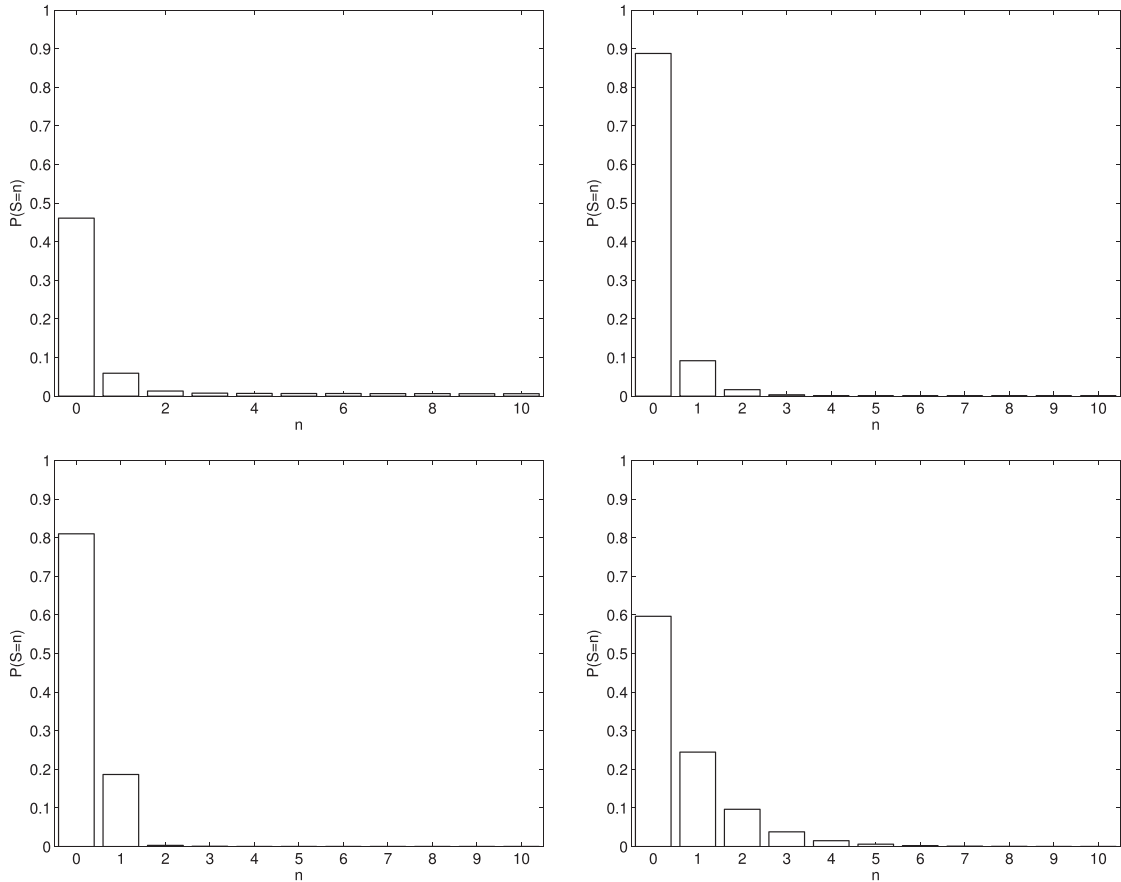


Fig. 5. $P(S = n)$ for the four MAP_2 representations: R1 (top left panel), R2 (top right panel), R3 (bottom left panel) and R4 (bottom right panel).

The distribution of the inter-loss times

The methodology from Section 2.3 was applied to fit a MAP_2 to the sample of inter-loss times, and the obtained results were as follows. First, the estimated process was found as

$$\hat{D}_0 = \begin{pmatrix} -0.0063 & 0.0011 \\ 0 & -0.1036 \end{pmatrix}, \quad \hat{D}_1 = \begin{pmatrix} 0.0052 & 0 \\ 0.0016 & 0.1020 \end{pmatrix}. \tag{25}$$

Very closed values were obtained if instead the Bayesian algorithm by [60] is considered.

The mean, median, coefficient of variation and correlation coefficient under the estimated model were 22.0047, 7.52, 2.8205 and 0.3545, all close to the empirical values commented in Section 5.1. The estimated MAP_2 (25) defines a theoretical cdf as in (8) which is plotted in dashed line in Fig. 8 together with the empirical cdf of the data (solid line). The performance of the exponential distribution is also shown in dotted line. The outperformance of the PH-distribution (which defines the inter-loss times distribution in a MAP_2) over the classic exponential distribution is clear from this figure.

The distribution of the number of losses per year

We turn out to the estimation of the counting process. As commented in Section 5.1 losses were recorded in a period of length between 13 and 14 years. It was found that the average number of losses per year is equal to 16.6154, and the variance is 342.9231. In this respect, the Poisson process presents again a poor performance, since it would notably underestimate the sample variance. Given the estimated MAP_2 (25), the expected number of losses as well as the variance were computed according to (14)-(15), and found as 16.5874 and 240.0192.

Since the modeling of the frequency in the OpRisk context usually refers to the one year horizon, the numerical algorithm described in [66] was applied to obtain the estimates of the probabilities as in (13), for $\tau = 365$. Hence, from now on, N will denote $N(365)$. Fig. 9 depicts the obtained mass function of N , the annual number of losses, which will be used in Section 5.4 for approximating the loss aggregate distribution. Note that the obtained probabilities cannot be compared against their empirical versions because of the small sample size (a sample of 13 years). We finally would like to point out the differences between the adjusted mass function and that of a Poisson distribution. The first is unimodal around 0, but

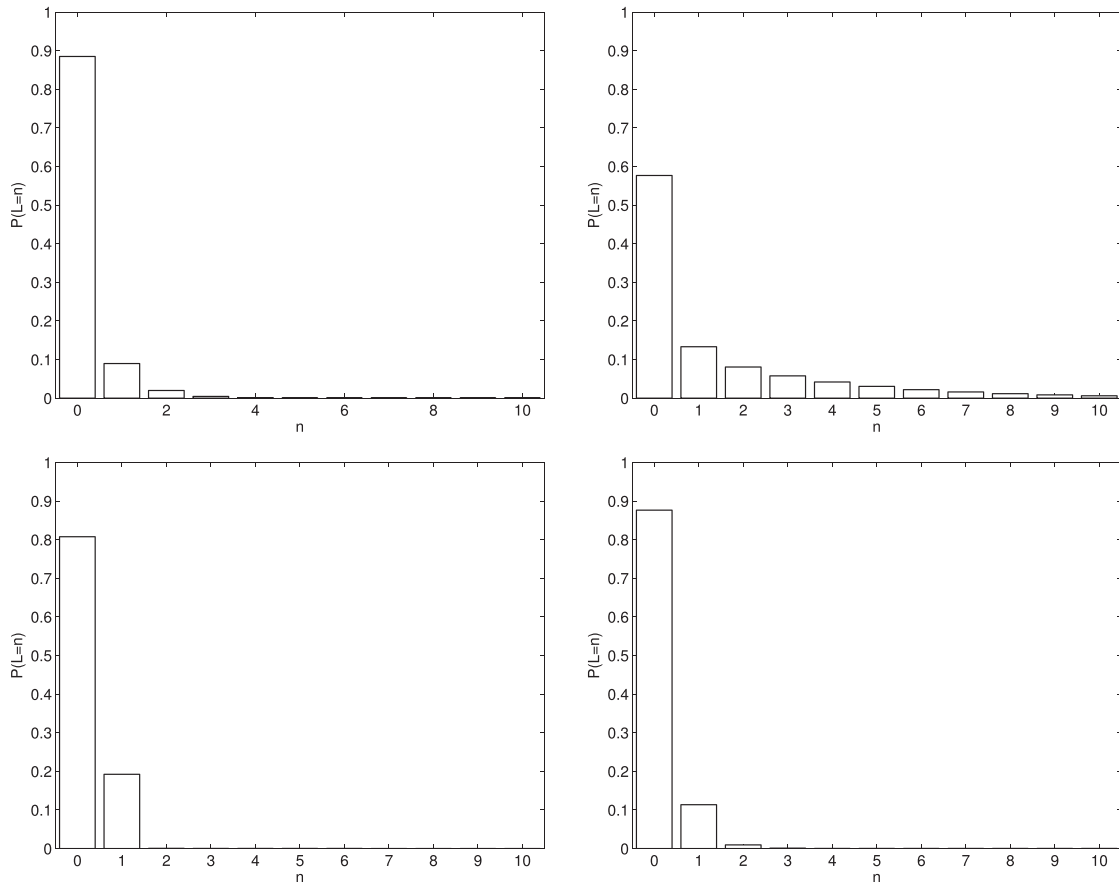


Fig. 6. $P(L = n)$ for the four MAP_2 representations: R1 (top left panel), R2 (top right panel), R3 (bottom left panel) and R4 (bottom right panel).

it is non-negligible for high losses; in contrast, the Poisson mass function of rate 16.6154 (equal to the average number of annual losses) concentrates around its mean with a much smaller variance.

Persistence measures

Here we obtain the estimates of (17),(18) and (21),(22) for the real inter-loss times, according to the MLE estimates (25).

Consider first the estimation of the conditional probability of a next *short (long)* inter-loss time assuming that the previous inter-loss time was *short (long)*. It is important to remark that due to the small sample sizes, the empirical values of the transition probabilities $1 - p_{01}(s)$ were not calculated for values of s less than the 30th percentile of the inter-loss times distribution. In analogous way, $p_{11}(s)$ was not estimated from the sample for values of s larger than the times distribution's 70th percentile. As a result, the empirical values of $1 - p_{01}(s)$ were estimated for s equal to the 30th, 35th,...,50th percentiles. These percentiles define for each case the length of the *short* inter-loss time. Similarly, the empirical probabilities $p_{11}(s)$ were estimated for s ranging between the 50th, 55th,...,70th percentiles (defining the *long* times). The empirical values (represented with a triangle) and the estimated values (represented by a circle) are depicted by Fig. 10. In the figure, the white bars refer to the $1 - p_{01}(s)$ values while the black bars concern the probabilities $p_{11}(s)$. Consider for example, the 30-th percentile of the dataset (found as $s = 3$). Then, the empirical value of $P(T_{n+1} < 3 | T_n < 3)$ was 0.3731 (that is, the 37.31% of inter-loss times satisfying $T < 3$ were followed by times that also satisfied the condition). This value is represented by the first white bar ending in a triangle. In this case, the fit given by the MAP_2 (equal to 0.262 and represented by the next white bar ending in a circle), clearly underestimated the empirical value. However, there are examples where the MAP_2 provides good estimates. For instance, if s is chosen as the 60th percentile (found as $s = 11$), then the empirical probability $P(T_{n+1} > 11 | T_n > 11)$ was 0.4787, and the MAP_2 provided an estimated value equal to 0.4340.

Consider next the estimation of the probabilities of short and long spells, as in (21)-(22). As in the previous case, $s = 3$ is considered for the study of S (sequence of *short* spells), and $s = 11$ for that of L (sequence of *long* spells). Fig. 11 depicts the estimated probabilities for values of $n = 0, 1, \dots, 10$. For both examples, the functions are unimodal at 0 but with non-negligible values for the cases $n = 1, 2$. The empirical values for the probabilities $P(S = 0)$, $P(S = 1)$ and $P(S = 2)$ were 0.7366, 0.0804 and 0.0670. The estimated values, shown by the first three bars in the left panel of the figure, where 0.7535,

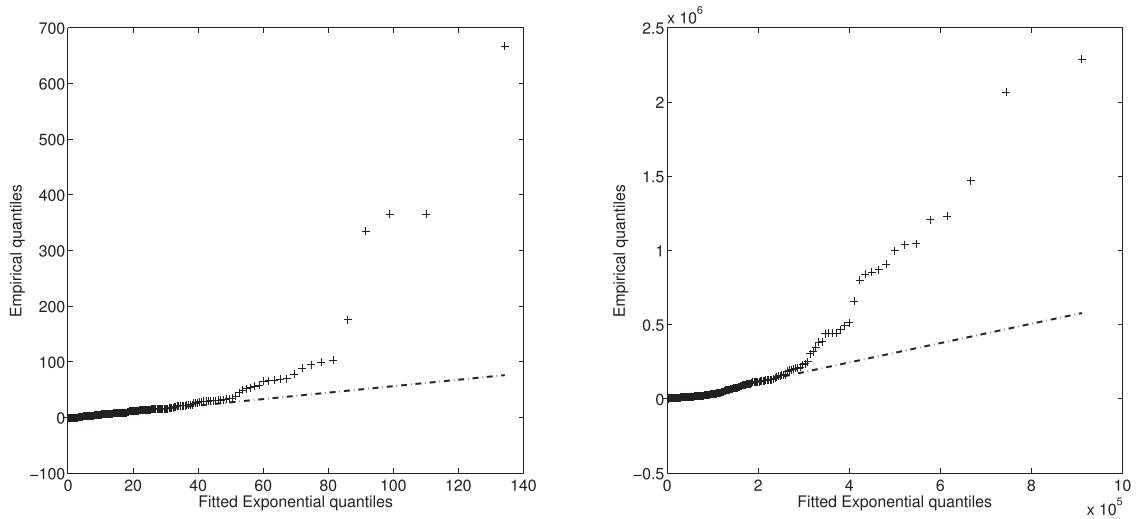


Fig. 7. QQ-plots of the empirical inter-loss times (left panel) and severities (right panel) quantiles versus those of the fitted exponential distributions.

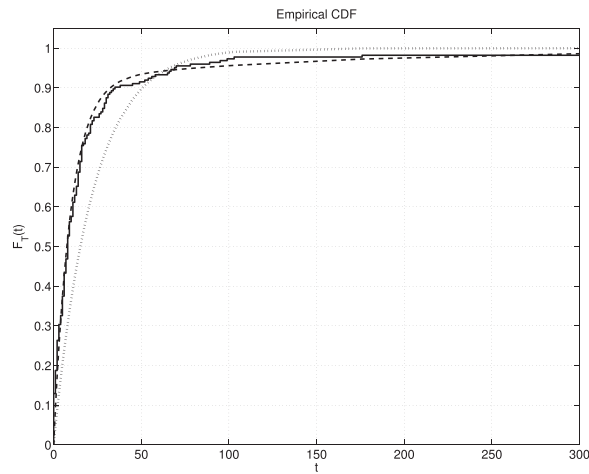


Fig. 8. Inter-loss times cdf: empirical (solid line) versus estimated by the MAP_2 (dashed line) and, by the exponential distribution (dotted line).

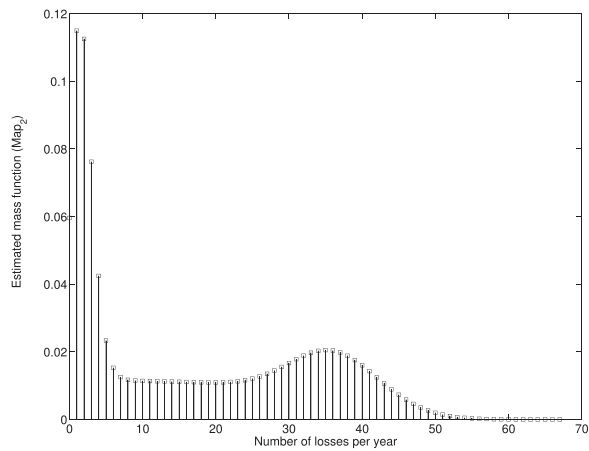


Fig. 9. Number of losses per year. Estimated mass function under the MAP_2 .

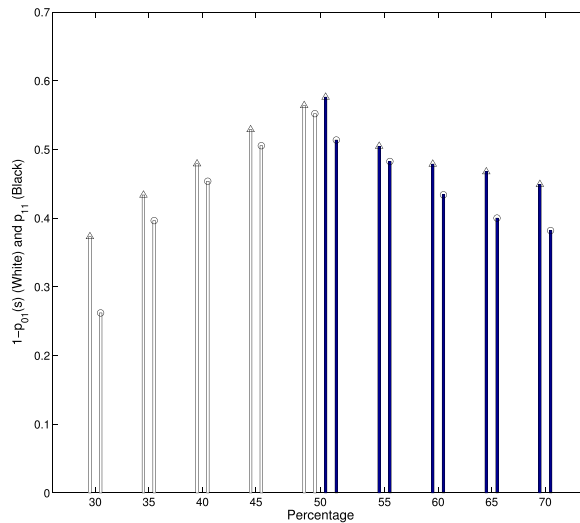


Fig. 10. White bars: $P(T_{n+1} < s | T_n < s)$, for $s = 30, 35, 40, 45, 50$ th percentiles. Blues bars: $P(T_{n+1} > s | T_n > s)$, for $s = 50, 55, 60, 65, 70$ th percentiles. Triangles denote empirical values and circles denote estimated values.

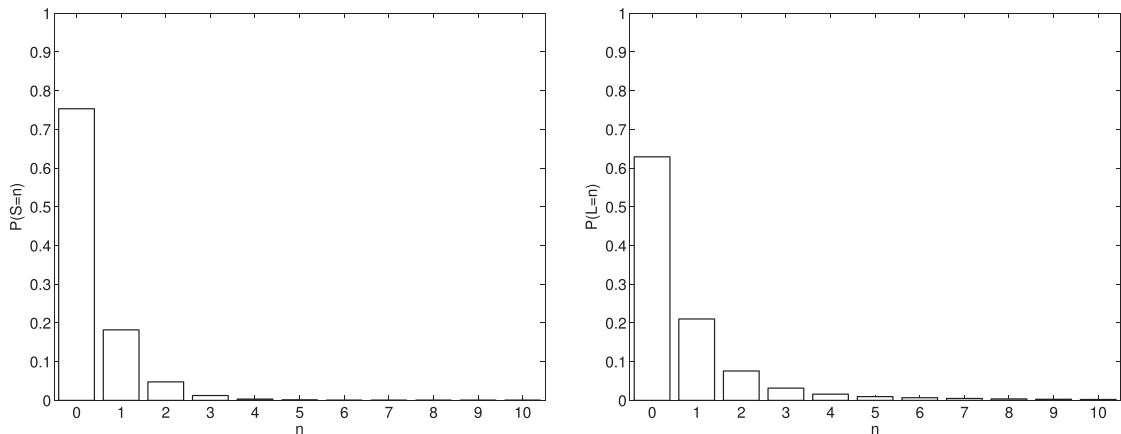


Fig. 11. Estimated values for the mass probability functions of the short (S , left panel) and long (L , right panel) spells of the OpRisk dataset.

0.1419 and 0.0476, respectively. If the interest is in the sequence of long spells, the empirical values for the probabilities $P(L = 0)$, $P(L = 1)$ and $P(L = 2)$ were 0.6161, 0.1875 and 0.0580 and the estimated ones were 0.6291, 0.2101 and 0.0759.

5.3. The estimated severities' distribution

Concerning the severity distribution, we fit the data by a double Pareto Lognormal ($dPIN$) distribution. Such distribution has been suggested in the literature as a versatile heavy-tailed model able to correctly model both the tail and body of the distribution and also, to capture different forms of asymmetry. See for example, [63] or [61]. We briefly recall its definition. A random variable X follows a double Pareto Lognormal distribution with parameters $(\alpha, \beta, \mu, \sigma^2)$ if $X = e^Y$ where Y is random variable distributed according a Normal Laplace distribution, that is, $Y = Z + W$, where $Z \sim N(\mu, \sigma^2)$, and W is distributed according a skewed Laplace model, that is,

$$f_W(w|\alpha, \beta) = \begin{cases} \frac{\alpha\beta}{\alpha+\beta} e^{\beta w} & \text{if } w \leq 0, \\ \frac{\alpha\beta}{\alpha+\beta} e^{-\alpha w} & \text{if } w > 0, \end{cases}$$

independent of Z , for $\alpha, \beta > 0$.

[63] illustrate the form of the $dPIN$ density function for various different groups of parameter values. In particular, they show that it exhibits heavy-tail behavior, since $f_X(x) \rightarrow kx^{-\alpha-1}$ as $x \rightarrow \infty$. The $dPIN$ distribution lacks a closed form expres-

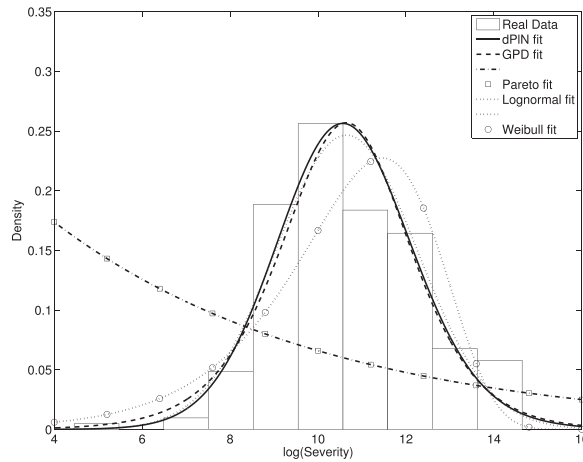


Fig. 12. Histogram and fitted pdf for the real severities in log-scale under the *dPIN* model (solid line) and benchmark approaches.

Table 1

Monte Carlo algorithm to sample from the loss aggregate distribution.

1. For $k = 1, \dots, K$,
 - (a) Simulate the number of annual losses, N , from the estimated MAP_2 given by (25). The estimated mass function of N to be used is depicted by Fig. 9.
 - (b) Given that $N = n$, simulate independent severities X_1, \dots, X_n from the *dPIN* distribution estimated in Section 5.3, with parameters $(\alpha, \beta, \mu, \sigma^2) = (1.24, 1.8, 10.4, 1.29^2)$.
 - (c) Compute $Z_k = \sum_{i=1}^n X_i$.
2. Set $k = k + 1$ and return to 1.

sion for its moment generating function. Nevertheless, if $r < \alpha$, the moment of order r is given by

$$E(X^r \mid \alpha, \beta, \mu, \sigma^2) = \frac{\alpha \beta}{(\alpha - r)(\beta + r)} e^{r\mu + r^2\sigma^2/2}.$$

The approach for Bayesian estimation of the *dPIN* distribution in [61] was applied to fit the severity data sample. The (posterior) expected values for the parameters of the model were $(\alpha, \beta, \mu, \sigma^2) = (1.24, 1.8, 10.4, 1.29^2)$. Note that, since $\alpha < 2$ then, the estimated distribution has a finite mean but an infinite variance. The fit to the histogram of the severities sample (in log-scale) by the *dPIN* distribution is shown by Fig. 12, as well as the fits by other heavy-tailed models typically used in the OpRisk contexts, namely, the Lognormal, Weibull, Pareto and Generalized Pareto. It can be seen how the first three have a poorer performance than the *dPIN*, while the Generalized Pareto was comparable.

5.4. The estimated loss aggregate distribution and risk measures

From the estimates obtained in Sections 5.2 and 5.3 we make inference here for the loss aggregate model (1). As commented in Section 4, a number of numerical strategies do exist in order to estimate (1). Comparing the different approaches is out of the scope in this paper, and here for simplicity, we shall approximate the loss aggregate distribution by the classic, Monte Carlo (MC) method. The steps to implement the MC algorithm are shown in Table 1.

Once the samples (Z_1, \dots, Z_k) from the loss aggregate distribution are obtained, the risk measures (23) and (24) can be easily estimated by their empirical counterparts.

One of the main drawbacks of the MC algorithm is its typical slow convergence, specially in presence of heavy tails, see for example [9,71]. In order to choose an appropriate number of simulations K , the following numerical exercise was considered. A total of K samples of Z , for $K \in \{10^4, 10^5, 10^6, 10^7\}$, were obtained 500 times, from which the corresponding estimated $Var_{99.9\%}$ s were recorded. The 500 samples of the $Var_{99.9\%}$ under the four considered values of K are depicted in Fig. 13. It can be noted that under $K = 10^7$, the convergence is achieved.

Section 5.2 showed the outperformance of the MAP_2 with respect to the Poisson process when fitting the loss occurrence process. It is natural to wonder if such encountered differences affect the estimated compound models and related risk measures. To address this issue, the algorithm in Table 1 was implemented, also for $K = 10^7$, but here the distribution of the annual losses N in point 1.(a) was set to a Poisson distribution of rate equal to the average number of annual losses (16.6154). The obtained results evidence the sensitivity of the aggregate loss distribution to the choice of the occurrence process. Table 2 summarizes both aggregate distributions by means of an assortment of descriptive statistics (minimum and maximum values, mean, standard deviation, skewness and different quantiles). The values have been divided by 10^7 for abbreviation reasons. A first remark to be made is that the minimum value and the 0.025 quantile of Z under the MAP_2

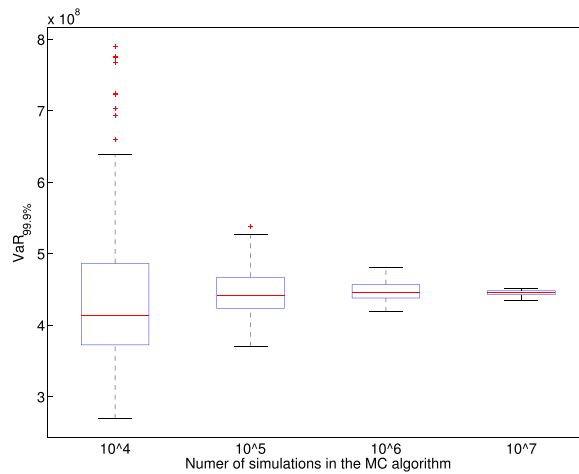


Fig. 13. Boxplot of the samples of $VaR_{99.9\%}$ (500 values each), for different values of K .

Table 2
Values (divided by 10^7) for the descriptive statistics of the aggregate loss distribution Z under the MAP_2 and Poisson process.

	min	max	mean	SD	γ	$Q_{0.025}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.975}$
MAP_2	0	8420	1.4	16	321.9	0	0.72	0.99	4.47
Poisson	0.00045	5000	0.41	8.6	431.6	0.0525	0.1303	0.2073	1.44

Table 3
 VaR_p and ES_p values (divided by 10^8 and 10^9 , respectively) under the MAP_2 and Poisson process, for $p \in \{0.95, 0.99, 0.999\}$.

	VaR_p		ES_p	
	$p = 0.99$	$p = 0.999$	$p = 0.99$	$p = 0.999$
MAP_2	0.8068	4.405	0.3181	1.6887
Poisson	0.27	1.575	0.1209	0.613

are equal to zero. Indeed, there exists a proportion (around 6%) of null samples of Z . This is a consequence of the non-negligible probability that $N = 0$, under the MAP_2 (see Fig. 9). This phenomenon is rarely observed under the considered Poisson process for which $P(N = 0) \approx 6 \times 10^{-8}$ is obtained. Second, as expected (see [27]), both distributions are positive skewed with a long right tail as a consequence of the heavy-tailed severities. In all cases (with the exception of the 0.025 quantile), the quantiles and mean of Z are larger under the MAP_2 than under the Poisson process. This can be explained again by the mass function of the annual occurrences under the two considered point processes: while for the Poisson process $P(N \geq 30) = 0.002$, this probability changes to 0.2836, under the MAP_2 . This implies that for some years, a very high number of losses may occur under the MAP_2 , and this makes the loss aggregate distribution take more extreme values than under the Poisson process.

Finally, Table 3 shows the risk measures of interest for the financial entities: the VaR_p and ES_p (see Eq. (23)–(24)), for $p \in \{0.99, 0.999\}$ and under the two point processes. The shown values are divided by 10^8 and 10^9 , respectively. Here again, the results suggest that under the MAP_2 process, higher potential losses than under the Poisson process may occur, and in consequence, the capital charge increases in significant way.

6. Conclusions

In this paper, an aggregate loss model with occurrence process governed by a type of Markov renewal process has been considered. The selected stochastic process is the MAP_2 which allows for correlated and non-exponential inter-loss times and also for overdispersed loss counts. We have shown in Section 3 how the MAP_2 mostly produces loss counts presenting a variance significantly higher than the mean, a feature that highlights its applicability in financial contexts. One of the novelties of the paper is the derivation of some persistence measures that may help the financial decision makers when evaluating the historical trace of loss times. The aggregate loss model has been estimated by modeling in separate way the frequency from the severity distributions. For the first, an approach based on the direct maximization of the likelihood function of the inter-loss times distributions, has been explored. The application of the methodology to the OpRisk modeling constitutes a significant contribution of the work. A real OpRisk data set representing inter-loss times and associated severities is fitted:

for comparison reasons, the times are modeled by both the MAP_2 and the Poisson process, and for the severities the same heavy-tailed distribution is considered. The outperformance of the MAP_2 in fitting the occurrence process leads to important differences in the estimated loss aggregate distribution and related risk measures, which points out the sensitivity of the aggregate loss distribution to the choice of the frequency distribution in this case.

The authors aim to test the potential for modeling aggregate loss distributions of the Batch Markovian arrival process ($BMAP$), an extension of the MAP which allows for simultaneous (and correlated) losses, see [42,43]. Several perspectives may be considered in this respect. First, and similarly as in the present paper, the $BMAP$ can be thought as a model for the frequency while any type of heavy-tailed distribution is chosen for the severity. In principle, this approach would show more versatility and applicability than the MAP because in real risk scenarios different losses may occur at the same time. Second, the previous approach could be extended to model aggregate loss distributions more than one cell via the superposition of differently parametrized $BMAPs$. From the statistical inference viewpoint the $BMAP$ represents a partially solved problem, mainly because of its non-identifiability, lack of a known canonical representation and consequent numerical troubles. Finally, as already commented throughout the manuscript, a more detailed inspection of the formulae for the persistence measures should be undertaken. In particular, given the aspect of the mass probability functions for the sequence of spells, it is of interest to study the possible connection with zero-inflated models. Work on these issues is underway.

Acknowledgments

This research has been financed in part by research projects FQM-329 and P18-FR-2369 (Junta de Andalucía, Spain); PID2019-110886RB-I00, PID2019-104901RB-I00 (Ministerio de Economía, Industria y Competitividad, Spain); PR2019-029 (Universidad de Cádiz, Spain); This support is gratefully acknowledged.

Appendix A. Derivation of the expressions for $p_{01}(s)$ and $p_{11}(s)$

The proof is a consequence of the theory of Markov renewal processes. All results used in this Appendix can be found in Ch. 10 in [12].

First, $p_{01}(s)$ is written as

$$p_{01}(s) = \frac{P(T_{n+1} > s, T_n < s)}{P(T_n < s)}, \tag{26}$$

where according to (8), $P(T_n < s) = F_T(s) = 1 - \phi e^{D_0 s} \mathbf{e}$. Recall that since the MAP_2 is a Markov renewal process, then the sequences of states $\{Y_n\}_{n=1}^\infty$ at the times of the losses is a Markov chain with transition matrix P^* . Consider the semi-Markov kernel $Q(\cdot)$ as in (3) over the state space $\mathcal{S} = \{1, 2\}$. Then, by definition of semi-Markov kernel, for $i, j \in \mathcal{S}, t > 0$,

$$Q_{i,j}(t) = P(Y_{n+1} = j, T_{n+1} \leq t \mid Y_n = i),$$

which represents the (i, j) th element of the semi-Markov kernel $Q(t)$. Also, define

$$G(i, j, t) = \frac{Q_{i,j}(t)}{P^*(i, j)}, \quad i, j \in \mathcal{S}, t > 0,$$

where $G(i, j, t) = 1$ in case $P^*(i, j) = Q_{i,j}(t) = 0$. Then, $G(i, j, \cdot)$ is a cumulative distribution function

$$G(i, j, t) = P(T_{n+1} \leq t \mid Y_n = i, Y_{n+1} = j).$$

Finally, the following paramount property is key for our proof. This states that given the Markov chain, then the inter-loss times are conditionally independent, or in other words,

$$P(T_1 \leq u_1, \dots, T_n \leq u_n \mid Y_0, \dots, Y_n) = G(Y_0, Y_1, u_1)G(Y_1, Y_2, u_2), \dots, G(Y_{n-1}, Y_n, u_n).$$

Consider the numerator in (26). This can be rewritten as

$$P(T_n < s, T_{n+1} > s) = \tag{27}$$

$$\begin{aligned} & \sum_{i,j,k \in \{1,2\}} \{P(T_n < s, T_{n+1} > s \mid Y_{n-1} = i, Y_n = j, Y_{n+1} = k) \times \\ & \quad \times P(Y_{n-1} = i, Y_n = j, Y_{n+1} = k)\} = \\ & \sum_{i,j,k \in \{1,2\}} \{P(T_n < s, T_{n+1} > s \mid Y_{n-1} = i, Y_n = j, Y_{n+1} = k) \times \\ & \quad \times P(Y_{n+1} = k \mid Y_n = j, Y_{n-1} = i)P(Y_n = j, Y_{n-1} = i)\} = \\ & \sum_{i,j,k \in \{1,2\}} \{P(T_n < s, T_{n+1} > s \mid Y_{n-1} = i, Y_n = j, Y_{n+1} = k) \times \end{aligned}$$

$$\begin{aligned}
 & \times P(Y_{n+1} = k \mid Y_n = j)P(Y_n = j \mid Y_{n-1} = i)P(Y_{n-1} = i) \} = \\
 & \sum_{i,j,k \in \{1,2\}} \{ P(T_n < s, T_{n+1} > s \mid Y_{n-1} = i, Y_n = j, Y_{n+1} = k) \times \\
 & \quad \times P^*(i, j) P^*(j, k) \phi_i \} = \\
 & \sum_{i,j,k \in \{1,2\}} G(i, j, s)(1 - G(j, k, s))P^*(i, j) P^*(j, k) \phi_i = \\
 & \sum_{i,j,k \in \{1,2\}} \frac{Q_{i,j}(s)}{P^*(i, j)} \left(1 - \frac{Q_{j,k}(s)}{P^*(j, k)} \right) P^*(i, j) P^*(j, k) \phi_i = \\
 & \sum_{i,j,k \in \{1,2\}} \frac{Q_{i,j}(s)}{P^*(i, j)} \left(\frac{P^*(j, k) - Q_{j,k}(s)}{P^*(j, k)} \right) P^*(i, j) P^*(j, k) \phi_i = \\
 & \sum_{i,j,k \in \{1,2\}} Q_{i,j}(s) (P^*(j, k) - Q_{j,k}(s)) \phi_i. \tag{28}
 \end{aligned}$$

It is a straightforward computation to check that (Appendix A) can be expressed in matrix form as

$$\sum_{i,j,k \in \{1,2\}} Q(i, j, s) (P^*(j, k) - Q(j, k, s)) \phi_i = \boldsymbol{\phi} Q(s) (P^* - Q(s)) \mathbf{e}. \tag{29}$$

From (3) and (4) one obtains

$$\boldsymbol{\phi} Q(s) (P^* - Q(s)) \mathbf{e} = \boldsymbol{\phi} (I - e^{D_0 s}) P^* e^{D_0 s} P^* \mathbf{e},$$

which concludes the proof for the expression of p_{01} . Finally, the proof for the matrix formulation of p_{11} is obtained in analogous way.

Appendix B. Derivation of the expressions for $P(S = n)$ and $P(L = n)$

Consider first the case $n = 0$. Then, from (8)

$$P(S = 0) = P(T_{n+1} > s) = 1 - F_T(s) = \boldsymbol{\phi} e^{D_0 s} \mathbf{e},$$

and similarly for the variable L . If instead $n \geq 1$, the proof follows closely the derivation of (27). Indeed, by a recursive argument,

$$\begin{aligned}
 P(S = n) &= P(T_1 < s, T_2 < s, \dots, T_n < s, T_{n+1} > s) = \\
 & \sum_{i_1, i_2, \dots, i_{n+2} \in \{1,2\}} \{ P(T_1 < s, T_2 < s, \dots, T_n < s, T_{n+1} > s \mid Y_0 = i_1, Y_1 = i_2, \dots, Y_n = i_{n+1}, Y_{n+1} = i_{n+2}) \times \\
 & \quad \times P(Y_0 = i_1, Y_1 = i_2, \dots, Y_n = i_{n+1}, Y_{n+1} = i_{n+2}) \} = \\
 & \dots = \sum_{i_1, i_2, \dots, i_{n+2} \in \{1,2\}} G(i_1, i_2, s) G(i_2, i_3, s) \dots G(i_n, i_{n+1}, s) (1 - G(i_{n+1}, i_{n+2}, s)) P^*(i_1, i_2) \dots P^*(i_{n+1}, i_{n+2}) \phi_{i_1} \\
 &= \sum_{i_1, i_2, \dots, i_{n+2} \in \{1,2\}} \left(\prod_{l=1}^n G(i_l, i_{l+1}, s) \right) (1 - G(i_{n+1}, i_{n+2}, s)) \left(\prod_{l=1}^n P^*(i_l, i_{l+1}) \right) P^*(i_{n+1}, i_{n+2}) \phi_{i_1} \\
 &= \sum_{i_1, i_2, \dots, i_{n+2} \in \{1,2\}} \left(\prod_{l=1}^n Q_{i_l, i_{l+1}}(s) \right) (P^*(i_{n+1}, i_{n+2}) - Q_{i_{n+1}, i_{n+2}}(s)) \phi_{i_1}.
 \end{aligned}$$

Then,

$$\sum_{i_1, i_2, \dots, i_{n+2} \in \{1,2\}} \left(\prod_{l=1}^n Q_{i_l, i_{l+1}}(s) \right) (P^*(i_{n+1}, i_{n+2}) - Q_{i_{n+1}, i_{n+2}}(s)) \phi_{i_1} = \boldsymbol{\phi} Q^n(s) (P^* - Q(s)) \mathbf{e},$$

and from (3) and (4),

$$\boldsymbol{\phi} Q^n(s) (P^* - Q(s)) \mathbf{e} = \boldsymbol{\phi} [(I - e^{D_0 s}) P^*]^n e^{D_0 s} P^* \mathbf{e}.$$

To see that (21) is indeed a mass probability function, we next show that $\sum_{n=0}^{\infty} P(S = n) = 1$. First, from Ch. 4 in [37], one has

$$\lim_{n \rightarrow \infty} Q^n(s) = \mathbf{0},$$

which implies that $\sum_{n=1}^{\infty} Q^n(s) = Q(s)(I - Q(s))^{-1}$. Then,

$$\begin{aligned} \sum_{n=0}^{\infty} P(S = n) &= \phi e^{D_0 s} \mathbf{e} + \sum_{n=1}^{\infty} \phi Q^n(s) e^{D_0 s} P^* \mathbf{e} \\ &= \phi e^{D_0 s} \mathbf{e} + \phi Q(s)(I - Q(s))^{-1} e^{D_0 s} P^* \mathbf{e} \\ &= \phi e^{D_0 s} \mathbf{e} + \phi Q(s)(I - Q(s))^{-1} (P^* - Q(s)) \mathbf{e} \\ &= \phi e^{D_0 s} \mathbf{e} + \phi Q(s)(I - Q(s))^{-1} (P^* \mathbf{e} - Q(s) \mathbf{e}) \\ &= \phi e^{D_0 s} \mathbf{e} + \phi Q(s)(I - Q(s))^{-1} (\mathbf{e} - Q(s) \mathbf{e}) \\ &= \phi e^{D_0 s} \mathbf{e} + \phi Q(s)(I - Q(s))^{-1} (I - Q(s)) \mathbf{e} \\ &= \phi e^{D_0 s} \mathbf{e} + \phi Q(s) \mathbf{e} \\ &= \phi (e^{D_0 s} + P^* - e^{D_0 s} P^*) \mathbf{e} \\ &= \phi e^{D_0 s} \mathbf{e} + \phi P^* \mathbf{e} - \phi e^{D_0 s} P^* \mathbf{e} \\ &= \phi e^{D_0 s} \mathbf{e} + 1 - \phi e^{D_0 s} \mathbf{e} \\ &= 1 \end{aligned}$$

In the previous set of formulae we applied that since P^* is a stochastic matrix, then $P^* \mathbf{e} = \mathbf{e}$. Also, since ϕ represents its stationary probability vector, then $\phi P^* = \phi$.

Finally, the proof for $P(L = n)$ is immediate following a similar reasoning.

References

- [1] S. Ahn, A.L. Badescu, On the analysis of the Gerber–Shiu discounted penalty function for risk processes with Markovian arrivals, *Insur.: Math. Econ.* 41 (2) (2007) 234–249.
- [2] M. Ausín, J. Vilar, R. Cao, C. González-Fragueiro, Bayesian analysis of aggregate loss models, *Math. Finance* 21 (2) (2011) 257–279.
- [3] B. Avanzi, B. Wong, X. Yang, A micro-level claim count model with overdispersion and reporting delays, *Insur.: Math. Econ.* 71 (2016) 1–14.
- [4] A. Badescu, S. Drešćić, D. Landriault, On the analysis of a multi-threshold Markovian risk model, *Scand. Actuar. J.* 2007(4) (2007) 248–260.
- [5] L. Bodrog, A. Heindlb, G. Horváth, M. Telek, A Markovian canonical form of second-order matrix-exponential processes, *Eur. J. Oper. Res.* 190 (2008) 459–477.
- [6] C. Bolancé, M. Guillén, J. Gustafsson, J.P. Nielsen, *Quantitative Operational Risk Models*, Chapman & Hall, CRC Press, 2012.
- [7] E. Brechmann, C. Czad, S. Paterlini, Flexible dependence modeling of operational risk losses and its impact on total capital requirements, *J. Bank. Finance* 40 (2014) 271–285.
- [8] L. Breuer, An EM algorithm for batch Markovian arrival processes and its comparison to a simpler estimation procedure, *Ann. Oper. Res.* 112 (2002) 123–138.
- [9] M. Brunner, F. Piacenza, F. Monti, D. Bazzarello, Fat tails, expected shortfall and the Monte Carlo method: a note, *J. Oper. Risk* 4 (1) (2009) 81–88.
- [10] E. Carrizosa, P. Ramírez-Cobo, Maximum Likelihood Estimation in the Two-State Markovian Arrival Process, 2014. arXiv preprint arXiv:1401.3105 [stat.CO].
- [11] G. Casale, E. Z. Zhang, E. Simirni, Trace data characterization and fitting for Markov modeling, *Perform. Eval.* 67 (2010) 61–79.
- [12] E. Çinlar, *Introduction to Stochastic Processes*, Prentice-Hall, Usa, 1975.
- [13] S. Chakravarty, The Batch Markovian arrival process: a review and future work, in: A.K. et al. (Ed.), *Advances in Probability and Stochastic Processes*, Notable Publications, New Jersey, 2001, pp. 21–49.
- [14] S. Chakravarty, A disaster queue with Markovian arrivals and impatient customers, *Appl. Math. Comput.* 214 (2009) 48–59.
- [15] A. Chapelle, Y. Crama, G. Hubner, J. Peters, Basel II and operational risk: Implications for risk measurement and management in the financial sector, 2004, National Bank of Belgium Working Paper, No. 51).
- [16] V. Chavez-Demoulin, P. Embrechts, M. Hofert, An extreme value approach for modeling operational risk losses depending on covariates, *J. Risk Insur.* (2015).
- [17] V. Chavez-Demoulin, P. Embrechts, J. Nešlehová, Quantitative models for operational risk: extremes, dependence and aggregation, *J. Bank. Finance* 30 (10) (2006) 2635–2658.
- [18] A.S. Chernobai, S.T. Rachev, F.J. Fabozzi, *Operational Risk: a Guide to Basel II Capital Requirements, Models, and Analysis*, 180, John Wiley & Sons, 2008.
- [19] E. Cheung, D. Landriault, A generalized penalty function with the maximum surplus prior to ruin in a MAP risk model, *Insur.: Math. Econ.* 46 (2010) 127–134.
- [20] E. Cheung, H. Liu, G. Willmot, Joint moments of the total discounted gains and losses in the renewal risk model with two-sided jumps, *Appl. Math. Comput.* 331 (2018) 358–377.
- [21] E.C. Cheung, D. Landriault, et al., Perturbed MAP risk models with dividend barrier strategies, *J. Appl. Probab.* 46 (2) (2009) 521–541.
- [22] E. Cope, G. Antonini, Observed correlations and dependencies among operational losses in the ORX consortium database, *J. Oper. Risk* 3 (4) (2008) 47–74.
- [23] M. Cruz, *Modeling, Measuring and Hedging Operational Risk*, John Wiley & Sons, 2002.
- [24] D. Dimitrova, V. Kaishev, S. Zhao, On the evaluation of finite-time ruin probabilities in a dependent risk model, *Appl. Math. Comput.* 275 (2016) 268–286.
- [25] K. Dutta, J. Perry, A tale of tails: an empirical analysis of loss distribution models for estimating operational risk capital, 2007, Working paper 06–13, Federal Reserve Bank of Boston).
- [26] P. Embrechts, A. Höing, A. Juri, Using copulae to bound the Value-at-Risk for functions of dependent risks, *Finance Stochast.* 7 (2003) 145–167.
- [27] P. Embrechts, C. Klüppelberg, T. Mikosch, *Modelling Extremal Events: for Insurance and Finance*, 33, Springer Science & Business Media, 1997.
- [28] S. Eum, R. Harris, I. Atov, A matching model for MAP-2 using moments of the counting process, in: *Proceedings of the International Network Optimization Conference, INOC 2007, 2007, Spa, Belgium*
- [29] P. Fearnhead, C. Sherlock, An exact Gibbs sampler for the Markov-modulated Poisson process, *J. Royal Stat. Soc.: Ser. B (Stat. Methodol.)* 68 (5) (2006) 767–784.
- [30] J. Fera-Domínguez, E. Jiménez-Rodríguez, O. Sholarin, Tackling the over-dispersion of operational risk: implications on capital adequacy requirements, *N. Am. J. Econ. Finance* 31 (2015) 206–221.
- [31] P. Fodra, H. Pham, High frequency trading and asymptotics for small risk aversion in a Markov renewal model, *SIAM J. Financ. Math.* 6 (1) (2015) 656–684.

- [32] E. Frostig, On ruin probability for a risk process perturbed by a Lévy process with no negative jumps, *Stochast. Models* 24 (2) (2008) 288–313.
- [33] T. Fung, A. Badescu, X. Lin, Multivariate Cox hidden Markov models with an application to operational risk, *Scand. Actuar. J.* 2019 (8) (2019) 686–710.
- [34] S. Gerhold, U. Schmock, R. Warnung, A generalization of Panger's recursion and numerically stable risk aggregation, *Finance Stochast.* 14 (2010) 81–128.
- [35] Groupe Consultatif Actuariel Européen, Solvency II Glossary: European Commission, Technical Report, 2007.
- [36] H. Heffes, D. Lucantoni, A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance, *IEEE J. Sel. Areas Commun.* 4 (1986) 856–868.
- [37] J. Janssen, R. Manca, *Applied Semi-Markov Processes*, Springer, 2006.
- [38] C. Kim, V. Klimenok, A. Dudin, Analysis of unreliable BMAP/PH/N type queue with Markovian flow of breakdowns, *Appl. Math. Comput.* 314 (2017) 154–172.
- [39] A. Klemm, C. Lindemann, M. Lohmann, Modeling IP traffic using batch Markovian arrival process, *Perform. Eval.* 54 (2) (2003) 149–173.
- [40] S.A. Klugman, H.H. Panjer, G.E. Willmot, *Loss Models: from Data to Decisions*, John Wiley & Sons, 2012.
- [41] J. Lindsey, *Modelling Frequency and Count Data*, 15, Oxford University Press, 1995.
- [42] D. Lucantoni, New results for the single server queue with a Batch Markovian Arrival Process, *Stoch. Models* 7 (1991) 1–46.
- [43] D. Lucantoni, The BMAP/G/1 queue: a tutorial, in: L. Donatiello, R. Nelson (Eds.), *Models and Techniques for Performance Evaluation of Computer and Communication Systems*, Springer, New York, 1993, pp. 330–358.
- [44] D. Lucantoni, K. Meier-Hellstern, M. Neuts, A single-server queue with server vacations and a class of nonrenewal arrival processes, *Adv. Appl. Probab.* 22 (1990) 676–705.
- [45] A. Maegebier, Valuation and risk assessment of disability insurance using a discrete time trivariate Markov renewal reward process, *Insur.: Math. Econ.* 53 (3) (2013) 802–811.
- [46] A.J. McNeil, R. Frey, P. Embrechts, *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton University Press, 2015.
- [47] S. Mitnik, T. Yener, Estimating operational risk capital for correlated, rare events, *J. Oper. Risk* 4 (4) (2009) 1–23.
- [48] S. Narayana, M.F. Neuts, The first two moment matrices of the counts for the Markovian arrival process, *Commun. Stat. Stoch Models* 8 (3) (1992) 459–477.
- [49] W. Nasr, A. Charanek, B. Maddah, MAP fitting by count and inter-arrival moment matching, *Stoch. Model.* (2018). (in Press)
- [50] M. Neuts, J. Li., An Algorithm for the $P(n, t)$ Matrices of a Continuous BMAP, vol. 183 of *Lectures notes in Pure and Applied Mathematics*, Srinivas R. Chakravorthy and Attahiru, S. Alfa, editors. NY: Marcel Dekker, Inc., pp. 7–19.
- [51] M.F. Neuts, A versatile Markovian point process, *J. Appl. Probab.* 16 (1979) 764–779.
- [52] A.C. Ng, H. Yang, On the joint distribution of surplus before and after ruin under a Markovian regime switching model, *Stoch. Process. Their Appl.* 116 (2) (2006) 244–266.
- [53] K. Nystrom, J. Skoglund, *Quantitative operational risk management*, 2002, Working paper, Swedbank, Group Financial Risk Control.
- [54] H. Okamura, T. Dohi, Fitting phase-type distributions and Markovian arrival processes: Algorithms and tools, in: *Principles of Performance and Reliability Modeling and Evaluation*, 2016, pp. 49–75.
- [55] H.H. Panjer, *Aggregate Loss Modeling*, John Wiley & Sons, Ltd. 10.1002/9780470012505.taa016
- [56] H.H. Panjer, *Operational Risk: Modeling Analytics*, John Wiley & Sons, 2006.
- [57] G.W. Peters, P.V. Shevchenko, *Advances in Heavy Tailed Risk Modeling: A Handbook of Operational Risk*, John Wiley & Sons, 2015.
- [58] D. Pfeifer, J. Nešlehová, Modeling and generating dependent risk processes for IRM and DFA, *Astin Bull.* 34 (02) (2004) 333–360.
- [59] P. Ramírez-Cobo, R. Lillo, M. Wiper, Nonidentifiability of the two-state Markovian arrival process, *J. Appl. Probab.* 47 (3) (2010) 630–649.
- [60] P. Ramírez-Cobo, R. Lillo, M. Wiper, Bayesian analysis of the stationary MAP₂, *Bayesian Anal.* 12(4) (2017) 1163–1194.
- [61] P. Ramírez-Cobo, R.E. Lillo, S. Wilson, M.P. Wiper, Bayesian inference for Double Pareto lognormal queues, *Ann. Appl. Stat.* 4(3) (2010) 1533–1557.
- [62] P. Ramírez-Cobo, X. Marzo, A.V. Olivares-Nadal, J.-A. Francoso, E. Carrizosa, M.F. Pita, The Markovian arrival process: a statistical model for daily precipitation amounts, *J. Hydrol.* 510 (0) (2014) 459–471 <https://doi.org/10.1016/j.jhydrol.2013.12.033>.
- [63] W.J. Reed, M. Jorgensen, The Double Pareto-lognormal distribution - a new parametric model for size distributions, *Commun. Stat. Theory Methods* 33(8) (2004) 1733–1753.
- [64] J. Ren, A multivariate aggregate loss model, *Insur.: Math. Econ.* 51 (2) (2012) 402–408.
- [65] G. Reshetar, Dependence of operational losses and the capital at risk, 2008, (Available at <https://doi.org/10.2139/ssrn.1081256>).
- [66] J. Rodríguez, R. Lillo, P. Ramírez-Cobo, Failure modeling of an electrical N-component framework by the non-stationary Markovian arrival process, *Reliab. Eng. Syst. Saf.* 134 (2015) 126–133.
- [67] T. Rydén, An EM algorithm for estimation in Markov-modulated Poisson processes, *Comput. Stat. Data Anal.* 21 (1996) 431–447.
- [68] S. Scott, Bayesian analysis of a two-state Markov modulated Poisson process, *J. Comput. Graph. Stat.* 8 (3) (1999) 662–670.
- [69] J. Shao, A. Papaioannou, A. Pantelous, Pricing and simulating catastrophe risk bonds in a markov-dependent environment, *Appl. Math. Comput.* 309 (2017) 68–84.
- [70] S. Sharma, Operational risk modeling - approaches and responses, *Bimaquest* 20(1) (2020) 15–31.
- [71] P.V. Shevchenko, Calculation of aggregate loss distribution, *J. Oper. Risk* 5 (2) (2010) 3–40.
- [72] P.V. Shevchenko, Implementing loss distribution approach for operational risk, *Appl. Stoch. Models Bus. Ind.* 26 (3) (2010) 277–307, doi:10.1002/asmb.v26:3.
- [73] M. Stutzer, Persistence of averages in financial Markov Switching models: a large deviations approach, *Phys. A* (2020) 124237.
- [74] M. Telek, G. Horváth, A minimal representation of Markov arrival processes and a moments matching method, *Perform. Eval.* 64 (2007) 1153–1168.
- [75] L. Valle, P. Giudici, A Bayesian approach to estimate the marginal loss distributions in operational risk management, *Computational Statistics & Data Analysis* 52 (2008) 3107–3127.
- [76] P. Vallois, C. Tapiero, A claims persistence process and insurance, *Insur.: Math. Econ.* 44 (3) (2009) 367–373.
- [77] C. Xu, C. Zheng, D. Wang, J. Ji, N. Wang, Double correlation model for operational risk: evidence from Chinese commercial banks, *Phys. A* 516 (2019) 327–339.
- [78] Y. Yera, R. Lillo, P. Ramírez-Cobo, Fitting procedure for the two-state Batch Markov modulated Poisson process, *Eur. J. Oper. Res.* 279 (1) (2019) 79–92.
- [79] Z. Zhang, H. Yang, H. Yang, On the absolute ruin in a MAP risk model with debit interest, *Adv. Appl. Probab.* 43 (1) (2011) 77–96.