Multivariate Hawkes process for cyber insurance

Yannick Bessy-Roland, Alexandre Boumezoued, Caroline Hillairet

Multivariate Hawkes process for cyber insurance

Yannick Bessy-Roland[1], Alexandre Boumezoued[2], Caroline Hillairet[3]

April 6, 2020

Abstract

In this paper, we propose a multivariate Hawkes framework for modelling and predicting cyber-attacks frequency. The inference is based on a public dataset containing features of data-breaches targeting the US industry. As a main output of this paper, we demonstrate the ability of Hawkes models to capture self-excitation and interactions of data-breaches depending on their type and targets. In this setting we detail prediction results providing the full joint distribution of future cyber attacks times of occurrence. In addition we show that a non-instantaneous excitation in the multivariate Hawkes model, which is not the classical framework of the exponential kernel, better fits with our data. In an insurance framework, this study allows to determine quantiles for number of attacks, useful for an internal model, as well as the frequency component for a data breach guarantee.

Keywords: Hawkes process; Cyber risk; Data breaches; Clustering; Lasso Inference; Prediction; Thinning algorithm.

[1]Milliman R&D, 14 Avenue de la Grande Armée, 75017 Paris, France.
Email: yannick.bessyroland@gmail.com

[2]Milliman R&D, 14 Avenue de la Grande Armée, 75017 Paris, France.
Email: alexandre.boumezoued@milliman.com

[3]CREST, UMR CNRS 9194, Ensae Paris, Avenue Henry Le Chatelier, 91120 Palaiseau, France.
Email: caroline.hillairet@ensae.fr. The author acknowledges funding from the project Cyber Risk Insurance: actuarial modeling, Joint Research Initiative under the aegis of Risk Fundation, with partnership of AXA, AXA GRM, ENSAE and Sorbonne Université

1 Introduction

With the rise of digital economy, cyber risk has become a major concern for all customer segments. Most research programs on cyber risk focus mainly on cyber security and physical means, in view of developing protection against hacking and data-breaches. Although the development of such strategies is fundamental, no protection is perfect and insurers are intended to play a crucial role in providing financial protection. This explains the expansion of cyber insurance contracts.

In the meantime, few works exist on the consequences of cyber attacks from an insurance point of view, and the scientific literature on pricing and reserving of cyber insurance contracts is not very vast. Topics recently addressed in cyber-insurance are reviewed in Biener et al. [BEW15], Eling and Schnell [ES16], or Marotta et al. [MMN$^+$17]. Most of the work on cyber insurance comes from the field of computer science, or from the economic science. For instance, in the field of computer science, Fahrenwaldt et al. [FWW18] study the topology of infected networks, and Rios et al. [RICVR$^+$19] gather expert judgments using an Adversarial Risk Analysis. Noel et al. [NJWS10] and Homer et al. [HZO$^+$13] consider a modeling through attack graphs to measure the security risk of networks (in [NJWS10]) or to propose an aggregating vulnerability metrics for enterprise networks (in [HZO$^+$13]). Johnson et al. [JBG11] provide analytical models of security games to compute adjusting incentives in order to improve network security. In the field of economic science, one may mention the contributions of Böhme and his co-authors, such as [BK06], [BS$^+$10], [RBC$^+$16]. Saini et al. [SARH11] use the utility theory to compute an insurance premium for cyber risk insurance, while a gametheoric approach is proposed by Wang et al. [Wan19], who investigate a mix between optimal investments in information security and cyber insurance innovation.
Herath and Herath [HH11], Eling and Loperfido [EL17] and Farkas et al. [FLT] studied statistical properties and developed more established insurance modeling methods illustrated on the Privacy Rights Clearinghouse (PRC) database. This database has also been studied by Maillart et al. [MS10] to quantify the distribution and time evolution of cyber risks, and by Edwards et al. [EHF16] who developed Bayesian Generalized Linear Models to investigate trends in data breaches. This public dataset is considered as a benchmark for cyber event analysis.

This paper puts this dataset at the cornerstone of the evaluation approach, using statistical techniques to model dynamic dependence and evolving events and providing an operational tool for insurance companies to quantify cyber risk. It takes into account the evolution of the information, while also quantifying the uncertainty of predictions. This is all the more important given that the threat of cyber risk is rapidly growing and evolving, making it one of the most important social and economic risks.
The modelling of cyber attacks frequency requires to take account of complex dependence effects since the majority of systems have the same flaws and are interconnected. Some works have been done in this direction, as it is the case in [PXXH17] in the cyber security

field. In the insurance framework one can cite [BK06] and [HH11] for a model using copulas; another possible approach is to use network contagion models, as in [XH19]. Due to the presence of accumulation phenomena and contagion, the use of Hawkes processes to understand the frequency of the claims will be proposed. Baldwin et al. [BGI$^+$17] claim that Hawkes processes provides the adequate modelling of cyber attacks into information systems because they capture both shocks and persistence after shocks that may form attack contagion. These processes, through their self-exciting property, are adapted to model aftershocks of cyber attacks.

In this paper we propose to use Hawkes processes, motivated by applications in the field of insurance, including pricing and Solvency Capital Requirement calculation. Hawkes processes have been introduced in [Haw71], they have the peculiarity to model excitation effects. Historically there was a first boom in their application in seismology, since then they have been widely used in many different fields, among which finance, neurology, population dynamics or social network modelling. Amongst the recent papers one can cite, for instance, Boumezoued [Bou16] for population dynamics modeling, [ELL11], [BMM15] and [Hai16] in finance, [EGG10] in credit risk, [ST10], [JD13] and [BMS16] in insurance, and [RLMX17] in the field of Social Media. Such processes have been recently used in the cyber security field, for instance by [BGI$^+$17] who studied the self and mutually exciting properties of the threats to 10 important IP services, using industry standard SANS data or by [PXXH17] who focused on extreme cyber attacks rates. Up to our knowledge, Hawkes processes have not been already used to model cyber-attacks in the insurance framework. This paper proposes to study the application of these processes on the public Privacy Rights Clearinghouse database, that makes a census of data breaches happening in the United-States.

In addition to the fact that computer systems are interconnected between companies and that attacks may be driven by common sources, two statistical arguments motivate our choice to use Hawkes processes. The first one is the rejection of the Poissonian hypothesis. Indeed, a Kolmogorov-Smirnov test for the suitability of the inter-arrival times to an exponential distribution leads to a clear rejection with a p-value close to zero. The second one is the detection of autocorrelation in the number of attacks. Indeed, drawing the number of attacks in a month $t+1$ as a function of the number of attacks in a month t, by type of attacks, leads to a correlation coefficient of 65%; this is depicted in Figure 1.

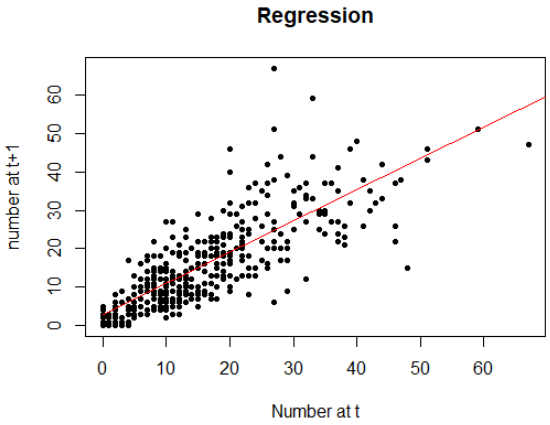

Figure 1: Regression of the number of one-month attacks on the previous one - by type of attacks - $R^2 = 0.6548$

To reproduce autocorrelation between inter-arrival times, two natural choices are the Cox processes and the Hawkes processes (see e.g. Daley [DVJ07] for a survey on point processes). In Cox processes (also known as doubly stochastic Poisson processes), the autocorrelation is captured through the time-dependent intensity that is itself a stochastic process. However, the appropriate specification of the stochastic intensity dynamics remains challenging. Therefore, we resort to the class of Hawkes self-exciting processes which benefit from an interpretable and rather parsimonious parametric representation. In addition to autocorrelation, the Hawkes processes allow to take into account excitation effects, by making arrival rate of events depends on the past events. This seems to make sense in the context of cyber risk, for example, a software flaw discovered will probably generate many attacks in a short time. Another example is the contagion of a virus on other computers, for instance the ransomware Wannacry attack in 2017, that led to a contagion of more than 300 000 computers over more than 150 countries.

In this paper, we propose a stochastic modeling to analyse and predict the arrivals of cyber events. The study is carried out on the PRC database. We specify and infer multivariate Hawkes processes with specific kernel choices to model the dynamics of data breaches times, depending on their characteristics (type, target, location). This modelling framework allows for a causal analysis of the autocorrelation between inter-arrival times according to each data breach feature, and provides forecasts of the full joint distribution of the future times of attacks. This is a first step in the actuarial quantification of cyber risk, more especially on its frequency component. Nevertheless, it is almost impossible to know from the PRC database which part of the variation along time of the reported claims is caused by an evolution of the risk, and which part is caused by an instability in the way the data are collected. Indeed, to fully capture the frequency component, one should also have information about the exposure. However, capturing this component remains challenging since one should track the number of entities by sector exposed to cyber risk over time, along with ensuring the exhaustiveness of the claims reporting process within

the PRC dataset. That is why the exposure quantification will not be addressed in this work. Concerning the severity component, the PRC dataset provides a proxy in terms of the number of records breached for each event. This variable is expected to be strongly correlated with the real loss of the event. Nevertheless, the standard formulas used to deduce a loss from the number of breached records are questionable, such as Jacobs formula [Jac14] proposed in 2014 (using data gathered by Ponemon Institute for the publication of the 2013 and 2014 Cost of Data Breach), which has been recently updated by Farkas et al. [FLT]. Again, a better assessment of the severity risk would require information about the characteristics of the type of breach and of the breached entity, and that are not available on the PRC database.

The remainder of this paper is as follows. In Section 2, we describe the data breach dataset from the Privacy Rights Clearinghouse and the classification of data breach features as used in this study. Section 3 describes the multivariate Hawkes modelling framework, the kernel specifications considered as well as the likelihood. The main inference and prediction results are detailed in Section 4, while our supporting results on the closed-form expectation of the multivariate Hawkes model are given in Appendix B.

2 Dataset

The analysis is based on the dataset from the Privacy Rights Clearinghouse (PRC) that is described below, as well as the different classes of data breach that will be considered.

2.1 Description

The dataset from the Privacy Rights Clearinghouse (PRC)[1] contains 8871 data breaches which have been made public since 2005. Our study focuses on the period 2010-2019, to avoid too much heterogeneity in the type of sources reporting the cyber breaches to the PRC dataset. Indeed, although data breach notification laws have been enacted at different dates in different states, many of them were enacted before 2010, therefore we decided to study the database from this date. For each breach, the following information is available:

- Name of the covered entity
- Type of the covered entity
- Localization of the breached entity
- Breach submission date
- Type of breach
- Number of individuals affected
- Localization of the breached information

[1] see `https://www.privacyrights.org/data-breaches`

- Source of information (US Government Agencies, Non-profit organizations, Media...)

In Table 1, we present the classification of the types of breaches within the PRC (column "Origin"), as well as our preliminary aggregation as used in this paper (column "Aggregation"), which will be further discussed in Section 4. The main types of breaches and entities recorded in the dataset are depicted in Figure 2 and Figure 3.

Aggregation	Origin	Description
HACK	**HACK**	Hacked by outside party or infected by malware
OTHER	**CARD**	Fraud involving debit and credit cards that is not accomplished via hacking
OTHER	**INSD**	Insider (someone with legitimate access intentionally breaches information, such as an employee, contractor or customer)
THEFT/LOSS	**PHYS**	Includes paper documents that are lost, discarded or stolen (non electronic)
THEFT/LOSS	**PORT**	Lost, discarded or stolen laptop, PDA, smartphone, memory stick, CDs, hard drive, data tape, etc
THEFT/LOSS	**STAT**	Stationary computer loss (lost, inappropriately accessed, discarded or stolen computer or server not designed for mobility)
DISC	**DISC**	Unintended disclosure, for example: sensitive information posted publicly, mishandled or sent to the wrong party
	UNKNBREACH	Unknown

Table 1: Type of breaches - Origin refers to the PRC classification - Aggregation refers to our classification

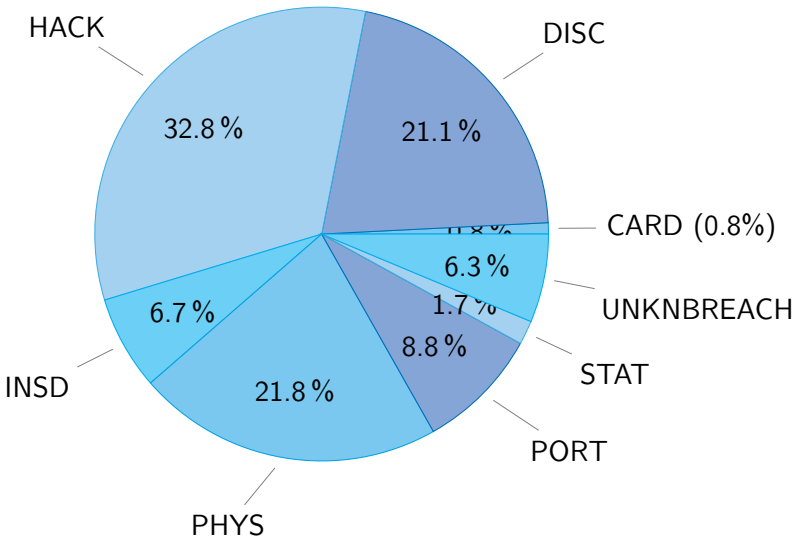

Figure 2: PRC types of breaches

Figure 2 shows three main categories that contain around 75.7% of the data. One can also remark the high level of unintended disclosure (21.1%) and non electronic data (21.8%).

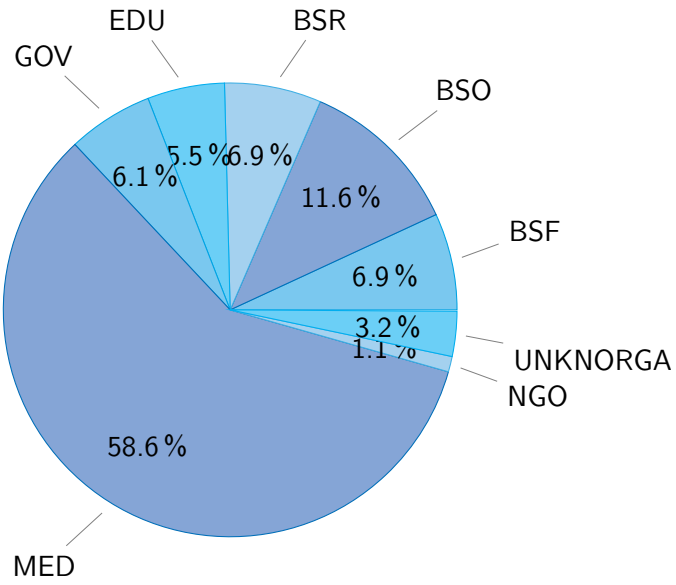

Figure 3: PRC types of organizations breached

The Healthcare industry seems to experience far more breaches than others, even than the Businesses, despite they have a lower exposure in terms of number of entities. It may be explained by the fact that personal health information is often more valuable on the black market than other data as it is the case for credit card credentials. The second industry mainly targeted is the Businesses. The PRC classification as well as our own aggregation are depicted in Table 2.

Aggregation	Origin	Description
BUSINESSES	**BSF**	Businesses-Financial and Insurance Services
BUSINESSES	**BSO**	Businesses - Other
BUSINESSES	**BSR**	Businesses-Retail/Merchant - Including Online Retail
OTHERORGA	**EDU**	Educational Institutions
OTHERORGA	**GOV**	Government & Military
OTHERORGA	**ONG**	Nonprofits
MED	**MED**	Healthcare, Medical Providers & Medical Insurance Services
	UNKNORGA	Unknown

Table 2: Type of organization breached - Origin refers to the PRC classification - Aggregation refers to our classification

In Appendix A, the table with the number of breaches reported in each state shows a clear heterogeneity. While California counts more than 15% of the total data available, many states as Alabama or Arkansas are under 1%. This fact could be linked with the age of the notification law in each state as well as the exposure in terms of number of entities at risk.

2.2 Aggregation

In order to get larger groups for the following study, we decided to group the attack types according to their similarities, see Table 1. The types **PHYS**, **PORT** and **STAT** are gathered in a new category named **Theft/Loss**. In the same way we grouped **CARD** and **INSD** in a new category named **Other**. These categories represent respectively 32.3% and 7.5% of the database.
Concerning the organization types, with the same arguments, we grouped **BSF**, **BSO** and **BSR** in a category named **BUSINESSES**. **NGO**, **EDU** and **GOV** are put in a **OtherOrga** category. It leads to a representation of 25.5% and 12.8% of the total database; these are detailed in Table 2.
Finally, because of the high heterogeneity of the number of breaches reported in each state, we made a main group named **OtherStates** (71.1%) and kept the three biggest ones, namely, **California** (15.7%), **Texas** (6.9%) and **New-York** (6.3%).
This granularity which has been derived at this stage of the analysis will be further discussed and aggregated in Section 4.

3 Multivariate Hawkes model

A multivariate Hawkes framework is proposed to model the clustering and autocorrelation of times of cyber attacks in the different group. In this section we present the model and the kernel specifications and we compute the likelihood.

3.1 Model specification

To fix the idea, we recall briefly the definition and main properties of a Hawkes process. A standard (one-dimensional) Hawkes process $(N_t)_{t\geq 0}$ is a self-exciting point process defined by its intensity function $(\lambda_t)_{t\geq 0}$ characterized by a baseline intensity $(\mu_t)_{t\geq 0}$ plus a self-exciting part $\sum_{T_n<t} \phi(t-T_n)$ where T_n is the jump time number n, and ϕ is a function which governs the clustering density of $(N_t)_{t\geq 0}$, also called the excitation function or kernel of the Hawkes process. Recall that the intensity $(\lambda_t)_{t\geq 0}$ represents the "instantaneous probability" to have a jump at time t, given all the past. This basically means that there is a baseline rate $(\mu_t)_{t\geq 0}$ to have a spontaneous jump at t but that also all the previous jumps influence the apparition of a jump at t. The existence (using Picard iteration) and the construction (using a thinning procedure) of such processes can be found in Brémaud and Massoulié [BM96], [BM02]. One could also find in Daley et al. [DVJ07] the main definitions, constructions and models related to point processes in general and Hawkes processes in particular.

In what follows, a multivariate Hawkes process is consider to capture the clustering and the autocorrelation between inter-arrival times, according to each data breach feature. We consider d groups of data breaches; these groups can be defined by crossing the several covariate dimensions as described in Section 2. For example, a given group can relate to

data breaches of:

- the same type (e.g. **THEFT/LOSS**)
- towards same entities (e.g. **MED**)
- in same location (e.g. **California**)

We consider a time origin at zero being set as the beginning of the first year of the historical period. From this time, data breaches occur in each group $i \in \{1, ..., d\}$ at random times denoted $(T_n^{(i)})_{n\geq 1}$. This sequence defines a counting process $(N_t^{(i)})_{t\geq 0}$ as

$$N_t^{(i)} = \sum_{n\geq 1} \mathbf{1}_{T_n^{(i)} \leq t}.$$

Therefore, $N_t^{(i)}$ is the number of data breaches which occurred for group i in the time interval $[0, t]$. The intensity process of $(N_t^{(i)})_{t\geq 0}$ is denoted by $(\lambda_t^{(i)})_{t\geq 0}$.
We propose a Hawkes process to model the self-excitation in each group as well the excitation between groups. It is specified as follows:

- For each $i \in \{1, ..., d\}$, the base intensity is a deterministic, continuous and non-negative map $t \mapsto \mu_t^{(i)}$,
- For each $(i, j) \in \{1, ..., d\}^2$, self and mutually-exciting maps $t \mapsto \phi_{i,j}(t)$ (commonly called kernels) are introduced and also assumed to be deterministic, continuous and non-negative,
- For each $i \in \{1, ..., d\}$, the intensity process of $(N_t^{(i)})_{t\geq 0}$ is specified as follows:

$$\lambda_t^{(i)} = \mu_t^{(i)} + \sum_{j=1}^{d} \sum_{T_n^{(j)} < t} \phi_{i,j}(t - T_n^{(j)}) = \mu_t^{(i)} + \sum_{j=1}^{d} \int_{[0,t[} \phi_{i,j}(t - s) \mathrm{d}N_s^{(j)}. \tag{1}$$

In this model, the maps $\phi_{i,i}$ quantify the self excitation in the group i, whereas for $i \neq j$, the map $\phi_{i,j}$ quantifies the contagion **in group i caused by a data breach in group j**. Note here that each intensity process $\lambda^{(i)}$ is adapted to the canonical filtration associated with the whole set of processes $(N^{(j)})_{j\in\{1,...,d\}}$; in this way the behavior of a given group may (generally) depend on that of the others.

The framework of interest in this paper is made of the two following kernels specifications:

$$\begin{aligned} \phi_{i,j}(t) &= \alpha_{i,j} \exp\left(-\beta_{i,j} t\right). \\ \phi_{i,j}(t) &= \alpha_{i,j} t \exp\left(-\beta_{i,j} t\right). \end{aligned} \tag{2}$$

Examples of such kernels are provided in Figure 4. The first one models an instantaneous excitation when an event occurs, it then decreases exponentially toward zero. It is the most widely used in the literature as it allows the intensity to be Markovian in the univariate case, as well as in the multivariate setting under some restrictions, see the next remark. The second one models a progressive excitation, that reaches its highest level after a time $1/\beta_{i,j}$, after this, the impact decreases toward zero.

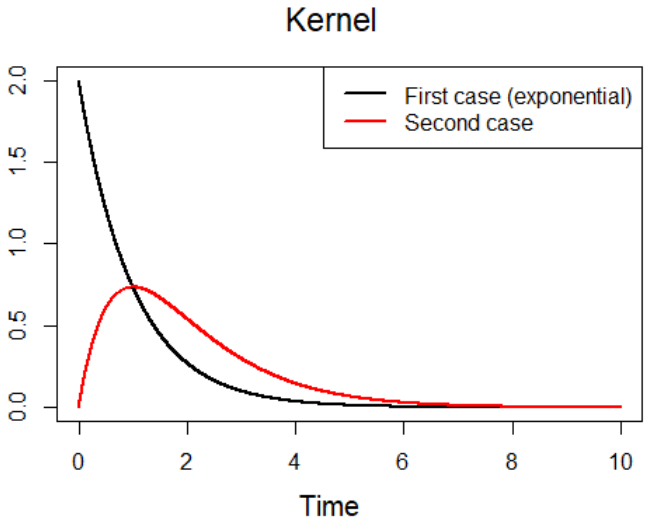

Figure 4: Evolution of kernels through time

Remark 1. *Although other forms of kernels can be considered (see e.g. [Bou16] for the computation of Hawkes distribution for general kernel assumptions), we restrict here our attention to these cases; extension to other kernels encompassing more parameters and related optimal selection is left for further research. Note that this framework is still rich enough. Indeed, the Hawkes process intensity is not Markov with the specification* $\phi_{i,j}(t) = \alpha_{i,j} t \exp(-\beta_{i,j} t)$. *Moreover, for the exponential specification* $\phi_{i,j}(t) = \alpha_{i,j} \exp(-\beta_{i,j} t)$ *with parameters* $\beta_{i,j}$ *depending on the interacting groups* i *and* j*, the vector* $(\lambda^{(1)}, ..., \lambda^{(d)})$ *of the Hawkes intensity processes is not Markov; it is only the case when for any* i *all the* $\beta_{i,j}$ *are constant equal to some* β_i*, that is when the memory of impacts from a group* j *on a group* i *only depends on group* i*. These general frameworks will lead us to consider an extended process with additional well chosen components to recover a higher dimensional dynamics and tractable formulas, see Appendix B. Finally, recall that the multivariate Hawkes process* $(N^{(1)}, ..., N^{(d)})$ *is not Markov in both parametrizations.*

3.2 Likelihood

The aim of this section is to detail the likelihood of the multivariate Hawkes model. This likelihood is known as for any counting process with stochastic intensity, and can be found in related references such as [Oza79]; we still provide its computation here for sake of completeness. Note that we implement a 'brute force' multi-dimensional optimization for maximum likelihood estimation (using the 'Nelder-Mead' algorithm from the 'optim' function in R). There exists alternatives to this approach, for example using a stochastic descent algorithm, see [Jai15] and [BMM15] for a general discussion on inference strategies in a high dimensional framework in the context of financial applications.
Let us consider that the processes are observed on a given interval $[0, \tau]$.

Proposition 1. *The log-likelihood of a multidimensional Hawkes process* $\left(N^{(i)}\right)_{1\leq i\leq d}$ *with baseline intensity* $\left(\mu^{(i)}\right)_{1\leq i\leq d}$ *and kernels* $\left(\phi_{i,j}(t)=\alpha_{i,j}\exp\left(-\beta_{i,j}t\right)\right)_{1\leq i,j\leq d}$ *is*

$$\begin{aligned}\log\mathcal{L} = &-\sum_{i=1}^{d}\int_0^\tau(\mu_s^{(i)}+\sum_{j=1}^{d}\int_0^s\alpha_{i,j}\exp\left(-\beta_{i,j}(s-u)\right)\mathrm{d}N_u^{(j)})\mathrm{d}s\\ &+\sum_{i=1}^{d}\sum_{n=1}^{m_i}\log\left(\mu_{t_n^{(i)}}^{(i)}+\sum_{j=1}^{d}\int_0^{t_n^{(i)}}\alpha_{i,j}\exp\left(-\beta_{i,j}(t_n^{(i)}-s)\right)\mathrm{d}N_s^{(j)}\right)\end{aligned} \tag{3}$$

where $(t_n^{(k)})_{1\leq n\leq m_k}$ *are the* m_k *times of event observed for each group* $k\in\{1,...,d\}$.

Proof of Proposition 1 . (i) For ease of presentation, let us start with the single group case ($d=1$); we omit the group index for simplicity of notation. Note that calibrating such model on each single group amounts to specify $\phi_{i,j}\equiv 0$ for $i\neq j$ in Equation (1). We introduce the notation $\mathcal{H}_n=\{T_n=t_n,...,T_1=t_1\}$ the information on the first n times of event, and add the conventions $\mathcal{H}_0=\emptyset$ and $T_0=0$. Having observed the times $(t_n)_{1\leq n\leq m}$, the likelihood can be written as (by abuse of notation we keep the T_n):

$$\begin{aligned}\mathcal{L} &= \mathbb{P}(\forall 1\leq n\leq m, T_n=t_n, \text{ and } T_{m+1}>\tau),\\ &= \mathbb{P}(T_{m+1}>\tau\mid\mathcal{H}_m)\prod_{n=1}^{m}\mathbb{P}(T_n=t_n\mid\mathcal{H}_{n-1}),\\ &= \exp\left(-\int_{T_m}^{\tau}\lambda_s\mathrm{d}s\right)\prod_{n=1}^{m}\exp\left(-\int_{T_{n-1}}^{T_n}\lambda_s\mathrm{d}s\right)\lambda_{T_n},\\ &= \exp\left(-\int_0^{\tau}\lambda_s\mathrm{d}s\right)\prod_{n=1}^{m}\lambda_{T_n}.\end{aligned}$$

Then the log-likelihood can be written as

$$\log\mathcal{L}=-\int_0^\tau\lambda_s\mathrm{d}s+\sum_{n=1}^{m}\log\lambda_{T_n}=-\int_0^\tau\lambda_s\mathrm{d}s+\int_0^\tau\log\lambda_s\mathrm{d}N_s. \tag{4}$$

This is the standard log-likelihood for any counting process $(N_s)_{s\geq 0}$ with intensity process $(\lambda_s)_{s\geq 0}$. It now remains to further specify it in the context of the model introduced in Equation (1), as

$$\log\mathcal{L}=-\int_0^\tau\mu_s\mathrm{d}s-\int_0^\tau\sum_{T_n<s}\phi(s-T_n)\mathrm{d}s+\sum_{n=1}^{m}\log\left(\mu_{T_n}+\sum_{k=1}^{n-1}\phi(T_n-T_k)\right). \tag{5}$$

If we further specify the self-exciting map as $\phi(t)=\alpha\exp\left(-\beta t\right)$, with non-negative α and β in the present context, we obtain

$$\begin{aligned}\log\mathcal{L}(\mu,\alpha,\beta) = &-\int_0^\tau\mu_s\mathrm{d}s-\alpha\int_0^\tau\sum_{T_n<s}\exp\left(-\beta(s-T_n)\right)\mathrm{d}s\\ &+\sum_{n=1}^{m}\log\left(\mu_{T_n}+\alpha\sum_{k=1}^{n-1}\exp\left(-\beta(T_n-T_k)\right)\right).\end{aligned} \tag{6}$$

(ii) In the multivariate setting, the parameters for all groups are gathered into a vector $M(t) = (\mu_t^{(1)}, ..., \mu_t^{(d)})$, possibly time-dependent, and two $d \times d$ matrices $A = (\alpha_{i,j})_{1 \leq i,j' \leq d}$ and $B = (\beta_{i,j})_{1 \leq i,j \leq d}$. The multivariate intensity is fully specified in Equation (1), and using the same reasoning as above, we can derive the associated log-likelihood based on observation $((t_n^{(i)})_{1 \leq n \leq m_i})_{1 \leq i \leq d}$

$$\begin{aligned}
\log \mathcal{L}(M(.), A, B) &= -\sum_{i=1}^{d} \int_0^\tau \lambda_s^{(i)} \mathrm{d}s + \sum_{i=1}^{d} \sum_{n=1}^{m_i} \log \lambda_{t_n^{(i)}}^{(i)} \\
&= -\sum_{i=1}^{d} \int_0^\tau \left(\mu_s^{(i)} + \sum_{j=1}^{d} \int_0^s \phi_{i,j}(s-u) \mathrm{d}N_u^{(j)} \right) \mathrm{d}s \\
&\quad + \sum_{i=1}^{d} \sum_{n=1}^{m_i} \log \left(\mu_{t_n^{(i)}}^{(i)} + \sum_{j=1}^{d} \int_0^{t_n^{(i)}} \phi_{i,j}(t_n^{(i)} - s) \mathrm{d}N_s^{(j)} \right)
\end{aligned} \tag{7}$$

When specifying the self and mutually-exciting maps $\phi_{i,j}(t) = \alpha_{i,j} \exp(-\beta_{i,j} t)$, the final log-likelihood is given by Equation (3).

Remark 2. *In this form the number of parameters to be estimated is $2d^2$, in addition to the number of parameters of the functions $\left(\mu^{(i)}\right)_{1 \leq i \leq d}$ which will be specified in Section 4.*

Remark 3. *Due to the complexity of the log-likelihood, we resort in this paper to a simplex-type optimization procedure in the form of the Nelder-Mead algorithm. Furthermore, we provide in our paper the inference of the memory parameters $\beta_{i,j}$ in Equation (1), in addition to the self-excitation matrix $(\alpha_{i,j})$. The fact that the $\beta_{i,j}$ are often considered as fixed parameters is discussed in e.g. [BMM15].*

4 Inference and prediction

We decided to use the periods 2011-2015 and 2011-2016 for parameter inference, in order to predict the number of attacks for 2016 and 2017 respectively. The one-year horizon taken here is motivated by applications in internal models for insurance companies, aiming to quantify the 99.5% most adverse one-year loss related to cyber risk insurance covers. Note that the exclusion of the years 2018 and 2019 from the analysis is motivated by the fact that the database appears to be incomplete for these two years, since 2019 is not fully developed, and maybe due to some delays in reporting for year 2018.

4.1 Segmentation

The aim of our study is to calibrate a multivariate Hawkes process on the several groups obtained by crossing the covariates: **Type of breach & Type of the covered entity & State**. These covariates are those discussed in Section 2 and Tables 1, 2 and in Appendix Table A. In particular, we use the aggregated covariates as discussed in Section 2.2. In order to have sufficiently represented groups we kept the largest ones and removed the others in a **OTHER** group. We also put the group **MED & OTHER & OTHER**

in the **OTHER** group because it was to irregular over the period of interest. This is summarized in Table 3.[2]

Group	Number of breaches
OTHER (1)	2046
MED & DISC & OTHER (2)	497
BUSINESSES & HACK & OTHER (3)	386
MED & HACK & OTHER (4)	472
MED & THEFT/LOSS & CALIFORNIA (5)	214
MED & THEFT/LOSS & OTHER (6)	943

Table 3: Studied groups

Figure 5 shows the frequency of attacks over the calibration period (2011-2016). First, the different trends strengthen our idea that this segmentation could indeed make sense. These trends will be taken into account through a dedicated linear specification in the baseline intensity. Furthermore, the clustering behaviour of the data breach occurrences appear, alternating high and low activity periods, see in particular **MED & DISC & OTHER (2)** and **MED & HACK & OTHER (4)**.

[2]The OtherState and OtherOrga groups are simply called OTHER for the following results.

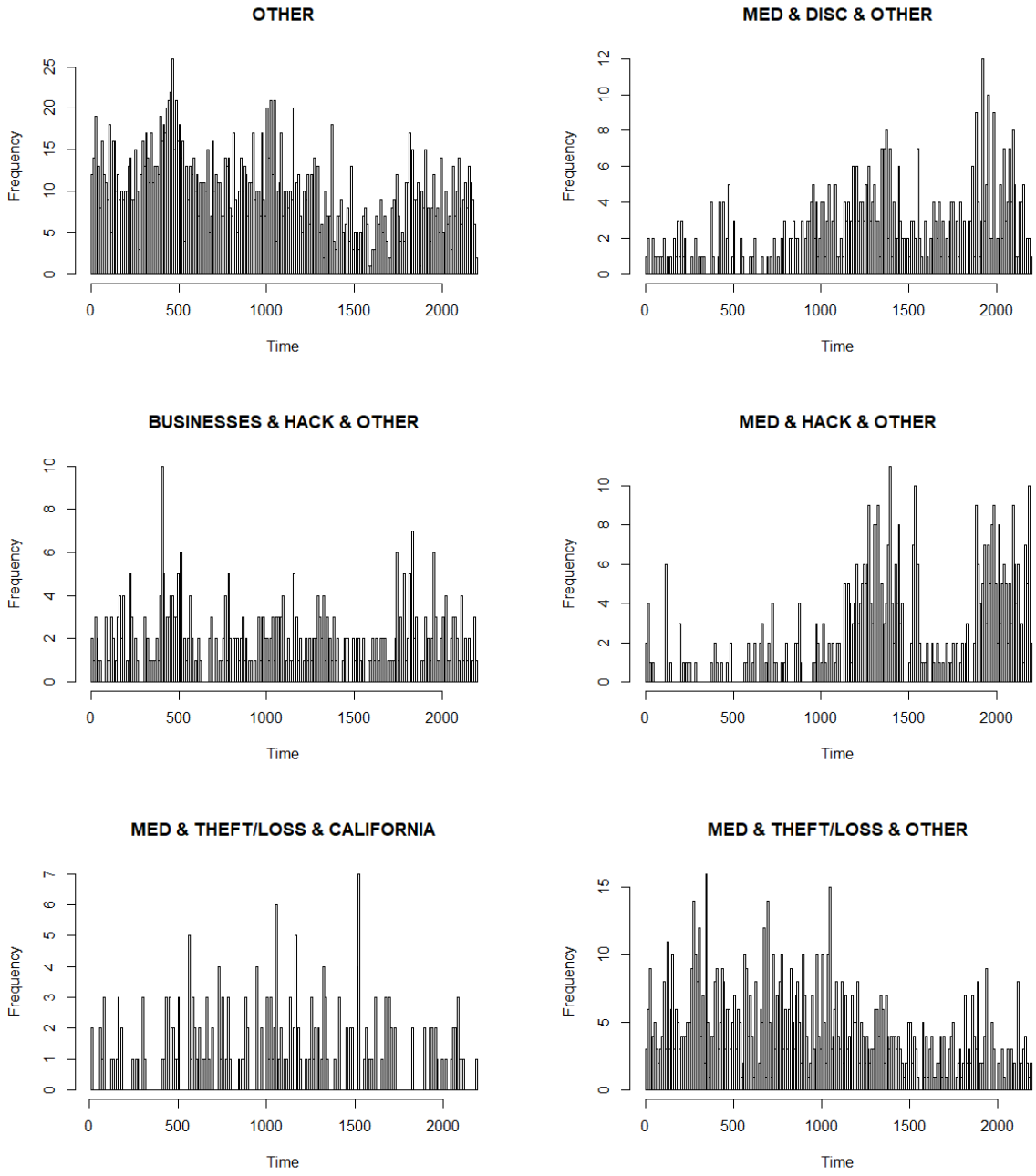

Figure 5: Numbers of attacks function of the time (in days) for each of the 6 groups

4.2 Models studied

Maximum likelihood inference has been performed for the three following kernels:

$$\begin{aligned} &\text{Kernel 1: } \phi_{i,j}(t) = \alpha_{i,j} \exp(-\beta_i t) \\ &\text{Kernel 2: } \phi_{i,j}(t) = \alpha_{i,j} \exp(-\beta_{i,j} t) \\ &\text{Kernel 3: } \phi_{i,j}(t) = \alpha_{i,j} t \exp(-\beta_i t) \end{aligned} \tag{8}$$

In order to take into account possible trends in the dynamics, as depicted in Figure 5, a linear baseline intensity $(\mu_s^{(i)})_{s\geq 0}$ has been specified as:

$$\mu_s^{(i)} = \mu_0^{(i)} + \gamma_i s$$

with $\mu_0^{(i)} \geq 0, \gamma_i \in \mathbb{R}$ for $i \in \{1, ..., d\}$. Note that in case of a negative trend, the parameters are constrained such that the baseline intensity remains positive at the end of the one-year forecasting period, that is for each $i \in \{1, ..., d\}$:

$$\mu_0^{(i)} + \gamma_i(\tau + 1) > 0. \tag{9}$$

This leads, in total, to a number of 54 parameters for kernels 1 and 3, and 84 parameters for kernel 2. Section 4.4 will test a Lasso method to reduce the dimension.
Recall that the first two kernels (exponential case) cause an instantaneous jump of the intensity when an event occur and then the impact decreases exponentially toward zero (the decrease speed is different for each pair (i, j) in the second case). The third kernel reaches its highest level after a time $1/\beta_i$ and then the impact decreases toward zero.

4.3 Inference

The likelihood obtained through the inference process is provided in Table 4.

	Kernel 1	Kernel 2	Kernel 3
Period 2011-2015	6513.19	6171.78	**6152.92**
Period 2011-2016	7639.44	7516.43	**7484.55**

Table 4: Opposite of the log-likelihood for each kernel.

Among the three kernels tested, see Equation (10), it appears that the use of kernel 2 with more parameters provides a better likelihood estimate compared to kernel 1, as expected. However, a key result is that kernel 3 with non-instantaneous excitation provides a better fit than kernel 2, with less parameters. Before detailing the parameter estimates and their interpretation, we present in the following the adequacy test performed.

Adequacy test. We use a statistical test of goodness of fit to evaluate the quality of adjustment of the model. One classical test in the theory of point processes uses the following result, which follows from Theorem 4.1 of Garcia and Kurtz [GK08]. Note first that this requires the intensity to remain positive, which is ensured by Equation (9). Remark also that Theorem 4.1 in [GK08] holds for general counting processes, beyond the class of Hawkes processes.

Proposition 2. *Let us define for any $i \in \{1, ..., d\}$ and $k \geq 1$,*

$$\tau_k^{(i)} = \int_0^{T_k^{(i)}} \lambda_t^{(i)} dt.$$

Then the $(\tau_k^{(i)})_{k\geq 1}$ are the jump times of an homogeneous Poisson process of intensity 1.

This result provides a way to test the adequacy of the Hawkes processes: if the underlying process is indeed a Hawkes process with this intensity, the times

$$\theta_k^{(i)} = \tau_k^{(i)} - \tau_{k-1}^{(i)},\ k \geq 1$$

are independent and distributed according to an exponential distribution with parameter 1. We can then assess the adequacy for each group $i \in \{1, ..., d\}$ of the time series $(\theta_k^{(i)})_{k\geq 1}$ to the exponential distribution with a standard Kolmogorov-Smirnov test. This test is based on comparing the distance between the empirical cumulative distribution and that of a reference specified distribution (here exponential) with known parameters (here 1). The adequacy tests for the different kernels are summarized in Table 5. The cases where the null hypothesis (adequacy) is not rejected at confidence level 5 % are highlighted in bold.

	Kernel 1	Kernel 2	Kernel 3
OTHER (1)	**0.0503**	**0.0865**	**0.9060**
MED & DISC & OTHER (2)	**0.5546**	**0.1300**	**0.5173**
BUSINESSES & HACK & OTHER (3)	**0.5558**	**0.5966**	**0.3363**
MED & HACK & OTHER (4)	0.0024	0.0361	0.0370
MED & THEFT/LOSS & California (5)	**0.1146**	**0.5669**	**0.4246**
MED & THEFT/LOSS & OTHER (6)	**0.0733**	**0.6341**	**0.5379**

Table 5: Adequacy test

From these tests, it appears that the adequacy is satisfactory except for group **MED & HACK & OTHER (4)**, whatever the kernel considered.
Given the quality of the adequacy and of the likelihood presented by kernel 3 ($\phi_{i,j}(t) = \alpha_{i,j} t \exp(-\beta_i t)$), we focus on this kernel for the parameter interpretation which follows.

Parameter estimates - Kernel 3 The parameter estimates for kernel 3 are detailed in Tables 6 and 7; they are interpreted in the following.

	$\mu_0^{(i)}$	β_i	γ_i
OTHER (1)	0.87	5.39	-2.53e-04
MED & DISC & OTHER (2)	0.02	6.88	9.52e-05
BUSINESSES & HACK & OTHER (3)	0.12	7.31	-3.56e-06
MED & HACK & OTHER (4)	0.02	5.75	9.65e-05
MED & THEFT/LOSS & CALIFORNIA (5)	0.05	5.96	-7.26e-06
MED & THEFT/LOSS & OTHER (6)	0.36	5.84	-1.07e-04

Table 6: Parameters $(\mu_0^{(i)})_{1\leq i\leq 6}$, $(\beta_i)_{1\leq i\leq 6}$ and $(\gamma_i)_{1\leq i\leq 6}$

	1	2	3	4	5	6
1	6.04	6.06	4.36	3.51	2.54	2.95
2	1.48	6.28	1.82	4.70	3.31	0.83
3	1.45	1.34	3.17	1.84	0.14	1.15
4	0.31	2.83	1.74	8.37	0.32	0.12
5	0.38	0.62	0.12	1.19	7.80	0.99
6	2.03	2.57	3.15	1.63	0.83	6.70

Table 7: Parameters $(\alpha_{i,j})_{1\leq i,j\leq 6}$

To further analyse the parameters, we also compute in Table 8 the maximum value $\Gamma_{i,j}$ of the influence of an event of group j on the intensity of the group i (for $1 \leq i, j \leq d$). This maximum influence is reached after a certain time, as specified before, up to the value $\Gamma_{i,j} := \frac{\alpha_{i,j}}{\beta_i} e^{-1}$.

Finally, the ratio $(\frac{\Gamma_{i,j}}{\mu_0^{(i)}})_{1\leq i,j\leq 6}$ between the maximum excitation and the baseline intensity is provided in Table 9. This ratio helps to understand the relative importance of the excitation phenomenon compared to the baseline dynamics.

	1	2	3	4	5	6
1	0.41	0.41	0.30	0.24	0.17	0.20
2	0.08	0.34	0.10	0.25	0.18	0.04
3	0.07	0.07	0.16	0.09	0.01	0.06
4	0.02	0.18	0.11	0.53	0.02	0.01
5	0.02	0.04	0.01	0.07	0.48	0.06
6	0.13	0.16	0.20	0.10	0.05	0.42

Table 8: Maximum excitations $(\Gamma_{i,j})_{1\leq i,j\leq 6}$

	1	2	3	4	5	6
1	0.47	0.47	0.3	0.28	0.19	0.23
2	4	**17**	5	**12.5**	**9**	2
3	0.58	0.58	1.33	0.75	0.08	0.5
4	1	**9**	5.5	**26.5**	1	0.5
5	0.4	0.8	0.2	1.4	**9.6**	1.2
6	0.36	0.44	0.56	0.28	0.14	1.17

Table 9: Ratios between the maximum excitation and the baseline intensity $(\frac{\Gamma_{i,j}}{\mu_0^{(i)}})_{1\leq i,j\leq 6}$. The largest values are highlighted in bold.

Interpretation A possible interpretation is the following.
The baseline intensity μ_0 is highest for groups 1 (**OTHER**) and 6 (**MED** & **THEFT/LOSS** & **OTHER**), which simply reflects the fact that they are more represented. This link is not always verified, for example, groups 2 (**MED** & **DISC** & **OTHER**) and 4 (**MED** & **HACK** & **OTHER**) seem to owe their number of attacks more to excitation phenomena than group 3 (**Businesses** & **HACK** & **OTHER**) because they are more represented but do not have a higher base rate ($\mu_0^{(2)} < \mu_0^{(3)}$ and $\mu_0^{(4)} < \mu_0^{(3)}$).
Concerning the drifts, the model seems to have captured the trends visible in Figure 5, they are all decreasing ($\gamma < 0$) except for segments 2 and 4. Segments 3 and 5 (**MED** & **THEFT/LOSS** & **CALIFORNIA**) have very low trend parameters, which also correspond to the histograms.
Table 9 represents the maximum value of excitation $\Gamma_{i,j}$, relatively to the basic intensity $\mu_0^{(i)}$. For $1 \leq i, j \leq d$, the coefficient $\Gamma_{i,j} = \frac{\alpha_{i,j}}{\beta_i} e^{-1}$ represents the maximum value of the influence of an event in group j, on the intensity of the group i, and $\mu_0^{(i)}$ is the constant component of the baseline intensity of group i. Table 9 shows a strong self-excitation of groups 2 and 4, which corresponds to the remark made in the paragraph on baseline intensity, this is also the case for group 5 ($\Gamma_{i,i} >> \mu_0^{(i)}$ for $i = 2, 4$ and 5). The **OTHER** group is the least self-excited ($\Gamma_{1.1} < \mu_0^{(1)}$). The different types of attacks for the medical sector seem to excite each other, attacks of type **HACK** and **DISC** have a clear impact on the intensity of the other, they also trigger attacks of type **THEFT/LOSS**. On the other hand, attacks **THEFT/LOSS** do not seem to have a significant impact on the other two.
Concerning the coefficients $(\beta_i)_{1\leq i\leq 6}$ they are all of the same order of magnitude except for the group **BUSINESSES** & **HACK** & **OTHER** which is higher; this means that the excitation phenomenon is less strong, and vanishes quickly for this group. More specifically, the $(\beta_i)_{1\leq i\leq 6}$ parameters for this kernel 3 indicate that the maximal excitation is globally reached after approximately 4 or 5 hours ($\frac{1}{\beta_i}$ for β_i varying from 5.4 to 7.3) .

From an actuarial perspective, the model parameter estimates allow first to identify sub-groups of entities and related insurance covers (groups 2 and 4 mainly, and 5 to a lesser extent) which still present a strong self-excitation despite that the joint dynamics

with the other groups is captured. This means that for actuarial applications (pricing, reserving) the recourse to self-exciting models as Hawkes processes can not be avoided to appropriately quantify the frequency risk of these groups.
In addition, we have seen that the different types of attacks in the medical sector interact with each other. This involves that an actuarial assessment possibly focusing on a specific guarantee (like theft/loss) will gain if other types of data losses are also involved in the modelling, even if not covered by the insurance contract; this is for example the case of unintended disclosure which has to be modeled since it is a strong explanatory driver to understand the pattern of the theft/loss risk.
These insights are even more important for actuaries when they derive capital requirements related to cyber risk uncertainty. Indeed, the identification of the self-exciting and mutually-exciting behaviors helps to refine the full distribution of the risk, again even if only a few set of entity types and covers (attack types) are of interest among those modelled.

4.4 Penalized likelihood

A first motivation in calibrating Hawkes processes is the natural interpretation given by the parameters. However the number of parameters and therefore the complexity, increases rapidly with the dimension of the Hawkes process. One way to reduce complexity consists in penalizing the likelihood with the norm of the vector of parameters. This penalization should shrink the potential values taken by the parameters. In our case we decided to penalize the $(\alpha_{i,j})_{1\leq i,j\leq d}$ parameters with the $\mathbb{L}_1$-norm[3] (Lasso method). Indeed, this choice should highlight the main interactions between the groups and provide parameter estimates with lower variance (at the price of an increase in the bias). One could have chosen instead a $\mathbb{L}_2$-norm penalization (Ridge method), that also shrinks the coefficients of less contributing variables towards zero, but without setting any of them exactly to zero. We therefore want to minimize the penalized log-likelihood

$$-\log \mathcal{L}(M(.), A, B)_{penalized} = -\log \mathcal{L}(M(.), A, B) + \nu \sum_{1\leq i,j\leq d} |\alpha_{i,j}|$$

were $\nu \geq 0$ is the penalization coefficient, and $\mathcal{L}(M(.), A, B)$ the likelihood of the Hawkes process. Different values of the coefficient ν will be tested, in order to observe how the estimated parameters react, and how the predictive capacity evolves. Let us recall that increasing the value of ν will increase the bias and decrease the variance of the predictions. In our experiment, we observe that the β parameters compensate the penalty constraint on the α parameters, therefore no parameters are forced to zero. This flexibility would be deleted if we would have extended the penalty to the β parameters as well; in such a case we expect we would have obtained the classical vanishing results of some coefficients, at the price of an expected decrease in the prediction capacity.

[3] Given $(d, p) \in \mathbb{N}^* \times \mathbb{N}^*$ and a matrix $A = (a_{i,j})_{1\leq i,j\leq d} \in \mathbb{R}^d$, we define the $\mathbb{L}_p$-norm $||.||_p$ of this matrix by : $||A||_p := \left(\sum_{1\leq i,j\leq d} |a_{i,j}|^p\right)^{1/p}$

The following penalization parameters have been tested, for the three kernels considered (see Section 4.2) : $\nu \in \{0, 100, 600, 900, 3000, 6000\}$.

Penalty coefficient ν	0	100	600	900	3000	6000
Kernel 1 (2011-2015)	6513.19	6375.83	6990.89	7385.11	13261.36	15799.48
Kernel 1 (2011-2016)	7639.44	7755.28	8421.69	8874.91	15143.42	17833.19
Kernel 2 (2011-2015)	6171.78	6278.31	6519.87	6673.23	8340.74	9563.34
Kernel 2 (2011-2016)	7516.43	7647.73	7959.91	8142.27	9835.47	11152.66
Kernel 3 (2011-2015)	6152.92	6528.24	6998.68	7407.18	10914.03	16016.69
Kernel 3 (2011-2016)	7484.55	7888.62	8253.16	8914.18	35829.76	78405.08

Table 10: Values of the opposites of the log-likelihood functions for each penalty coefficient - the upper line corresponds to the ν coefficient.

Table 10 shows that likelihood values are not too damaged up to a penalization coefficient of 900. Recall that kernel 3 has a better likelihood (with no penalization) than kernel 2, this indicates that our data is better represented with a latent excitation than an instantaneous one. Moreover, kernel 3 has a better likelihood than kernel 2 even if it is a more sparing model (54 parameters versus 84 parameters).
The analysis of the penalization in terms of prediction will be assessed in the next section.

4.5 Predictions

4.5.1 Computation of the mean predicted number of attacks

In order to compare the different calibrations, we study the gap between the expectation of the calibrated process and the real number of attacks. In Appendix B, we detail the computation for the expectation of a multivariate Hawkes process, for the three types of kernels considered in Equation (10). Recall that we are concerned by the non-stationary framework as we have chosen a temporal drift for our baseline intensity. It extends previous results of [Bou16] for which computations are done for non-stationary univariate Hawkes processes with a wide class of kernels. We thus compute the average expected number of events on a given period $\mathbb{E}[N_t^{(i)} \mid \mathcal{F}_{t_0}]$, $t_0 < t$, for the calibrated set of parameters. The numerical results are given in the first columns of Table 11, for different values ν of the Lasso penalty coefficient. It is compared with the real number of attacks of the PRC database in the last column.

4.5.2 Results

Table 11 below shows that an overestimated number of attacks on one segment could be "compensated" by an underestimated number on another segment. To provide a more detailed estimation of the errors in prediction, segment by segment, Table 12 computes the sum (over the six segments) of the absolute differences between the expected number of attacks predicted and the real number.

Penalty coefficients ν	0	100	600	900	3000	6000	Real number
Kernel 1 (2016)	476.0	537.2	708.9	780.1	1782.0	1944.5	809
Kernel 1 (2017)	721.4	433.6	587.7	628.2	1387.0	1408.5	655
Kernel 2 (2016)	588.4	549.0	620.8	637.3	896.1	1190.7	809
Kernel 2 (2017)	592.4	548.4	534.5	530.7	689.2	967.8	655
Kernel 3 (2016)	634.1	597.0	583.7	649.9	1311.2	2368.1	809
Kernel 3 (2017)	671.1	665.3	514.8	722.1	3823.9	7057.4	655

Table 11: Sum of the mean predicted number of attacks over each segment

Penalty coefficients ν	0	100	600	900	3000	6000
Kernel 1 (2016)	337.7	280.7	277.8	283.6	973.0	1135.5
Kernel 1 (2017)	170.3	249.5	240.4	261.7	732.0	792.8
Kernel 2 (2016)	259.5	282.2	202.7	180.4	256.8	430.4
Kernel 2 (2017)	127.3	160.9	159.5	141.7	148.1	346.5
Kernel 3 (2016)	201.8	285.1	262.1	283.2	502.2	1559.1
Kernel 3 (2017)	165.7	183.3	254.3	172.7	3168.9	6402.4

Table 12: Sum of the absolute differences between the expected number of attacks predicted (in 2016 and 2017) and the real number, over the six segments.

Tables 11 and 12 show that a penalty can improve the predictive capacities of the model, to a certain extent, this is the case for example for kernels 1 and 2 for the year 2016. Kernel 2 seems to be better improved by a penalization, which is in line with the fact that it is the less parsimonious (84 parameters). We observe in our experiment that as a large penalty coefficient makes the model closer to a Poisson process (without auto excitation), it is compensated with a larger baseline intensity μ, which explains the worsening of the prediction when ν becomes too large. Kernel 2 is less sensible to this degradation since the $\beta_{i,j}$ can also compensate a too strong penalization on the $\alpha_{i,j}$. It also appears that the classical trade-off train/test is difficult here, since the best predictions of 2017 are not made by the best models of 2016.
The model with kernel 3 ($\phi_{i,j}(t) = \alpha_{i,j} t \exp(-\beta_i t)$) and no penalization seems to be a good compromise in terms of predictions over the two years. This model appears to be the most reliable since it is also the one with the best likelihood.

4.6 Predictions of the whole distribution of the number of attacks

We use the so-called thinning algorithm in order to simulate trajectories of Hawkes processes that are projections of a Hawkes process with past occurrences corresponding to the historical data. A detailed discussion on the thinning procedure for Hawkes processes is given in Section 4 of [Bou16]. The histograms of predictions based on 10 000 simulations are depicted in Figures 6 and 7. These distributions could be used to determine a 99.5% percentile at a one-year horizon in the context of a Solvency II internal model.

The distributions seem to capture the main trends, excepted for two cases in 2016 and one case in 2017. Besides, we note a tendency for the model to underestimate the number of attacks, this a feature of exponential kernels (that decrease very fast) that has already been pointed out by Bouchaud et al. [HBB13] in a financial framework for modeling order books. The projection for group **OTHER** (group 1) is particularly bad, due to the lack of structure of this "catch-all" group.

Finally, the model with kernel 3 seems to capture a significant part of the dynamics, with a reasonable number of parameters. Due to the heterogeneity of the PRC dataset, it is impossible to work globally on the whole dataset, that we have split in different groups. The choice of the different groups, that should be the more homogeneous as possible, is determinant in the prediction accuracy, as illustrated by the group **OTHER** that performs very badly. The advantage of our approach is that this joint mutually exciting model focuses on the whole joint distribution of numbers of attacks for each group, which is more accurate - but also more complex - than modeling marginal distributions, group per group. It thus allows to globally analyse the arrival of events, given some characteristics (type of the breached entity, type of breach, localization). A counterpart of this joint model is that a lower fitting accuracy for a given group may propagate to other groups.
Besides, part of the discrepancy observed in Figures 6 and 7 is also due to the variation of the underlying exposure, that inevitably impacts the number of attacks and the accuracy of the predictions. Although it is very difficult to assess the exposure for such a public dataset, an insurer could enrich this joint mutually exciting model using its own data on exposure, that could then be integrated within the baseline intensity of the Hawkes processes (for example taking a baseline intensity proportional to the exposure).

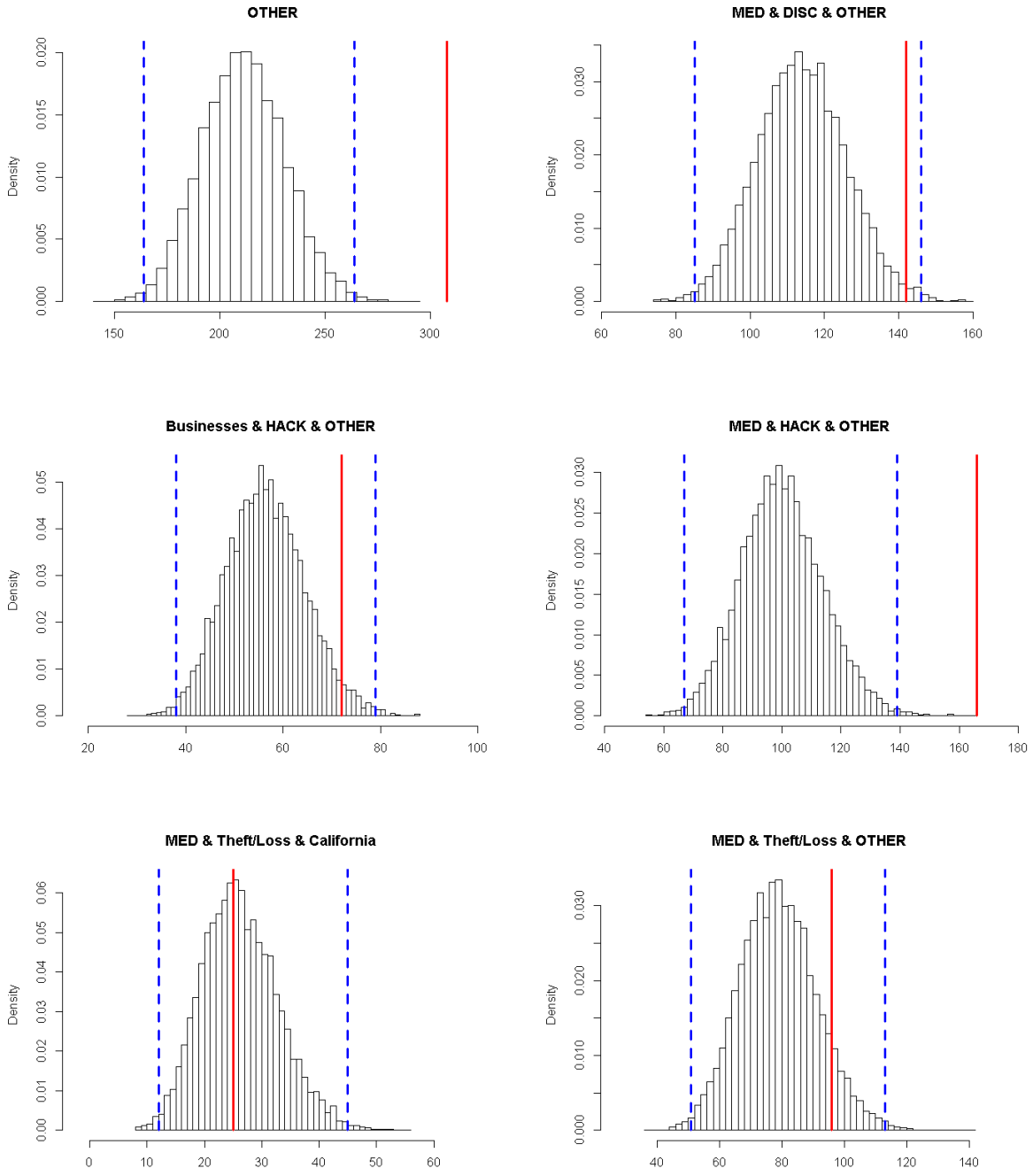

Figure 6: Distribution of the number of attacks predicted for 2016 with kernel 3 - In red the real number, in blue the 0.5% and 99.5% quantiles of the predicted distribution

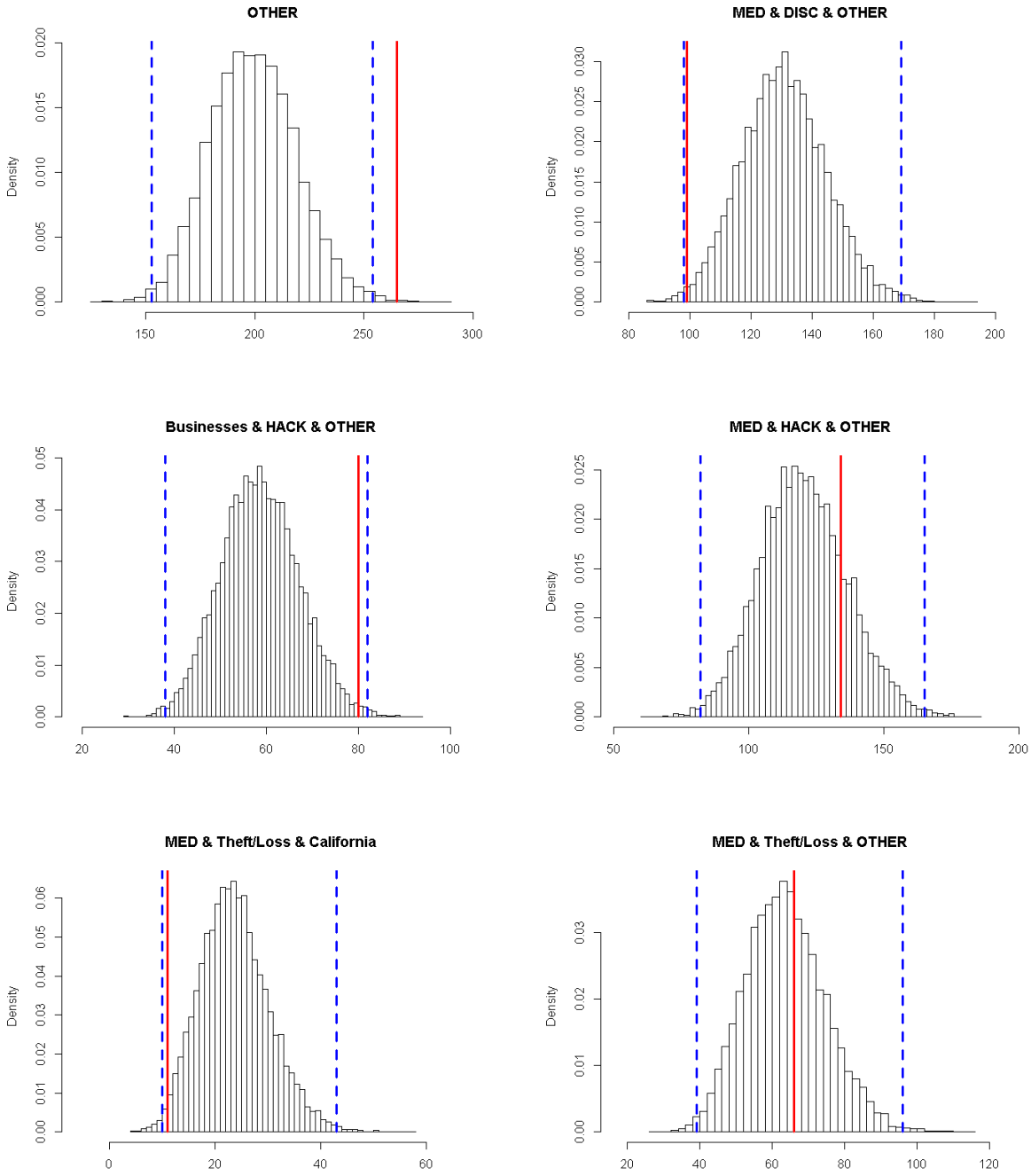

Figure 7: Distribution of the number of attacks predicted for 2017 with kernel 3 - In red the real number, in blue the 0.5% and 99.5% quantiles of the predicted distribution

5 Conclusion

This paper proposes a joint mutually exciting model to analyse and predict the arrivals of cyber events. It is achieved through multivariate Hawkes processes, that capture the clustering and autocorrelation of times of cyber events, depending on some characteristics. The analysis is conducted on the public dataset of Privacy Rights Clearinghouse, on which

different Hawkes kernels are calibrated. A kernel with a non-instantaneous excitation provides a better fit, compared to the standard exponential kernel. With such parsimonious parametric specifications, the model achieves reasonable forecasts which have been performed over a one-year horizon. Besides, the methodology can be easily extended to other types of cyber data.

This is a first step towards using advanced stochastic processes for the computation of a solvency capital for cyber-insurance or in the field of cyber risk covers pricing. An insurance pricing methodology requires the estimation of the frequency and severity of claims, while our methodology only focuses on the arrivals of cyber events.

First, our model should be completed with an exposure analysis in order to assess the pure frequency component of the risk. Unfortunately, since the PRC database is fed by various sources of information, the exposure - that is the number of entities exposed to risk within the PRC database - is difficult to handle. Therefore it is almost impossible to know from such data which part of the variation along time of the reported claims is caused by an evolution of the risk, and which part is caused by an instability in the way the data are collected. On the contrary, an insurance portfolio has a better knowledge of its exposition. But even for an insurance company with a cyber portfolio, the frequency would be poorly estimated if only based on internal historical data, since the number of reported claims would be too small to perform an accurate estimation. That is why analyzing public databases like PRC is important to improve the evaluation of the risk.

Second, for the severity component, the PRC dataset does not report directly the financial loss resulting from a data breach event, but still a severity indication is given through the volume of data breached. A projected financial loss can be then estimated from the number of records, accordingly to previous approaches such as in [EL17] or more recently in [FLT]. Let us note that Romanosky [Rom16] also studied the cost of data breaches using a private database gathering cyber events and associated losses. Again, an insurance company has a better knowledge of the effective losses of cyber events, but on a smaller dataset. To conclude, combining insurance portfolio data with external information - including public databases like PRC - seems to be essential to improve the evaluation of the risk.

A Repartition of attacks by state (PRC)

	Var1	Freq
1	Alabama	70
2	Alaska	22
3	Arizona	129
4	Arkansas	52
5	Beijing	1
6	Berlin	1
7	British Columbia	3
8	Buckinghamshire	2
9	California	1117
10	Cheshire	1
11	Colorado	122
12	Connecticut	109
13	Delaware	16
14	District Of Columbia	104
15	Dublin	1
16	Florida	386
17	Georgia	207
18	Grand Bahama	1
19	Guangdong	1
20	Hawaii	19
21	Idaho	18
22	Illinois	291
23	Indiana	170
24	Iowa	56
25	Kansas	49
26	Kentucky	95
27	London	2
28	Louisiana	50
29	Maine	25
30	Maryland	331
31	Massachusetts	200
32	Michigan	122
33	Minnesota	133
34	Mississippi	29
35	Missouri	131
36	Montana	25
37	Nebraska	33
38	Nevada	46
39	New Hampshire	30
40	New Jersey	123
41	New Mexico	42
42	New York	451
43	Noord Holland	1
44	North Carolina	146
45	North Dakota	10
46	Ohio	193
47	Oklahoma	54
48	Ontario	7
49	Oregon	101
50	Pennsylvania	217
51	Puerto Rico	31
52	Quebec	3
53	Rhode Island	31
54	South Carolina	61
55	South Dakota	11
56	Tennessee	130
57	Texas	489
58	Tokyo	1
59	UNKNSTATE	302
60	Utah	48
61	Vermont	27
62	Virginia	148
63	Washington	169
64	West Virginia	17
65	Wisconsin	84
66	Wyoming	12

B Computation of the expectation of the multivariate non-stationary Hawkes process

In this section, we derive closed-form formulas for the expectation of the Hawkes process under the three kernel specifications as given in Equation (10). The proof relies on deriving the dynamics of the underlying age-pyramid, as developed by [Bou16] to compute the distribution of non-stationary Hawkes processes for general kernel in the univariate case; we extend here the scope of application of such techniques to the multivariate Hawkes model for the kernels considered.

In the population representation, events are interpreted as arrivals or births of individuals in a population, while the Hawkes process measures the evolution of the total population size over time. Considering for example the univariate case ($d = 1$), immigrants arrive (with age zero) in the population according to a Poisson process with rate μ_t, then each immigrant with age a gives birth with rate $\phi(a)$; more generally, every individual with age a in the population gives birth with rate $\phi(a)$.
We introduce the random point measure $Z_t^{(i)}(\mathrm{d}a), i \in \{1, .., d\}$ defined as:

$$Z_t^{(i)}(\mathrm{d}a) = \int_{(0,t]} \delta_{t-s}(\mathrm{d}a)\mathrm{d}N_s^{(i)} = \sum_{n=1}^{N_t^{(i)}} \delta_{t-T_n^{(i)}}(\mathrm{d}a).$$

This measure allows to keep track of all ages in the population and can be used to integrate a function f, to do this we use the notation:

$$\langle Z_t^{(i)}, f\rangle = \int_{\mathbb{R}^+} f(a) Z_t^{(i)}(\mathrm{d}a) = \int_{(0,t]} f(t-s)\mathrm{d}N_s^{(i)}.$$

For instance it allows us to represent the Hawkes process itself with $N_t^{(i)} = \langle Z_t^{(i)}, 1\rangle$ or the intensity of the process $(N_t^{(i)})_{t\geq 0}$ with:

$$\lambda_t^{(i)} = \mu_t^{(i)} + \sum_{j=1}^{d} \langle Z_{t-}^{(j)}, \phi_{i,j}\rangle.$$

The computation of the expectation will make use of the following result, see Lemma 1 in [Bou16]:

Proposition 3. *For any differentiable function $f : \mathbb{R}_+ \to \mathbb{R}$ with derivative f':*

$$\langle Z_t^{(j)}, f\rangle = f(0)\langle Z_t^{(j)}, 1\rangle + \int_0^t \langle Z_s^{(j)}, f'\rangle \mathrm{d}s, \quad \textit{for any } j \in \{1, ..., d\}.$$

In this dynamics, the first term refers to the pure jump part of arrivals of individuals with age 0, whereas the second term of transport type illustrates the aging phenomenon: all ages are translated along the time axis.

In the following, we provide closed-form calculations for the expectation of the multivariate Hawkes process for the three kernel specifications studied in this paper:

$$\begin{aligned} &\text{Kernel 1: } \phi_{i,j}(t) = \alpha_{i,j} \exp(-\beta_i t) \\ &\text{Kernel 2: } \phi_{i,j}(t) = \alpha_{i,j} \exp(-\beta_{i,j} t) \\ &\text{Kernel 3: } \phi_{i,j}(t) = \alpha_{i,j} t \exp(-\beta_i t) \end{aligned} \tag{10}$$

Kernels 1 and 2 correspond to the exponential case and are studied in Section B.1, whereas kernel 3 is tackled separately in Section B.2.

B.1 Expectation with kernels 1 and 2 (exponential case)

Proposition 4. *Let us consider a d-variate Hawkes process $(N_t^{(1)})_{t\geq 0}, ..., (N_t^{(d)})_{t\geq 0}$ with exponential kernel, and let us denote $(\lambda_t^{(i)})_{t\geq 0}$ the intensity process of the process $(N_t^{(i)})_{t\geq 0}$:*

$$\lambda_t^{(i)} = \mu_t^{(i)} + \sum_{j=1}^{d} \int_{[0,t[} \phi_{i,j}(t-s) \mathrm{d}N_s^{(j)} = \mu_t^{(i)} + \sum_{j=1}^{d} \sum_{T_n^{(j)} < t} \alpha_{i,j} \exp(-\beta_{i,j}(t - T_n^{(j)}))$$

with $\mu^{(i)} : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$, $(\alpha_{i,j})_{1\leq i,j\leq d} \in \mathbb{R}_+^{d\times d}$, $(\beta_{i,j})_{1\leq i,j\leq d} \in \mathbb{R}_+^{d\times d}$.

We define the vector X_t and its expectation:

$$\begin{aligned} X_t &:= \Big(\langle Z_t^{(1)}, 1\rangle, ..., \langle Z_t^{(d)}, 1\rangle, \langle Z_t^{(1)}, \phi_{1,1}\rangle, ..., \langle Z_t^{(d)}, \phi_{1,d}\rangle, ..., \langle Z_t^{(1)}, \phi_{d,1}\rangle, ..., \langle Z_t^{(d)}, \phi_{d,d}\rangle\Big) \\ G(t) &:= \mathbb{E}\left[X_t\right] = \big(g_{0,1}(t), ..., g_{0,d}(t), g_{1,1}(t), ..., g_{1,d}(t), ..., g_{d,1}(t), ..., g_{d,d}(t)\big) \end{aligned}$$

Then the dynamics of X_t is Markovian and $G(t)$ can be expressed, for $t_0 < t$, as:

$$G(t) = G(t_0) \exp(A(t-t_0)) + \int_{t_0}^{t} \exp(A(t-s)) B(s) \mathrm{d}s$$

$B(t)$ is a vector with size $d(d+1)$ defined below:

- *for each $1 \leq i \leq d$, $[B(t)]_i = \mu_t^{(i)}$*
- *for each $d+1 \leq i \leq d^2+d$ such that $i = ad + b$ with integers a and b such that $1 \leq a \leq d$ and $1 \leq b \leq d$, $[B(t)]_i = \alpha_{a,b} \mu_t^{(b)}$*

A is a matrix with size $d(d+1) \times d(d+1)$ defined as follows:

- *For $1 \leq m \leq d$: $[A]_{m,n} = 1$ with $md + 1 \leq n \leq md + d$.*
- *For $d+1 \leq m \leq d^2+d$ such that $m = ad + b$, with integers a and b such that $1 \leq a \leq d$ and $1 \leq b \leq d$:*
  - *If $a \neq b$,*
    - *$[A]_{m,m} = -\beta_{a,b}$*
    - *$[A]_{m,n} = \alpha_{a,b}$ for any $bd + 1 \leq n \leq bd + d$*
  - *If $a = b$,*

* $[A]_{m,m} = \alpha_{a,a} - \beta_{a,a}$
* $[A]_{m,n} = \alpha_{a,a}$ *for any* $bd + 1 \leq n \leq bd + d$ *and* $n \neq m$

Note that non-specified components are zero.

Proof of Proposition 4 Let us prove in a first step that the dynamics of $(X_t)_{t\geq 0}$ is Markovian. In particular let us determine the dynamics of $\langle Z_t^{(j)}, \phi_{i,j}\rangle$ with $1 \leq i, j \leq d$, keeping in mind that $\phi_{i,j}$ represents the influence of $(N_t^{(j)})_{t\geq 0}$ on the intensity of $(N_t^{(i)})_{t\geq 0}$; by using Proposition 3,

$$\begin{aligned}\langle Z_t^{(j)}, \phi_{i,j}\rangle &= \alpha_{i,j} N_t^{(j)} + \int_0^t \langle Z_s^{(j)}, \phi'_{i,j}\rangle \mathrm{d}s \\ &= \alpha_{i,j} N_t^{(j)} - \beta_{i,j} \int_0^t \langle Z_s^{(j)}, \phi_{i,j}\rangle \mathrm{d}s\end{aligned}$$

By differentiation we obtain:

$$\mathrm{d}\langle Z_t^{(j)}, \phi_{i,j}\rangle = \alpha_{i,j} \mathrm{d}N_t^{(j)} - \beta_{i,j}\langle Z_t^{(j)}, \phi_{i,j}\rangle \mathrm{d}t \tag{11}$$

Therefore the dynamics of $(X_t)_{t\geq 0}$ is Markovian.

Let us find in a second step the form of G. Let us define $g_{i,j}(t) = \mathbb{E}\big[\langle Z_t^{(j)}, \phi_{i,j}\rangle\big]$ and let us take the expectation in (11) using the fact that $\mathbb{E}[\mathrm{d}N_t^{(j)}] = \mathbb{E}[\lambda_t^{(j)}]\mathrm{d}t$:

$$\begin{aligned}\mathrm{d}\mathbb{E}\big[\langle Z_t^{(j)}, \phi_{i,j}\rangle\big] &= \alpha_{i,j}\mathbb{E}\big[\lambda_t^{(j)}\big]\mathrm{d}t - \beta_{i,j}\mathbb{E}\big[\langle Z_t^{(j)}, \phi_{i,j}\rangle\big]\mathrm{d}t \\ &= \alpha_{i,j}\mu_t^{(j)}\mathrm{d}t + \alpha_{i,j}\sum_{k=1}^{d}\mathbb{E}\Big[\langle Z_{t-}^{(k)}, \phi_{j,k}\rangle\Big]\mathrm{d}t - \beta_{i,j} g_{i,j}(t)\mathrm{d}t\end{aligned}$$

By using the fact that Lebesgue measure charges no point, we obtain:

$$g'_{i,j}(t) = \alpha_{i,j}\mu_t^{(j)} + \alpha_{i,j}\sum_{k=1}^{d} g_{j,k}(t) - \beta_{i,j} g_{i,j}(t).$$

Now let us study the expectation of $N_t^{(k)} = \langle Z_t^{(k)}, 1\rangle$ for $1 \leq k \leq d$:

$$\mathrm{d}\mathbb{E}\big[\langle Z_t^{(k)}, 1\rangle\big] = \mathbb{E}\big[\lambda_t^{(k)}\big]\mathrm{d}t = \mu_t^{(k)}\mathrm{d}t + \sum_{l=1}^{d}\mathbb{E}\big[\langle Z_{t-}^{(l)}, \phi_{k,l}\rangle\big]\mathrm{d}t$$

Let us set $g_{0,k}(t) = \mathbb{E}\big[\langle Z_t^{(k)}, 1\rangle\big]$, then:

$$g'_{0,k}(t) = \mu_t^{(k)} + \sum_{l=1}^{d} g_{k,l}(t).$$

Therefore, the system of differential equations can be conveniently rewritten as:

$$G'(t) = AG(t) + B(t),$$

where A and $B(t)$ are specified in Proposition 4; this finally proves that G is of the form:

$$G(t) = G(t_0)\exp(A(t - t_0)) + \int_{t_0}^{t} \exp(A(t - s))B(s)\mathrm{d}s.$$

Remark 4. *The proof to obtain expectations conditional on information up to any time is similar.*

B.2 Expectation with kernel 3

The next proposition provides the computation for a d-variate Hawkes process with kernel $\phi_{i,j}(a) = \alpha_{i,j} a \exp(-\beta_{i,j} a)$, which is little bit more general than kernel 3 that corresponds to $\beta_{i,j} = \beta_i$ for all j.

Proposition 5. *Let us consider a d-variate Hawkes process with kernel $\phi_{i,j}(a) = \alpha_{i,j} a \exp(-\beta_{i,j} a)$. Let us still denote by $(\lambda_t^{(i)})_{t\geq 0}$ the intensity process of the process $(N_t^{(i)})_{t\geq 0}, i \in \{1, ..., d\}$:*

$$\lambda_t^{(i)} = \mu_t^{(i)} + \sum_{j=1}^{d} \int_{[0,t[} \phi_{i,j}(t-s) \mathrm{d}N_s^{(j)} = \mu_t^{(i)} + \sum_{j=1}^{d} \sum_{T_n^{(j)} < t} \alpha_{i,j}(t - T_n^{(j)}) \exp(-\beta_{i,j}(t - T_n^{(j)}))$$

with $\mu^{(i)} : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$, $(\alpha_{i,j})_{1\leq i,j\leq d} \in \mathbb{R}_+^{d\times d}$, $(\beta_{i,j})_{1\leq i,j\leq d} \in \mathbb{R}_+^{d\times d}$.

Now let us consider the following vector Y_t and its expectation H_t:

$$\begin{aligned} Y_t &:= \left((\langle Z_t^{(i)}, 1\rangle)_{1\leq i\leq d}, (\langle Z_t^{(j)}, \phi_{i,j}\rangle)_{1\leq i,j\leq d}, (\langle Z_t^{(j)}, \phi_{i,j}^{(e)}\rangle)_{1\leq i,j\leq d} \right) \\ H(t) &:= \mathbb{E}\left[Y_t\right] \end{aligned}$$

with same ordering as for X in Proposition 4, and using the notation $\phi_{i,j}^{(e)}(a) = \alpha_{i,j} \exp(-\beta_{i,j} a)$ for the exponential kernel. Then the dynamics of Y_t is Markovian and in particular $H(t)$ can be expressed for $t > t_0$ as:

$$H(t) = H(t_0) \exp(C(t - t_0)) + \int_{t_0}^{t} \exp(C(t-s)) D(s) \mathrm{d}s.$$

$D(t)$ is a vector with size $d(2d+1)$ defined below:

- *for each $1 \leq i \leq d$, $[D(t)]_i = \mu_t^{(i)}$*
- *for each $d^2 + d + 1 \leq i \leq 2d^2 + d$ such that $i = d^2 + ad + b$ with integers a and b such that $1 \leq a \leq d$ and $1 \leq b \leq d$, $[D(t)]_i = \alpha_{a,b} \mu_t^{(b)}$.*

C is a matrix with size $d(2d+1) \times d(2d+1)$ defined as follows:

- *For $1 \leq m \leq d$: $[C]_{m,n} = 1$ with $md + 1 \leq n \leq md + d$*
- *For $d + 1 \leq m \leq d^2 + d$ such that $m = ad + b$, with integers a and b such that $1 \leq a \leq d$ and $1 \leq b \leq d$:*
  - *$[C]_{m,m} = -\beta_{a,b}$*
  - *$[C]_{m,m+d^2} = 1$*
- *For $d^2 + d + 1 \leq m \leq 2d^2 + d$ such that $m = ad + b$, with integers a and b such that $1 \leq a \leq d$ and $1 \leq b \leq d$:*
  - *$[C]_{m,m} = -\beta_{a,b}$*
  - *$[C]_{m,n} = \alpha_{a,b}$ with $bd + 1 \leq n \leq bd + d$*

Note that non-specified components are zero.

Proof of Proposition 5. Let us first notice that $\phi'_{i,j}(a) = \phi^{(e)}_{i,j}(a) - \beta_{i,j}\phi_{i,j}(a)$ where we redefine the exponential kernel as:

$$\phi^{(e)}_{i,j}(a) = \alpha_{i,j} \exp(-\beta_{i,j} a).$$

The dynamics of $(\langle Z^{(i)}_t, 1\rangle)_{1\leq i\leq d}$ is the same as in the proof for the exponential case and we get

$$g'_{0,k}(t) = \mu^{(k)}_t + \sum_{l=1}^{d} g_{k,l}(t), \tag{12}$$

where we set $g_{0,k}(t) = \mathbb{E}\big[\langle Z^{(k)}_t, 1\rangle\big]$ and $g_{i,j}(t) = \mathbb{E}\big[\langle Z^{(j)}_t, \phi_{i,j}\rangle\big]$.
Let us now study the dynamics of $\langle Z^{(j)}_t, \phi_{i,j}\rangle$ for $1 \leq i, j \leq d$:

$$\begin{aligned}
\mathrm{d}\langle Z^{(j)}_t, \phi_{i,j}\rangle &= \phi_{i,j}(0)\mathrm{d}N^{(j)}_t + \langle Z^{(j)}_t, \phi'_{i,j}\rangle \mathrm{d}t, \\
&= \langle Z^{(j)}_t, \phi^{(e)}_{i,j}\rangle \mathrm{d}t - \beta_{i,j}\langle Z^{(j)}_t, \phi_{i,j}\rangle \mathrm{d}t.
\end{aligned}$$

By taking expectation we then get:

$$g'_{i,j}(t) = g^{(e)}_{i,j}(t) - \beta_{i,j} g_{i,j}(t), \tag{13}$$

where we set $g^{(e)}_{i,j}(t) = \mathbb{E}\big[\langle Z^{(j)}_t, \phi^{(e)}_{i,j}\rangle\big]$. Finally, let us study the dynamics of $\langle Z^{(j)}_t, \phi^{(e)}_{i,j}\rangle$:

$$\mathrm{d}\langle Z^{(j)}_t, \phi^{(e)}_{i,j}\rangle = \alpha_{i,j}\mathrm{d}N^{(j)}_t - \beta_{i,j}\langle Z^{(j)}_t, \phi^{(e)}_{i,j}\rangle \mathrm{d}t.$$

By taking expectation we get:

$$\begin{aligned}
\mathrm{d}\mathbb{E}\big[\langle Z^{(j)}_t, \phi^{(e)}_{i,j}\rangle\big] &= \alpha_{i,j}\mathbb{E}\big[\lambda^{(j)}_t\big]\mathrm{d}t - \beta_{i,j}\mathbb{E}\big[\langle Z^{(j)}_t, \phi^{(e)}_{i,j}\rangle\big]\mathrm{d}t \\
&= \alpha_{i,j}\mu^{(j)}_t \mathrm{d}t + \alpha_{i,j}\sum_{k=1}^{d} \mathbb{E}\Big[\langle Z^{(k)}_{t-}, \phi_{j,k}\rangle\Big]\mathrm{d}t - \beta_{i,j} g^{(e)}_{i,j}(t)\mathrm{d}t
\end{aligned}$$

leading to the differential equation:

$$g^{(e)\prime}_{i,j}(t) = \alpha_{i,j}\mu^{(j)}_t + \alpha_{i,j}\sum_{k=1}^{d} g_{j,k}(t) - \beta_{i,j} g^{(e)}_{i,j}(t) \tag{14}$$

From Equations (12), (13) and (14), the ordinary differential equation for $H(t) = \mathbb{E}[Y_t]$ can therefore be rewritten as

$$H'(t) = CH(t) + D(t),$$

where C and $D(t)$ are specified in Proposition 2, which finally proves that H can be computed as:

$$H(t) = H(t_0)\exp(C(t-t_0)) + \int_{t_0}^{t} \exp(C(t-s))D(s)\mathrm{d}s.$$

References

[BEW15] Christian Biener, Martin Eling, and Jan Hendrik Wirfs. Insurability of cyber risk: An empirical analysis. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 40(1):131–158, 2015. 2

[BGI+17] Adrian Baldwin, Iffat Gheyas, Christos Ioannidis, David Pym, and Julian Williams. Contagion in cyber security attacks. *Journal of the Operational Research Society*, 68(07):780–791, 2017. 3

[BK06] Rainer Böhme and Gaurav Kataria. Models and measures for correlation in cyber-insurance. In *WEIS*, 2006. 2, 3

[BM96] Pierre Brémaud and Laurent Massoulié. Stability of nonlinear hawkes processes. *The Annals of Probability*, pages 1563–1588, 1996. 8

[BM02] Pierre Brémaud and Laurent Massoulié. Power spectra of general shot noises and hawkes point processes with a random excitation. *Advances in Applied Probability*, 34(1):205–222, 2002. 8

[BMM15] Emmanuel Bacry, Iacopo Mastromatteo, and Jean-Francois Muzy. Hawkes processes in finance. *Market Microstructure and Liquidity*, 1(01):1550005, 2015. 3, 10, 12

[BMS16] Flavia Barsotti, Xavier Milhaud, and Yahia Salhi. Lapse risk in life insurance: Correlation and contagion effects among policyholdersâ behaviors. *Insurance: Mathematics and Economics*, 71:317–331, 2016. 3

[Bou16] Alexandre Boumezoued. Population viewpoint on Hawkes processes. *Advances in Applied Probability*, 48(2):463–480, 2016. 3, 10, 20, 21, 27

[BS+10] Rainer Böhme, Galina Schwartz, et al. Modeling cyber-insurance: Towards a unifying framework. In *WEIS*, 2010. 2

[DVJ07] Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes: volume II: general theory and structure*. Springer Science & Business Media, 2007. 4, 8

[EGG10] Eymen Errais, Kay Giesecke, and Lisa R Goldberg. Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics*, 1(1):642–665, 2010. 3

[EHF16] Benjamin Edwards, Steven Hofmeyr, and Stephanie Forrest. Hype and heavy tails: A closer look at data breaches. *Journal of Cybersecurity*, 2(1):3–14, 2016. 2

[EL17] Martin Eling and Nicola Loperfido. Data breaches: Goodness of fit, pricing, and risk measurement. *Insurance: mathematics and economics*, 75:126–136, 2017. 2, 25

[ELL11] Paul Embrechts, Thomas Liniger, and Lu Lin. Multivariate hawkes processes: an application to financial data. *Journal of Applied Probability*, 48(A):367–378, 2011. 3

[ES16] Martin Eling and Werner Schnell. What do we know about cyber risk and cyber risk insurance? *The Journal of Risk Finance*, 2016. 2

[FLT] Sébastien Farkas, Olivier Lopez, and Maud Thomas. Cyber claim analysis through generalized pareto regression trees with applications to insurance. *https://hal.archives-ouvertes.fr/hal-02118080v2/document*. 2, 5, 25

[FWW18] Matthias A Fahrenwaldt, Stefan Weber, and Kerstin Weske. Pricing of cyber insurance contracts in a network model. *ASTIN Bulletin: The Journal of the IAA*, 48(3):1175–1218, 2018. 2

[GK08] Nancy L Garcia and Thomas G Kurtz. Spatial point processes and the projection method. In *In and Out of Equilibrium 2*, pages 271–298. Springer, 2008. 15

[Hai16] Donatien Hainaut. A bivariate hawkes process for interest rate modeling. *Economic Modelling*, 57:180–196, 2016. 3

[Haw71] Alan G Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971. 3

[HBB13] Stephen J Hardiman, Nicolas Bercot, and Jean-Philippe Bouchaud. Critical reflexivity in financial markets: a hawkes process analysis. *The European Physical Journal B*, 86(10):442, 2013. 22

[HH11] Hemantha Herath and Tejaswini Herath. Copula-based actuarial model for pricing cyber-insurance policies. *Insurance markets and companies: analyses and actuarial computations*, 2(1):7–20, 2011. 2, 3

[HZO+13] John Homer, Su Zhang, Xinming Ou, David Schmidt, Yanhui Du, S Raj Rajagopalan, and Anoop Singhal. Aggregating vulnerability metrics in enterprise networks using attack graphs. *Journal of Computer Security*, 21(4):561–597, 2013. 2

[Jac14] Jay Jacobs. Analyzing ponemon cost of data breach. *Data Driven Security*, 11, 2014. 5

[Jai15] Thibault Jaisson. *Market activity and price impact throughout time scales*. PhD thesis, Ecole Polytechnique, 2015. 10

[JBG11] Benjamin Johnson, Rainer Böhme, and Jens Grossklags. Security games with market insurance. In *International Conference on Decision and Game Theory for Security*, pages 117–130. Springer, 2011. 2

[JD13] Jiwook Jang and Angelos Dassios. A bivariate shot noise self-exciting process for insurance. *Insurance: Mathematics and Economics*, 53(3):524–532, 2013. 3

[MMN+17] Angelica Marotta, Fabio Martinelli, Stefano Nanni, Albina Orlando, and Artsiom Yautsiukhin. Cyber-insurance survey. *Computer Science Review*, 24:35–61, 2017. 2

[MS10] Thomas Maillart and Didier Sornette. Heavy-tailed distribution of cyber-risks. *The European Physical Journal B*, 75(3):357–364, 2010. 2

[NJWS10] Steven Noel, Sushil Jajodia, Lingyu Wang, and Anoop Singhal. Measuring security risk of networks using attack graphs. *International Journal of Next-Generation Computing*, 1(1):135–147, 2010. 2

[Oza79] Tohru Ozaki. Maximum likelihood estimation of hawkes' self-exciting point processes. *Annals of the Institute of Statistical Mathematics*, 31(1):145–155, 1979. 10

[PXXH17] Chen Peng, Maochao Xu, Shouhuai Xu, and Taizhong Hu. Modeling and predicting extreme cyber attack rates via marked point processes. *Journal of Applied Statistics*, 44(14):2534–2563, 2017. 2, 3

[RBC+16] Markus Riek, Rainer Böhme, Michael Ciere, Carlos Ganan, and Michel van Eeten. Estimating the costs of consumer-facing cybercrime: A tailored instrument and representative data for six eu countries. In *Workshop on the Economics of Information Security (WEIS), University of California at Berkeley*, 2016. 2

[RICVR+19] David Rios Insua, Aitor Couce-Vieira, Jose A Rubio, Wolter Pieters, Katsiaryna Labunets, and Daniel G. Rasines. An adversarial risk analysis framework for cybersecurity. *Risk Analysis*, 2019. 2

[RLMX17] Marian-Andrei Rizoiu, Young Lee, Swapnil Mishra, and Lexing Xie. A tutorial on hawkes processes for events in social media. *arXiv preprint arXiv:1708.06401*, 2017. 3

[Rom16] Sasha Romanosky. Examining the costs and causes of cyber incidents. *Journal of Cybersecurity*, 2(2):121–135, 2016. 25

[SARH11] Dinesh Kumar Saini, Imran Azad, Nitin B Raut, and Lingaraj A Hadimani. Utility implementation for cyber risk insurance modeling. In *Proceedings of the World Congress on Engineering*, volume 1, 2011. 2

[ST10] Gabriele Stabile and Giovanni Luca Torrisi. Risk processes with non-stationary hawkes claims arrivals. *Methodology and Computing in Applied Probability*, 12(3):415–429, 2010. 3

[Wan19] Shaun S Wang. Integrated framework for information security investment and cyber insurance. *Pacific-Basin Finance Journal*, 57:101173, 2019. 2

[XH19] Maochao Xu and Lei Hua. Cybersecurity insurance: Modeling and pricing. *North American Actuarial Journal*, 23(2):220–249, 2019. 3