A stochastic SIR model for cyber contagion: application to granular growth of firms and to insurance portfolio

Caroline Hillairet and Olivier Lopez
CREST, UMR CNRS 9194, Ensae Paris, Avenue Henry Le Chatelier, 91120 Palaiseau, France

Lionel Sopgoui
CREST, UMR CNRS 9194, Ensae Paris, Avenue Henry Le Chatelier, 91120 Palaiseau, France and Institut Louis Bachelier

(Dated: January 13, 2026)

The aim of this work is to evaluate the impact of a cyber episode on firm's financial health and on cyber insurance portfolio. We stand on key empirical facts in economics and in cybersecurity. In economics, firm size and growth rate distributions are non-Gaussian and exhibit heavy tails. In cybersecurity, contagion dynamics depends heavily on firm's size and environmental conditions. Therefore, taking into account these behaviors, we propose a stochastic multigroup SIR model integrating a granular model of firm growth. This allows us to define the financial impact of attacks on firm' revenue and insurance portfolio. In our model, the revenue of a subunit incurs a random loss upon cyber attack. The arrival time and the duration of this cyber attack are defined using a sum of Cox process and a Bernoulli random variable. The Cox process whose intensity is the force of the cyber episode represents the external contagion while the Bernoulli random variable represents the contagion from an infected sister-subsidiary. We provide theoretical results that include existence, uniqueness, stability of the SIR model, and additional asymptotic properties. By governing the SIR parameters with CIR dynamics, we incorporate environmental variability. This stochastic approach enables scenario generation and the calculation of the aggregate exceedance probability – a metric used in catastrophe modeling that gives an insurer immediate feedback on the financial nature of an event. We apply this framework to the Lockbit ransomware attacks between May and July 2024. For a portfolio of 2,929 firms located in Île-de-France, our model predicts that the insurer will have to compensate, with a 50% probability, up to 2 days of revenue for a 100-day cyber incident.

Keywords: Stochastic modeling; cyber risk; granular model of firm growth; Aggregate exceedance probability; stochastic stress scenarios; SIR model

I. INTRODUCTION

Cybersecurity risk is, along with climate change and geopolitical instability, on top of the list of emerging concerns for experts and the general population (see [3]). Cybersecurity risk entails malicious or accidental events that can compromise the confidentiality, availability, or integrity of data or IT services. [33] notes that global cyber attacks increased by 38% between 2021 and 2022. Malwares, service-provider outages, or data breaches are among the types of attacks. Ransomware – a category of malware that prevents or restricts users from accessing their system or exfiltrates sensitive data – with lower barriers to entry, has become one of the most significant threats. [19] reports a growth of 76.8% between 2022 and 2023, and 8.72% between 2023 and 2024 in ransomware attacks. This corresponds to 4,848 ransomware attacks posted by up to 88 tracked ransomware groups. Victim organizations often incur the costs of extortion ([3] mentions an increase of $1.1 billion paid to hackers in 2023), remediation expenses, legal fees, business interruption costs, and reputation costs. That could consequently represent an increasing part in cyber insurance claims. Therefore, there is a need – not only for businesses but also for financial institutions such as insurance companies, and even for governments – to assess the economic and financial impact of potential future attacks. Precisely, we are interested in evaluating the impact of a cyber episode on firm's financial health and on cyber insurance portfolio.

Cyber risk modeling in finance and insurance, although relatively new, incorporates different approaches and models. [14] and [17] use traditional severity or frequency approaches. [13] investigate trends in data breaches using Bayesian Generalized Linear models. An emerging approach consists in using Hawkes processes such as [7] and [8] which introduce multivariate Hawkes processes to capture self-excitation and interactions of data-breaches, as well as impacts of vulnerabilities, to predict the frequency of cyber attacks. Furthermore, biological epidemics and contagious cyber attacks share alike mechanisms, and key parameters (susceptible population, infection rate, recovery rate, peak, prevalence, network/cluster structure). Therefore, a recent class of approaches draws inspiration from the literature on compartmental epidemiological models. They are particularly suitable for designing a contagious cyber event that behaves as an epidemic such as Covid-19. With this in mind, [22] adapts an epidemiological model for a cyber event such as a Wannacry-type scenario. They then apply it on an insurance portfolio to study the impact of reaction and remediation measures. This is extended in [23] to take into account the network structure of the population of firms. The present work draws inspiration from [22] and [23] in order to design a cyber-episode. Both use a particular example of a model widely used in epidemiology called SIR model (see [9, 12, 26]). In their simplest form, this model consists in dividing the population of firms/policyholders into 3 groups. S for susceptible – the exposed firms, that are not yet infected but are vulnerable if exposed to an infectious firm. I for infected – the companies affected by the malware and which can transmit. And R for removed – the cured, secured or patched systems. The relation between these groups and the dynamics of the epidemic are built thanks to a system of ordinary differential equations (ODEs).

Given that we are in the early stage of the application of epidemiological models to cyber insurance, there are several features that are not yet taken into account. In particular, our aim is to incorporate the following empirical stylized facts. The first one concerns the firm size and firm structure in economics: [4] shows that the firm size distribution feature heavy tails (especially Zipf's law) while [32] remarks that the growth rate distribution of the firm size is non-Gaussian and also features heavy tails. The second feature is that in cybersecurity, the dynamics of a cyber-episode depends on the firm size: [5, 27] demonstrate that larger firms are more often targeted by hackers in cyber-attack. Moreover, the contagion parameters can be subject to environmental variability (see [15]). For example, given a firm organized in subunits, the infection rate will vary if the contamination comes from inside or outside the corporation.

In order to take these features into account, we propose to model a cyber-attack episode by a stochastic multigroup SIR model and to model firm's size by a granular growth model. Precisely, we will extend the epidemiological model developed in [12] to study the effect of household size and size distribution on the dynamics of a disease. However, instead of transforming the system of ODEs of the deterministic SIR into a system of stochastic differential equations (SDE) as usually done, we adapt a more intuitive modeling approach by introducing stochasticity into the parameters that will be Cox-Ingersoll-Ross (CIR) processes, as in [2]. For the dynamics of firm's revenue, we adapt the granular model of firm growth of [29] that helps us to consider each firm as an ensemble of subunits with varying degrees of independence. By extending that in continuous time, we will state that the subunit's instantaneous revenue experiences a loss if she is attacked. The arrival of attacks will follow the sum of Cox process and Bernoulli random variable. The Cox process, whose intensity is the force of infection (depending on the dynamics of the epidemic) expresses the external contagion. The Bernoulli random variable represents the internal/secondary contagion from a sister subsidiary. This mix process will be used to define both the first arrival time and the duration of contagion inside a subunit. Afterward, we define the instantaneous cost of the cyber event for the insurance company as the difference between the revenue with and without the attack. At the portfolio level, we analyze the impact thanks to the aggregate exceedance probability (AEP) – the probability that the sum of losses during a certain period exceeds a given level.

The remainder of this paper proceeds as follows. In Section II, we introduce the jump diffusion model used to determine the financial health of a firm subject to cyber attacks and the aggregate exceedance probability as a risk measure on an insurance portfolio. Section III will be dedicated to the description of the cyber attack contagion model and some analytical results that include the existence and uniqueness of a non-negative solution of the stochastic multi-group SIR model. In Section IV, the impact of the contagious cyber episode is translated on the jump diffusion firms' revenues framework through the probabilities of internal and external transmission. Finally, in Section V, we realize calibration, describe the data, perform simulations, and discuss the results. Precisely, the contagion model is evaluated using the ransomware attack episode attributed to the Lockbit group between May and July 2024. The impact is then assessed on a portfolio of 2,929 firms located in the Île-de-France region.

II. FIRM VALUE AND INSURANCE PORTFOLIO DYNAMICS UNDER CYBERATTACKS

In this section, we extend the granular model of firm growth of [21, 29, 35] in a continuous time setting and in the context of cyber risk. This model has been introduced to fit some empirical facts, in particular the ones cited by [3, 32]: both the firm size distribution and the growth rate distribution are non-Gaussian and feature heavy tails.

Throughout this work we consider a filtered probability space $(\Omega, \mathbb{G}, (\mathcal{G}_t)_{t\geq 0}, \mathbb{P})$, where the filtration $(\mathcal{F}_t)_{t\geq 0}$ satisfies the usual conditions (i.e., it is right-continuous and complete). In the same spirit as [29], within a given insurance portfolio of $H \in \mathbb{N}^*$ firms characterized by exogenous production, we consider that firms are composed of a number of independent subunits, such as departments or production units, which operate within separate sub-markets and which are not necessarily of equal size. We have the following assumption.

Assumption II.1. The revenue of firm $i \in \{1, \dots, H\}$ at time $t \geq 0$ is denoted by $Z_{i,t}$ and satisfies

$$Z_{i,t} := \sum_{j=1}^{K_i} z_{ij,t}, \tag{II.1}$$

where $K_i \in \mathbb{N}^*$ represents the number of subunits and $z_{ij,t}$, $j = 1, \dots, K_i$, denotes the respective revenue.

We will note

$$\mathcal{I} := \{(i,j) \mid i \in \{1, \dots, H\} \text{ and } j \in \{1, \dots, K_i\}\}. \tag{II.2}$$

The firm's size is the number of its subunits, and the revenue is an indicator of firms' and subunits' financial health. Instead of revenue, we could consider its workforce, profits, etc. In addition, we could view the economy as a supra-firm subdivided in sectors. In the same way, each sector can be seen as a supra-firm subdivided in companies. This point of view could be interesting for calibration because public data are usually more easily available at economy/sectoral level than at firm's level.

The firm's size K_i, which is an integer, is a random variable assumed to be distributed according to Zipf's law (the discrete version of Pareto distribution) on $\{1, \dots, K\}$, $K \in \mathbb{N}^*$. For $k = 1, \cdots, K$

$$\mathbb{P}(K_i = k) = q\frac{\mathfrak{a}k^{-(1+\mathfrak{a})}}{1 - K^{-\mathfrak{a}}}; \quad 1 < \mathfrak{a} < 2. \tag{II.3}$$

The above assumption is motivated by the literature (see [29, 35]) and is easily verified on the data (see Figure 15a and Figure 15b where $\mathfrak{a} = 1.76$ and $q = 0.784$).

A. The impact of cyberattacks on firms' revenue

When a cyber attack succeeds, it causes financial loss to the company. This loss can result from paying a ransom, business disruption, repairing equipment, reputation costs, etc. Let $1 \leq i \leq H$ and $1 \leq j \leq K_i$. In the absence of a cyber attack, the revenue z_{ij} of subunit j of firm i is modeled as a Geometric Brownian motion, while in the presence of a cyber attack, it is modeled as a jump-diffusive process. We introduce the following assumption.

Assumption II.2. For $(i,j) \in \mathcal{I}$,

1. Without the cyber-event, the time evolution of each subunit i's revenue, $\overline{z}_{ij}$, is characterized by the diffusion process

$$\frac{d\overline{z}_{ij,t}}{\overline{z}_{ij,t}} = \mu_{ij}\mathrm{d}t + \sigma_{ij}\mathrm{d}B_{ij,t}, \quad \overline{z}_{ij,0} = z_{ij,0}, \tag{II.4}$$

where B_{ij} is a Brownian motion. For a given firm i, $1 \leq i \leq H$, the Brownian motions $(B_{ij})_{1\leq j\leq K_i}$ driving the revenues of the subunits are correlated with correlation coefficient $\rho_i \in \mathbb{R}$. But for different firms $i_1 \neq i_2$, $B_{i_1j_1}$ and $B_{i_2j_2}$ are assumed to be independent. Moreover, $\sigma_{ij}, z_{ij,0} > 0$, $\mu_{ij} \in \mathbb{R}^*$. The solution of (II.4) is

$$\overline{z}_{ij,t} = z_{ij,0}\exp\left(\left(\mu_{ij} - \frac{1}{2}\sigma_{ij}^2\right)t + \sigma_{ij}B_{ij,t}\right).$$

2. The random time τ_{ij} of the arrival of a cyber attack affecting the subunit j of firm i, is defined as

$$\tau_{ij} = \inf\{t \geq 0 | N_{ij,t} \geq 1\}, \tag{II.5}$$

where $N_{ij,t}$ is a point process, whose intensity will be defined later.

3. δ_{ij} is a random time representing the duration of an infection.

4. We introduce the process π_{ij}, taking values in $[0,1]$, such that at each time t during the infection, the firm loses a random fraction $\pi_{ij,t}$ of its revenue.

Therefore, the firm's revenue $z_{ij,t}$ evolves as follows

$$z_{ij,t} = \begin{cases} \overline{z}_{ij,t} & \text{if } t < \tau_{ij} \\ (1 - \pi_{ij,t})\,\overline{z}_{ij,t} & \text{if } \tau_{ij} \leq t < \tau_{ij} + \delta_{ij} \\ \overline{z}_{ij,t} & \text{if } t \geq \tau_{ij} + \delta_{ij} \end{cases} \tag{II.6}$$

Moreover, $(\pi_{ij})_{(i,j)\in\mathcal{I}}$, $(N_{ij})_{(i,j)\in\mathcal{I}}$, and $(B_{ij})_{(i,j)\in\mathcal{I}}$ are independent.

Let $(i,j) \in \mathcal{I}$. In the above assumption, the constant parameters σ_{ij} (resp. μ_{ij}) define the order of magnitude (resp. the drift) of the growth fluctuations at the level of a subunit. Regarding the point process N_{ij}, the event $\{N_{ij,t} \geq 1\}$ corresponds to a successful breach caused by an attack; otherwise, $\{N_{ij,t} = 0\}$, means no attack or the potential attack is blocked and no loss occurs at time t, $t \geq 0$. The K_i-dimensional process $(N_{ij})_{1\leq j\leq K_i}$ is potentially correlated. Moreover, the property on the Brownian motion $(B_{ij})_{ij}$ means that inside the same firm, the subunits' revenues are correlated but different firms' revenues are not.

Finally, the process π_{ij} represents the severity (i.e. the instantaneous random monetary loss) of the attack. This loss may be potentially covered by an insurance contract. It is a $[0,1]$-value process because a cyber attack is detrimental to the subunit, but its losses remain capped at its potential earnings. The time-behavior of the loss process may depend on the dynamics of the epidemic, on the firm's investments in cybersecurity, and also on cyber insurance premium: the severity might begin high when infection starts, and progressively drop to zero (full recovery). Or, it could start slowly, peak, and then decline. Having no explicit empirical input on this feature, we limit ourselves in this work to the case where the severity does not depend on time, so π_{ij} is a random variable instead of a stochastic process. The sequence $(\pi_{ij})_{ij}$ are assumed independent and identically distributed with common distribution f_π having support in $[0,1]$, mean $\pi_\star$ and standard deviation $\sigma_\star^2$.

Example II.3 (Beta distribution)**.** For the numerical analysis, we will take $\pi_{ij} \sim B(\alpha^\pi, \beta^\pi)$ whose density probability function is $f(x;\alpha^\pi,\beta^\pi) = \frac{x^{\alpha^\pi-1}(1-x)^{\beta^\pi-1}}{B(\alpha^\pi,\beta^\pi)}$ for $x \in [0,1]$ where given the Gamma function Γ, $B(\alpha^\pi,\beta^\pi) = \frac{\Gamma(\alpha^\pi)\Gamma(\beta^\pi)}{\Gamma(\alpha^\pi+\beta^\pi)}$.

If we are dealing with the whole portfolio (or a whole sector or an entire economy), we can also derive the total output $\mathbf{O}$ so that for all $t \geq 0$,

$$\mathbf{O}_t := \sum_{i=1}^{H} Z_{i,t}, \tag{II.7}$$

where the individual firm output (revenue) is given by

$$Z_{i,t} = \sum_{j=1}^{K_i} \left(1 - \pi_{ij}\mathbf{1}_{\tau_{ij}\leq t<\tau_{ij}+\delta_{ij}}\right) \overline{z}_{ij,t}.$$

We also note

$$\mathbb{F}^B := (\mathcal{F}_t^B)_{t\geq 0}, \tag{II.8}$$

the natural filtration of $(B_{ij})_{(i,j)\in\mathcal{I}}$.

B. The impact of cyberattack on insurance portfolio

Based on the description of the attack at the policyholder level, the insurance company is interested in aggregating these risks. We focus on the total claims of the cyber event for the insurance company over a period. For a policyholder experiencing a cyber attack, each day spent infected results in financial losses: their revenue decreases. The insurance company potentially compensates a part (or all) of these losses. Consider the policyholder $i \in \{1,\ldots,H\}$. For all $j \in \{1,\ldots,K_i\}$, recall the dynamics of the subunit's revenue without and with the cyber risk $\overline{z}_{ij}$ and z_{ij} respectively in (II.4) and in (II.6). We thus write the following definition.

Definition II.4. For $1 \leq i \leq H$, the process $\mathfrak{c}_i$, corresponding to the instantaneous claim for policyholder i at time $t \geq 0$, is defined as

$$\mathfrak{c}_{i,t} := -\sum_{j=1}^{K_i} \left(z_{ij,t} - \overline{z}_{ij,t}\right). \tag{II.9}$$

This represents the sum of the differences in income for each sub-unit, with and without a cyberattack. An explicit expression of this cost is straightforward using the definition of subunit revenue in (II.6). We have

$$\mathfrak{c}_{i,t} = \sum_{j=1}^{K_i} z_{ij,0}\pi_{ij}\mathbf{1}_{\tau_{ij}\leq t<\tau_{ij}+\delta_{ij}} \exp\left(\left(\mu_{ij} - \frac{1}{2}\sigma_{ij}^2\right)t + \sigma_{ij}B_{ij,t}\right). \tag{II.10}$$

In order to measure the feedback on the financial nature of the attack, we introduce a widely used metric in catastrophe risk modeling: the Aggregate Exceedance Probability (AEP) [18, 24]. It is the probability that the sum of losses over a period exceeds a certain amount. In general, AEP is computed on one year, but we consider here an instantaneous version.

Definition II.5. For $t, u \geq 0$, the Aggregate Exceedance Probability in the period $[t, t+u)$ is

$$\mathrm{AEP}_{[t,t+u)}(x) := \mathbb{P}\left(\sum_{i=1}^{H} \mathfrak{C}_{i,[t,t+u)} > x\right),$$

for all $x \geq 0$ and where for all $1 \leq i \leq H$,

$$\begin{aligned}\mathfrak{C}_{i,[t,t+u)} &:= \int_t^{t+u} \mathfrak{c}_{i,s}\mathrm{d}s \\ &= \sum_{j=1}^{K_i} z_{ij,0} \int_{t\vee\tau_{ij}}^{(t+u)\wedge(\tau_{ij}+\delta_{ij})} \pi_{ij} e^{\left(\mu_{ij}-\frac{1}{2}\sigma_{ij}^2\right)s+\sigma_{ij}B_{ij,s}}\mathrm{d}s,\end{aligned} \tag{II.11}$$

is the total claim of firm i between t and $t+u$.

However, not all the policyholders are experiencing cyber losses at time t. Precisely, let us note

$$\mathcal{I}^\star_{[t,t+u)} = \left\{ i \text{ if } \sum_{j=1}^{K_i} \mathbf{1}_{t\leq\tau_{ij}<t+u} \geq 1 \right\}, \tag{II.12}$$

the set of policyholders that contains at least one subunit infected in the period $[t, t+u)$. Clearly the cardinal $|\mathcal{I}^\star_{[t,t+u)}| \leq H$ firms. Then a simplified expression of AEP can be written as

$$\mathrm{AEP}_{[t,t+u)}(x) := \mathbb{P}\left(\sum_{i\in\mathcal{I}^\star_{[t,t+u)}} \mathfrak{C}_{i,[t,t+u)} > x\right), \quad \text{(II.13)}$$

Remark II.6. In the very special case where all the claims $(\mathfrak{C}_{i,[t,t+u)})_{i\in\mathcal{I}^\star_{[t,t+u)}}$ are homogeneous, in the sense that they have the same cumulative distribution function $F_{C_{1,[t,t+u)}}$, we can obtain a simplified expression of AEP using convolution. We have

$$\begin{aligned}\mathrm{AEP}_{[t,t+u)}(x) &= \mathbb{P}\left(\sum_{i\in\mathcal{I}^\star_{[t,t+u)}} \mathfrak{C}_{i,[t,t+u)} > x\right)\\ &= 1 - \mathbb{P}\left(\sum_{i\in\mathcal{I}^\star_{[t,t+u)}} \mathfrak{C}_{i,[t,t+u)} \leq x\right)\\ &= 1 - F^{(|\mathcal{I}^\star_{[t,t+u)}|)}_{C_{1,[t,t+u)}}(x),\end{aligned}$$

where $|\mathcal{I}^\star_{[t,t+u)}|$ denotes the cardinal of the set $\mathcal{I}^\star_{[t,t+u)}$, and for $m \in \mathbb{N}^*$, $F^{(m)}_{C_{1,[t,t+u)}}(x)$ is the m-fold convolution of $F_{\mathfrak{C}_{1,[t,t+u)}}(x)$, defined as

$$F^{(m)}_{C_{1,[t,t+u)}}(x) = \int_0^x F^{(m-1)}_{C_{1,[t,t+u)}}(x-y)\,\mathrm{d}\mathbb{P}(C_{1,[t,t+u)} \in \mathrm{d}y).$$

Other metrics to quantify the impact of the attack on the portfolio are the quantiles of the total portfolio loss. For a cyber episode lasting until date $T \geq 0$, the total loss $\mathfrak{C}_T$ of the portfolio on $[0, T]$ is given by

$$\mathfrak{C}_T := \sum_{i=1}^{H}\left(\int_0^T \mathfrak{c}_{i,t}\mathrm{d}t\right), \quad \text{(II.14)}$$

where for each $1 \leq i \leq H$,

$$\int_0^T \mathfrak{c}_{i,t}\mathrm{d}t = \sum_{j=1}^{K_i} z_{ij,0}\pi_{ij}\int_{\tau_{ij}\wedge T}^{(\tau_{ij}+\delta_{ij})\wedge T} e^{\left(\mu_{ij}-\frac{1}{2}\sigma_{ij}^2\right)s+\sigma_{ij}B_{ij,s}}\mathrm{d}s, \quad \text{(II.15)}$$

is the total claim for firm i. All these metrics require to know the distribution of the processes $(\mathfrak{c}_{i,t})_t$ for each i. The sources of randomness of the latter are from $(B_{ij,t})$, $(N_{ij,t})$, and (π_{ij}). Even if assuming $(\pi_{ij,t})$ deterministic, $(\mathfrak{c}_{i,t})_{i,t}$ would not have an explicit distribution, since it is the sum of integral of log-normal with random bounds. We will return to this point later.

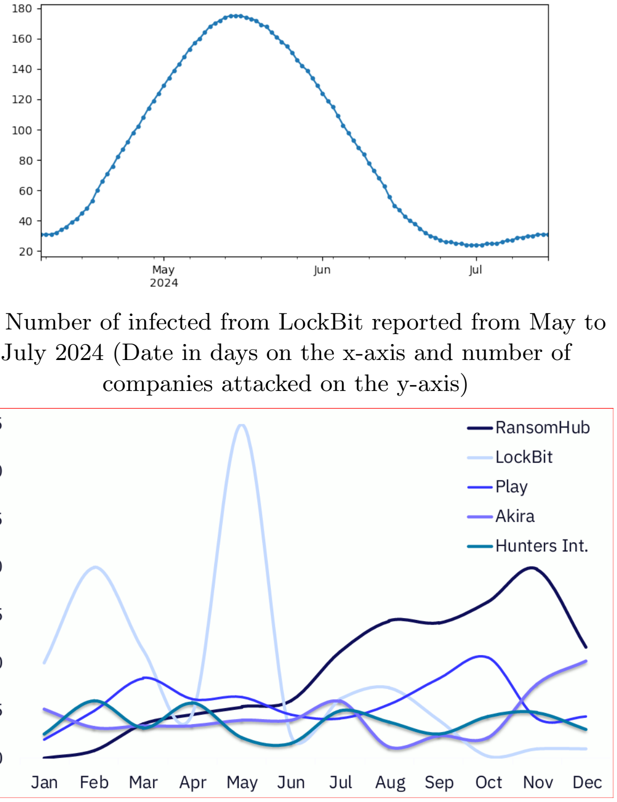
In order to evaluate the impact of cyber incidents on an economy (using (II.7)) or on an insurance portfolio (using (II.13)), one need to describe the arrival of cyber contagion or in other words, the dynamics of the process $(N_{ij})_{ij}$. Early works such as [31] or [36] model the arrival of cyberattacks using (in)homogeneous Poisson processes. Thereafter, [7] and [8] show that Hawkes processes are particularly suitable to capture the contagion phenomena, the clustering, and the autocorrelation of the arrival times of cyber incidents. Last but not least, by noting that a cyber incident and a biological epidemic have similar characteristics, [22] and [23] use compartmental epidemiological models to describe cyber contagion by linking the hazard rate function of arrival times with the dynamics of infected individuals. This paper follows the latter approach.

III. THE CONTAGION MODEL

As mentioned by [22] and [23], and as suggested by Figure 1 (from the 2025 Ransomware & Cyber Threat Report of *GuidePoint Security's Research and Intelligence Team's [GRIT]*), biological epidemics and cyber contagion episodes have fairly similar characteristics. Much like certain contagious biological outbreaks, a firm (just as a person) could be susceptible i.e. not yet infected but vulnerable if exposed to an infectious firm, infected i.e. affected by the malware and which can transmit, and removed or recovered i.e. whose systems are cured, secured, or patched. Moreover, when examining the dynamics of companies infected by the LockBit ransomware between May and July 2024 on Figure 1a, the number of infections starts out low, then escalates rapidly until it reaches a peak, before beginning to decline. Finally, just as in-household and out-household infection rates are different in the context of biological epidemics, the same applies to cyber epidemics when companies are viewed as collections of interconnected subunits. Subsidiaries of a given company are, by nature, more interconnected together than with external organizations.

A. The model

To model the infection process within a firm, we adapt the SIR for households proposed by [12]. Precisely, we extend their deterministic SIR into a stochastic SIR in the sense that the model parameters are no longer deterministic but instead follow a Cox–Ingersoll–Ross dynamic. As [22], we assume that contagion does not come from inside the portfolio itself, but from the outside, so that the spread of the attack is defined as a global level. We view a cyber infectious event in a fully



(a) Number of infected from LockBit reported from May to July 2024 (Date in days on the x-axis and number of companies attacked on the y-axis)

(b) Most Impactful Ransomware in 2024 (the date in months on the x-axis and the number of companies attacked on the y-axis)

FIG. 1: GRIT 2025 Ransomware & Cyber Threat Report (see [19])

susceptible firm of size $k \in \{1, \dots, K\}$ as a splitting process generating a infected firm of size j, where $1 \leq j \leq k$ and a remaining susceptible firm of size $k - j$. Let s_j, i_j and r_j denote the number of firms susceptible, infected, and removed, of size j, where $1 \leq j \leq K$. The quantity K, introduced in Equation (II.3), represents the maximal size of a firm. We denote

- $h_j = s_j + i_j + r_j$ equals to the total number of current firms of size j;
- $h = \sum_{j=1}^{K} h_j$ as the total number of firms;
- The total population (number of subunits) is given by $n = \sum_{j=1}^{K} jh_j$.

Rather than working with the number of susceptible/infected/removed firms which are integers and not fully adapted to the ODE, we introduce for all $1 \leq j \leq K$, $S_j := s_j/h$, $I_j := i_j/h$, $R_j := r_j/h$, and $H_j := h_j/h$ the fraction of susceptible, infected, removed, and all firms of size j respectively. We thus write the average size of firm as

$$N = \sum_{j=1}^{K} jH_j. \tag{III.1}$$

Let us now describe how the cyber-infection spreads into the subunits. Let j range from 1 to K.

1. If an initial infection is brought into a susceptible firm of size j, secondary infections will occur inside the firm.
2. Each of the remaining $j - 1$ firm members can get infected with equal probability a. This represents the "in-firm" infection rate.
3. The probability that a primary infection in a firm of size j generates in total k infections inside this firm is $b_{j,k}$, where $1 \leq k \leq j$.
4. The secondary infections give rise to a splitting of the initial firm of size j into a new, fully infected firm of size k and another still susceptible firm of size $j - k$.
5. An infected firm of size k recovers with a rate γ_k and contributes to the overall force of infection between different firms.
6. β_k is the "out–firm" infection rate which refers to infection events occurring between different firms and hence outside a given single firm.

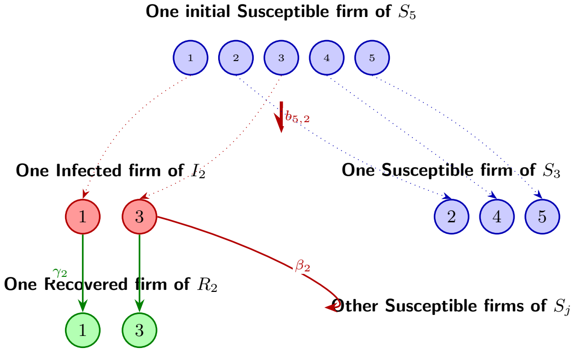

One initial Susceptible firm of S_5
1
2
3
4
5
$b_{5,2}$
One Infected firm of I_2
One Susceptible firm of S_3
1
3
2
4
5
γ_2
β_2
One Recovered firm of R_2
Other Susceptible firms of S_j
1
3

FIG. 2: Infection and Splitting: $S_5 \to I_2 + S_3$

The graph in Figure 2 above gives an example where $k = 5$ and $j = 2$. We summarize and precise these features in the following assumption.

Assumption III.1. The dynamic system governing the dynamics of the susceptible, infected and removed firms of size k reads as

$$\frac{\mathrm{d}S_{k,t}}{\mathrm{d}t} = Y_t \left(-kS_{k,t} + \sum_{j=k+1}^{K} jS_{j,t} \cdot b_{jj-k,t} \right), \quad \text{(III.2a)}$$

$$\frac{\mathrm{d}I_{k,t}}{\mathrm{d}t} = -\gamma_{k,t} I_{k,t} + Y_t \sum_{j=k}^{K} jS_{j,t} \cdot b_{jk,t}, \quad \text{(III.2b)}$$

$$\frac{\mathrm{d}R_{k,t}}{\mathrm{d}t} = \gamma_{k,t} I_{k,t}, \quad \text{(III.2c)}$$

where

$$Y_t = \frac{1}{N_t} \sum_{k=1}^{K} \beta_{k,t} \cdot kI_{k,t}, \quad \text{(III.3)}$$

is the total force of infection − the instantaneous risk that a susceptible individual becomes infected. Let us note the set of parameters

$$\Psi := \{\beta_1, \dots, \beta_K, \gamma_1, \dots, \gamma_K, \\ b_{11}, \dots, b_{1K}, b_{21}, \dots, b_{K-1K}, b_{KK}\}.$$

We then have $2K + K^2$ model parameters. For each parameter $\varphi \in \Psi$, we assume that the $[0,1]$-valued process $\varphi = \frac{1}{1+e^{-\tilde{\varphi}}} \in [0,1]$ where $\tilde{\varphi}$ evolves according to

$$\mathrm{d}\tilde{\varphi}_t = \kappa_\varphi(\mu_\varphi - \tilde{\varphi}_t)\mathrm{d}t + \Sigma_\varphi \sqrt{\tilde{\varphi}_t} \mathrm{d}\mathcal{W}_{\varphi,t} \quad \text{(III.4)}$$

with the constants $\mu_\varphi \in \mathbb{R}$, $\Sigma_\varphi > 0$, and $\kappa_\varphi, \varphi_0 \geq 0$. Moreover, $(\mathcal{W}_\varphi)_{\varphi \in \Psi}$ is a $2K + K^2$-dimensional Brownian motion.

Remark III.2. In this SIR model, the total population

$$N = \sum_{k=1}^{K} k(S_k + I_k + R_k),$$

defined in (III.1), is conserved i.e. for all $t \geq 0$, $N_t = N_0$. From here on, we will write N_0 instead of N_t to refer to the total population.

We could have used a SIRS model that allows an infected-and-removed unit to become susceptible again and be reinfected, but we can ignore this feature since we are dealing with a short-lived cyber incident. In other words, there is not enough time for a unit to be infected multiple times.

The details on the deterministic version of this model are detailed in [12]. The stochasticity in the epidemiological model is motivated by the fact that fluctuations in the environment affect models' parameters. Several works in the literature introduce randomness by simply transforming the model from a system of ODEs to a system of SDEs. This implies that the randomness comes from a single parameter (for example on transmission coefficient between compartments for [26] or on the the decay rate of immunity for [9]). The resulting form of this alternative are described in Appendix A.

However, since all the parameters can be subject to environmental variability, it appears more natural to consider each parameter having its own stochastic dynamics.

Remark III.3 (Feller condition)**.** For $\varphi \in \Psi$, according to (III.4), $\tilde{\varphi}$ follows a Cox-Ingersoll-Ross process (see [11] for details). This ensures that φ is a $\mathbb{R}_+$-value process. Precisely,

- for $\varphi_0 > 0$, the process will never touch zero, if $2\kappa_\varphi \mu_\varphi \geq \Sigma_\varphi^2$;
- otherwise it can occasionally touch 0.

Let $t \geq 0$, given φ_0, we also have

$$\mathbb{E}[\tilde{\varphi}_t | \varphi_0] = \varphi_0 e^{-\kappa_\varphi t} + \mu_\varphi(1 - e^{-\kappa_\varphi t}),$$

and

$$\mathbb{V}[\tilde{\varphi}_t | \varphi_0] = \varphi_0 \frac{\Sigma_\varphi^2}{\kappa_\varphi} \left(e^{-\kappa_\varphi t} - e^{-2\kappa_\varphi t} \right) + \frac{\mu_\varphi \Sigma_\varphi^2}{2\kappa_\varphi} \left(1 - e^{-\kappa_\varphi t} \right)^2.$$

We define the (right continuous and complete) filtration $\mathbb{F}^{\mathcal{W}} = (\mathcal{F}_t^{\mathcal{W}})_{t \geq 0}$ on the set of parameters as: for $t \geq 0$,

$$\mathcal{F}_t^{\mathcal{W}} := \sigma\left(\{\mathcal{W}_{\varphi,s}, \ s \leq t \text{ and } \varphi \in \Psi\} \right). \quad \text{(III.5)}$$

Remark III.4. For the numerical analysis, the infections inside the firm is modeled by a Bernoulli-process with in-firm attack rate (infection probability) $a \in [0,1]$, the total number of infected people inside the firm follows a binomial distribution, namely, for $1 \leq j, k \leq K$,

$$b_{jk} = \binom{j-1}{k-1} a^{k-1}(1-a)^{j-k}. \quad \text{(III.6)}$$

Next, we also assume that $a = \frac{1}{1+e^{-\tilde{a}}} \in [0,1]$ where the process $\tilde{a}$ is governed by a CIR dynamics. This assumption reduces the dimension of the set Ψ from $2k + k^2$ to $2k + 1$ (namely $\Psi := \{a, \beta_1, \dots, \beta_K, \gamma_1, \dots, \gamma_K\}$), and ease the calibration process.

B. Existence and uniqueness of the nonnegative solution

This section is dedicated to the existence and uniqueness result of a nonnegative solution to the system (III.2a)-(III.2b)-(III.2c).

When $\mathfrak{s}, \mathfrak{i}, \mathfrak{r} \in \mathbb{R}_+^K$, we first consider the feasible region Γ defined as follows:

$$\Gamma := \left\{ (\mathfrak{s}, \mathfrak{i}, \mathfrak{r}) \in \mathbb{R}_+^{3K} \middle| \sum_{k=1}^{K} k(\mathfrak{s}_k + \mathfrak{i}_k + \mathfrak{r}_k) \leq N_0 \right\}, \tag{III.7}$$

where N_0 is the average size of firm defined in eq. (III.1). The following theorem is inspired by [26].

Theorem III.5 (Existence and uniqueness)**.** *Consider the system* (III.2a)-(III.2b)-(III.2c).

1. *For $\omega \in \Omega$, $t_0 \geq 0$, and for any initial value $(S_{t_0}, I_{t_0}, R_{t_0}) \in \Gamma$, there is a unique global solution $(S_t(\omega), I_t(\omega), R_t(\omega))_{t \geq t_0}$ in Γ of system* (III.2a)-(III.2b)-(III.2c).

2. *S, I, and R are $\mathbb{F}^{\mathcal{W}}$-adapted.*

Proof. The differential system (III.2a)-(III.2b)-(III.2c) is an ordinary differential equation (ODE) with stochastic coefficients sometimes abusively called stochastic differential equation (SDE).

1. Consider the "associated" deterministic ODE i.e. the one associated with a fixed sample path. For a fixed $\omega \in \Omega$, the system becomes

$$\frac{\mathrm{d}S_{k,t}}{\mathrm{d}t} = Y_t \left(-kS_{k,t} + \sum_{j=k+1}^{K} jS_{j,t} \cdot b_{jj-k,t}(\omega) \right), \tag{III.8a}$$

$$\frac{\mathrm{d}I_{k,t}}{\mathrm{d}t} = -\gamma_{k,t}(\omega) I_{k,t} + Y_t \sum_{j=k}^{K} jS_{j,t} \cdot b_{jk,t}(\omega), \tag{III.8b}$$

$$\frac{\mathrm{d}R_{k,t}}{\mathrm{d}t} = \gamma_{k,t}(\omega) I_{k,t}, \tag{III.8c}$$

where

$$Y_t = \frac{1}{N_0} \sum_{k=1}^{K} \beta_{k,t}(\omega) \cdot kI_{k,t}. \tag{III.9}$$

The system can be rewritten in the following form

$$(\dot{S}, \dot{I}, \dot{R}) = F(S, I, R, \omega),$$

where $F(S, I, R, \omega)$ corresponds to the right hand side of the system (III.8a)-(III.8b)-(III.8c). The function F is locally Lipschitz continuous, for any given initial value $(S_{t_0}, I_{t_0}, R_{t_0}) \in \mathbb{R}_+^{3K}$: according to Picard-Lindelöf theorem, there is a unique local solution (S_t, I_t, R_t) on $t \geq t_0$, for any $t_0 \geq 0$.

From (III.8b), we have

$$\frac{\mathrm{d}I_{k,t}}{\mathrm{d}t} = -\gamma_{k,t}(\omega) I_{k,t} + Y_t \sum_{j=k}^{K} jS_{j,t} \cdot b_{jk,t}(\omega)$$
$$\geq -\gamma_{k,t}(\omega) I_{k,t},$$

then

$$\frac{\mathrm{d}I_{k,t}}{I_{k,t}} \geq -\gamma_{k,t}(\omega)\mathrm{d}t,$$

which implies that

$$I_{k,t} \geq I_{k,t_0} \exp\left(-\int_{t_0}^{t} \gamma_{k,s}(\omega)\mathrm{d}s \right) \geq 0.$$

From (III.8c), by noticing that $\frac{\mathrm{d}R_{k,t}}{\mathrm{d}t} \geq -\alpha_{k,t}(\omega)R_{k,t}$, and from (III.8a), by noticing that $Y_t \leq 1$ and $\frac{\mathrm{d}S_{k,t}}{\mathrm{d}t} \geq -kS_{k,t}$, we also conclude that $S_{k,t} \geq 0$ and $R_{k,t} \geq 0$. However, from Theorem III.2, for $t \geq 0$, $N_t = N_0$. This means that both forward limit and backward limits (i.e. $lim_{t \to +\infty} N_t$ and $\lim_{t_0 \to 0} N_t$ as defined in [10]) of N_t are bounded consequently the local solution can be extended to a global one and belongs to Γ.

2. When a solution exists, the second item is straightforward. Since each coefficient $\varphi \in \Psi$ satisfy Lipschitz and linear growth conditions, there exists an unique strong solution, namely,

$$\tilde{\varphi}_t = \varphi_0 e^{-\kappa_\varphi t} + \mu_\varphi \left(1 - e^{-\kappa_\varphi t}\right)$$
$$+ \Sigma_\varphi \int_0^t e^{-\kappa_\varphi (t-s)} \sqrt{\tilde{\varphi}_s}\, d\mathcal{W}_{\varphi,s}, \quad \text{for all } t \geq 0 \tag{III.10}$$

which is $\mathbb{F}$-adapted. Therefore all coefficients φ of the system (III.2a)-(III.2b)-(III.2c) are $\mathbb{F}^{\mathcal{W}}$-adapted, whose unique solution (S, IR) is also $\mathbb{F}^{\mathcal{W}}$-adapted.

□

It can also be noted that the region Γ is positively invariant of our stochastic model equation.

To analyze the spread of the cyber attack, we can introduce additional quantities such as the prevalence i.e. the fraction of recovered subunits,

$$\mathfrak{P} := \frac{1}{N_0} \sum_{k=1}^{K} kR_k, \tag{III.11}$$

and the peak i.e. the maximum of total number of infected subunits,

$$\mathfrak{I} := \max_{t \geq 0} \sum_{k=1}^{K} kI_{k,t}. \tag{III.12}$$

Also in epidemiology modeling literature, the basic reproduction number is the number of new infections produced by a typical infective individual in a population at a Disease Free Equilibrium (DFE), the state at which a population remains in the absence of disease. It is used to determine if the disease always dies out (after some time) or if it persists (around an endemic equilibrium). When the model parameters are constant, [12] shows that the basic reproduction number R_0 is

$$R_0 = \sum_{i=1}^{K} i \frac{\beta_i}{\gamma_i} \sum_{m=i}^{K} \frac{mS_m}{N_0} \cdot b_{m,i}. \tag{III.13}$$

Moreover, some works such as [20, 34] show that, if the model admits a DFE, then the latter is locally asymptotically stable if $R_0 < 1$. Similar results exist for the stochastic case but where the randomness is introduced by using a system of SDE instead of a system of ODE (see [26]). The following theorem provides a basic reproduction number and a local stability condition in the dynamic contagion model developed in this paper.

Theorem III.6 (Necessary condition for the local stability)**.** *Let $t \geq 0$. If*

$$\mathcal{R}_{max,t} := \left[\max_{1 \leq k \leq K} \frac{\beta_{k,t}}{\gamma_{k,t}}\right] \sum_{i=1}^{K} i \sum_{m=i}^{K} \frac{mS_{m,t}}{N_0} \cdot b_{mi,t} < 1, \tag{III.14}$$

then

$$\frac{\mathrm{d} \log \mathfrak{I}_t}{\mathrm{d}t} < 0.$$

Proof. Let $t \geq 0$, from (III.12), we have

$$\begin{aligned} \mathfrak{I}'_t &= \sum_{k=1}^{K} k I'_k(t) \\ &= \sum_{k=1}^{K} k \left(-\gamma_{k,t} I_{k,t} + Y_t \sum_{j=k}^{K} j S_{j,t} \cdot b_{jk,t} \right), \end{aligned}$$

By taking the Itô's formula

$$\begin{aligned} \frac{\mathrm{d} \log \mathfrak{I}_t}{\mathrm{d}t} &= \frac{\mathrm{d}\mathfrak{I}_t}{\mathrm{d}t} \frac{1}{\mathfrak{I}_t} \\ &= -\frac{1}{\mathfrak{I}_t} \sum_{k=1}^{K} k\gamma_{k,t} I_{k,t} + Y_t \sum_{k=1}^{K} k \sum_{j=k}^{K} jS_{j,t} \cdot b_{jk,t} \\ &= -\frac{1}{\mathfrak{I}_t} \left(\sum_{k=1}^{K} k\gamma_{k,t} I_{k,t} \right. \\ &\qquad \left. - \frac{1}{N_0} \sum_{k=1}^{K} \beta_{k,t} \cdot kI_{k,t} \sum_{i=1}^{K} i \sum_{m=i}^{K} mS_{m,t} \cdot b_{mi,t} \right) \\ &= -\frac{1}{\mathfrak{I}_t} \sum_{k=1}^{K} k\gamma_{k,t} I_{k,t} \left(1 - \frac{\beta_{k,t}}{\gamma_{k,t}} \sum_{i=1}^{K} i \sum_{m=i}^{K} \frac{mS_{m,t}}{N_0} \cdot b_{mi,t} \right) \\ &= -\sum_{k=1}^{K} \frac{k\gamma_{k,t} I_{k,t}}{\mathfrak{I}_t} \left(1 - \mathcal{R}_{k,t}\right). \end{aligned}$$

where we denote

$$\mathcal{R}_{k,t} := \frac{\beta_{k,t}}{\gamma_{k,t}} \sum_{i=1}^{K} i \sum_{m=i}^{K} \frac{mS_{m,t}}{N_0} \cdot b_{mi,t}. \tag{III.15}$$

With $\mathcal{R}_{max}$ defined in (III.14), we have

$$\mathcal{R}_{k,t} \leq \mathcal{R}_{max,t}.$$

Therefore $1 - \mathcal{R}_{max,t} \leq 1 - \mathcal{R}_{k,t}$ and since $\frac{-k\gamma_{k,t} I_{k,t}}{\mathfrak{I}_t} \leq 0$, we have

$$\frac{\mathrm{d} \log \mathfrak{I}_t}{\mathrm{d}t} \leq -\sum_{k=1}^{K} \frac{k\gamma_{k,t} I_{k,t}}{\mathfrak{I}_t} \left(1 - \mathcal{R}_{max,t}\right),$$

which is negative when $\mathcal{R}_{max,t} < 1$. □

In addition, if $\mathcal{R}_\infty := \max_{t \geq 0}\{\mathcal{R}_{max,t}\} < 1$, we can easily show that there exists $h < 0$ such that

$$\limsup_{T \to +\infty} \frac{\log \mathfrak{I}_t}{T} \leq h. \tag{III.16}$$

Remark III.7. We can also make the following remarks:

- In our framework, $\mathcal{R}_{max,t}$ (respectively $\mathcal{R}_\infty$) represents the local (respectively global) basic reproduction number.
- In addition to the model parameters, both $\mathcal{R}_{max,t}$ and $\mathcal{R}_\infty$ can depend on the trajectories of susceptible subunits $(S_{m,t})_{1 \leq m \leq K, t \geq 0}$. When $\frac{\beta_{k,t}}{\gamma_{k,t}}$ is independent of k and t and $b_{mi,t}$ independent of t as in [12], we retrieve exactly the expected R_0 given in (III.13).
- It is obvious that $\mathfrak{I} \geq 0$ and Theorem III.5 implies that $\mathfrak{I}' < 0$ therefore the total number of infected subunits strictly decreases locally with time, or in other words, the epidemic is set to disappear.

- Better, (III.16) means that $\mathfrak{J}$ almost surely converges exponentially to 0 or the epidemic disappears exponentially.
- These results mean that the contagion-free equilibrium point, if it exists, is locally asymptotically stable.

IV. FROM THE STOCHASTIC SIR TO THE IMPACT ON AN INSURANCE PORTFOLIO

Once the cyber contagion model is described, we are interested in the following questions concerning the distribution of the subunits' and firm's sizes, as well as the impact of the cyber contagion on the firm size growth, on a cyber insurance portfolio and more generally on the economy growth. Recall from (II.1) and (II.6), the revenue $Z_{i,t}$ of firm $i \in \{1, \dots, H\}$ at time $t \geq 0$, satisfies

$$Z_{i,t} = \sum_{j=1}^{K_i} z_{ij,t} = \sum_{j=1}^{K_i} \left(1 - \pi_{ij} \mathbf{1}_{\tau_{ij} \leq t < \tau_{ij} + \delta_{ij}}\right) \overline{z}_{ij,t}$$

where τ_{ij} is the first arrival time of the point process N_{ij} and δ_{ij} is the random recovery time. The arrival of a cyber attack can come either from outside or from a sister subunit. We thus have the following assumption.

Assumption IV.1. For $(i,j) \in \mathcal{I}$ and $t \geq 0$, we assume

$$N_{ij,t} := N^0_{ij,t} + \mathbf{1}_{\left\{N^0_{ij,\tau_i^-} = 0\right\}} \mathbf{1}_{\{t \geq \tau_i\}} U_{ij},$$

where

- $(N^0_{ij})_{ij}$ is a sequence of independent Cox processes with common intensity Y defined in (III.3).
- τ_i is a stopping time so that
$$\tau_i := \inf \left\{ t > 0 : \sum_{k=1}^{K_i} N^0_{ik,t} \geq 1 \right\}. \tag{IV.1}$$
- And $(U_{ij})_{ij}$ is a sequence of iid Bernoulli random variables with probability a_{τ_i}.

Let $(i,j) \in \mathcal{I}$ and $t \geq 0$, the previous assumption comes from the fact that the infection of j can come from inside or outside the firm i.

- When the infection of j comes from outside between t and $t + \mathrm{d}t$, its probability is
$$Y_t \mathrm{d}t$$
where Y_t is the force of infection defined in (III.3). In other words, the arrival of successful cyberattacks from outside $N^0_{ij,t}$ satisfies $\mathbb{E}[N^0_{ij,t+\mathrm{d}t} - N^0_{ij,t} | \mathcal{F}^{\mathcal{W}}_t] = Y_t \mathrm{d}t$ recalling that Y is $\mathbb{F}^{\mathcal{W}}$-adapted (with $\mathbb{F}^{\mathcal{W}}$ introduced in (III.5).
- When there exists at least one sister subunit of j infected from outside at time t i.e. $K_i > 1$ ans $\sum_{k=1}^{K_i} N^0_{ik,t} \geq 1$, and if the subunit j is not infected from outside at the time τ_i i.e. $\mathbf{1}_{\left\{N^0_{ij,\tau_i^-} = 0\right\}}$, then it may have an internal contagion U_{ij} with probability a_{τ_i} where a is a process defined in Assumption III.1.

For each firm $1 \leq i \leq H$, if $K_i > 1$, the vector of point processes $(N_{ij})_{1 \leq j \leq K_i}$ is strongly correlated. We describe in the following proposition the marginal law of first jump of subunit j, namely τ_{ij} defined in (II.5).

Proposition IV.2. *For $(i,j) \in \mathcal{I}$, if $K_i \geq 2$, then the conditional marginal cdf function of τ_{ij} is*

$$\begin{aligned} F_{ij}(t) &= \mathbb{P}[\tau_{ij} \leq t] \\ &= 1 - e^{-K_i \Lambda_t} - (K_i - 1) e^{-\Lambda_t} \int_0^t Y_s e^{-(K_i - 1)\Lambda_s} (1 - a_s) \mathrm{d}s, \end{aligned} \tag{IV.2}$$

for all $t \geq 0$ and where $\Lambda_t := \int_0^t Y_s \mathrm{d}s$.

Proof. Let $(i,j) \in \mathcal{I}$ and $t \geq 0$. If $K_i = 1$, there is not internal contagion. We directly have

$$\mathbb{P}[\tau_{ij} > t | \mathcal{F}^{\mathcal{W}}_t] = \mathbb{P}[N^0_{ij,t} = 0 | \mathcal{F}^{\mathcal{W}}_t] = e^{-\Lambda_t}.$$

If $K_i \geq 2$, knowing that $\{\tau_{ij} > t\} = (\{N^0_{ij,t} = 0\} \cap \{\tau_i > t\}) \cup (\{N^0_{ij,t} = 0\} \cap \{\tau_i \leq t\} \cap \{U_{ij} = 0\})$ where $\{N^0_{ij,t} = 0\} \cap \{\tau_i > t\}$ and $\{N^0_{ij,t} = 0\} \cap \{\tau_i \leq t\} \cap \{U_{ij} = 0\}$ are disjoint, we have

$$\begin{aligned} \mathbb{P}[\{N^0_{ij,t} = 0\} \cap \{\tau_i > t\} | \mathcal{F}^{\mathcal{W}}_t] &= \mathbb{P}\left[\cap_{k=1}^{K_i} \{N^0_{ik,t} = 0\} | \mathcal{F}^{\mathcal{W}}_t\right] \\ &= e^{-K_i \Lambda_t}, \end{aligned}$$

where the last equality comes from the fact $\left(N^0_{ik}\right)_{ik}$ are iid conditionally to $\mathcal{F}^{\mathcal{W}}_t$. We also have

$$\begin{aligned} &\mathbb{P}\left[\{N^0_{ij,t} = 0\} \cap \{\tau_i \leq t\} \cup \{U_{ij} = 0\} | \mathcal{F}^{\mathcal{W}}_t\right] \\ &\quad = \mathbb{P}\left[N^0_{ij,t} = 0, \tau_i \leq t, U_{ij} = 0 | \mathcal{F}^{\mathcal{W}}_t\right] \\ &\quad = \int_0^t \mathbb{P}\left[N^0_{ij,t} = 0, \tau_i \in \mathrm{d}s | \mathcal{F}^{\mathcal{W}}_t\right] (1 - a_s), \end{aligned}$$

where for $0 \leq s \leq t$, $1 - a_s = \mathbb{P}[U_{ij} = 0]$ because $\tau_i = s$. However, $\mathbb{P}\left[N^0_{ij,t} = 0, \tau_i \in \mathrm{d}s | \mathcal{F}^{\mathcal{W}}_t\right] = \mathbb{P}\left[N^0_{ij,t} = 0 | \mathcal{F}^{\mathcal{W}}_t\right] \mathbb{P}\left[\tau_i \in \mathrm{d}s | N^0_{ij,t} = 0, \mathcal{F}^{\mathcal{W}}_t\right]$. With

$$\mathbb{P}\left[N^0_{ij,t} = 0 | \mathcal{F}^{\mathcal{W}}_t\right] = e^{-\Lambda_t},$$

and

$$\begin{aligned} \mathbb{P}\left[\tau_i \in \mathrm{d}s | N^0_{ij,t} = 0, \mathcal{F}^{\mathcal{W}}_t\right] &= \frac{\mathrm{d}}{\mathrm{d}s} \mathbb{P}\left[\tau_i \leq s | N^0_{ij,t} = 0, \mathcal{F}^{\mathcal{W}}_t\right] \\ &= -\frac{\mathrm{d}}{\mathrm{d}s} \mathbb{P}\left[\cap_{k=1, k \neq j}^{K_i} \{N^0_{ik,t} = 0\} | \mathcal{F}^{\mathcal{W}}_t\right] \\ &= -\frac{\mathrm{d}}{\mathrm{d}s} e^{-(K_i - 1)\Lambda_s} \\ &= (K_i - 1) Y_s e^{-(K_i - 1)\Lambda_s}. \end{aligned}$$

Therefore,

$$\begin{aligned}&\mathbb{P}\left[\{N^0_{ij,t}=0\}\cap\{\tau_i\le t\}\cup\{U_{ij}=0\}|\mathcal{F}^{\mathcal{W}}_t\right]\\&\quad=\int_0^t (K_i-1)Y_s e^{-(K_i-1)\Lambda_s}(1-a_s)e^{-\Lambda_t}\mathrm{d}s.\end{aligned}$$

Finally

$$\begin{aligned}\mathbb{P}[\tau_{ij}>t|\mathcal{F}^{\mathcal{W}}_t]&=e^{-K_i\Lambda_t}\\&\quad+(K_i-1)e^{-\Lambda_t}\int_0^t Y_s e^{-(K_i-1)\Lambda_s}(1-a_s)\mathrm{d}s,\end{aligned}$$

and the conclusion follows. □

Remark IV.3. From (IV.2), we can make the following remarks:

1. The right term does not depend on the subunit so that the marginal distributions of first jump of all the subunits of the same firm are identical. We will note F_i the cumulative distribution function, instead of F_{ij}.
2. For a fixed firm i at a fixed date t, the function $a \mapsto F_i(t)$ is not decreasing: the probability that a subunit becomes infected increases with the probability of internal contagion.
3. For a fixed firm i at a fixed date t, if the function $K_i \mapsto (K_i-1)\Lambda_s - 1$ is negative for all $0\le s\le t$, then the function $K_i \mapsto F_i(t)$ is non-decreasing. In fact, by differentiating the function $K_i \mapsto F_i(t)$, we get

$$\begin{aligned}\frac{\mathrm{d}F_i(t)}{\mathrm{d}K_i}&=K_i e^{-K_i\Lambda_t}\\&\quad-e^{-\Lambda_t}\int_0^t Y_s e^{-(K_i-1)\Lambda_s}(1-a_s)\mathrm{d}s,\\&\quad+(K_i-1)e^{-\Lambda_t}\int_0^t \Lambda_s Y_s e^{-(K_i-1)\Lambda_s}(1-a_s)\mathrm{d}s,\\&=K_i e^{-K_i\Lambda_t}\\&\quad+e^{-\Lambda_t}\int_0^t[(K_i-1)\Lambda_s-1]Y_s e^{-(K_i-1)\Lambda_s}(1-a_s)\mathrm{d}t.\end{aligned}$$

   This means that the probability that a subunit becomes infected increases with the number of subunits. This is obvious in particular when the process (Λ_t) is bounded by $\frac{1}{K-1}$.
4. For all firms with the same size (number of subunits), all their subunits have the same probability to be infected at a given time.
5. Let $1\le i\le H$ and $t\ge 0$. Because

$$\mathbb{P}\left[\cap_{k=1}^{K_i}\{N_{ik,t}=0\}|\mathcal{F}^{\mathcal{W}}_t\right]=\mathbb{P}\left[\cap_{k=1}^{K_i}\{N^0_{ik,t}=0\}|\mathcal{F}^{\mathcal{W}}_t\right]=e^{-K_i\Lambda_t},$$

   the probability that at least one subunit of firm i gets infected is $1-e^{-K_i\Lambda_t}$ which increases with K_i. Therefore, the vulnerability of firms increases with their size.

The marginal laws (F_i) are obviously $\mathbb{F}^{\mathcal{W}}-$adapted because a and Y are. In order to fully determine the firm revenue (see (II.1)) and firm claims (see (II.9)) under cyber attack, we need to characterize the random recovery time (δ_{ij}) introduced in Assumption II.2. Given that we introduce the recovery rate of each group in the SIR mode in Theorem III.1, we have the following assumption.

Assumption IV.4. For $(i,j)\in\mathcal{I}$, we assume that

$$\delta_{ij}=\frac{1}{\gamma_{i,\tau_{ij}}}. \tag{IV.3}$$

This is motivated by the fact that when the recovery times of infected individuals are modeled as independent exponentially distributed random variables, the recovery rate is equal to the inverse of the expected recovery time (see [12]).

Moreover, we define the filtration $\mathbb{G}=(\mathcal{G}_t)_{t\ge0}$ by

$$\mathcal{G}_t=\sigma\left(\mathcal{F}^B_t\cup\mathcal{F}^{\mathcal{W}}_t\cup\sigma\left\{\pi_{ij,s},N_{ij,s}:s\in[0,t]\text{ and }(i,j)\in\mathcal{I}\right\}\right), \tag{IV.4}$$

where $\mathcal{F}^B_t$ and $\mathcal{F}^{\mathcal{W}}_t$ are defined in (II.8) and in (III.5) respectively.

Remark IV.5. It is straightforward that the subunit revenue $(z_{ij})_{(i,j)\in\mathcal{I}}$, the total revenue of the portfolio $\mathbf{O}$, and the firm's claims $(\mathfrak{c}_{i,t})_{1\le i\le H}$, defined respectively in (II.6), (II.7), and (II.9) are $\mathbb{G}$-adapted.

With the knowledge of $(\tau_{ij})_{ij}$ and $(\delta_{ij})_{ij}$, it is now possible to compute the costs of cyber-events at the firm level $(\mathfrak{c}_i)_i$ from (II.10) and at the portfolio level AEP from (II.13). But let us note that other insurance portfolio measures can be studied, and in particular, the impact of reaction and remediation measures as in [22] (see Remark A.1).

From (II.9), we can identify 4 sources of randomness in the cost functions $\mathfrak{c}$, $\mathfrak{C}$, or AEP. The first one is the Brownian motion B governing the economic risk and represented by $\mathcal{F}^B$. The second one is the random coefficients of SIR represented by $\mathcal{F}^{\mathcal{W}}$, the third one is the random arrivals (τ_{ij}) of cyber attacks, and the last one is their random severity π. In fact, all these quantities have a "systemic part" which is $\mathcal{F}$ coming from the cyberattack and an "idiosyncratic/economic part" which are $(N_{ij,t})$, (π_{ij}), $(B_{ij,t})$. If we assume that the insurer ignore the idiosyncratic/economic risk and only deals with the cyber risks i.e. we may focus on $\mathcal{F}^{\mathcal{W}}$ as the only source of randomness and ignore the economic part by taking the expectation conditional to the systemic factor.

This introduces the following version of AEP without the idiosyncratic risk (that we will use in application), a conditional AEP,

$$\widehat{\mathrm{AEP}}_{[t,t+u)}(x) := \mathbb{P}\left(\sum_{i=1}^{H} \widehat{\mathfrak{C}}_{i,[t,t+u)} > x\right). \tag{IV.5}$$

where for $1 \leq i \leq H$ and $t \geq 0$, the expected loss conditionally to the random SIR factors up to to t, $\mathcal{F}_t^{\mathcal{W}}$:

$$\widehat{\mathfrak{C}}_{i,[t,t+\mathrm{d}t]} := \mathbb{E}[\mathfrak{C}_{i,[t,t+\mathrm{d}t]}|\mathcal{F}_t^{\mathcal{W}}],$$

where $\mathfrak{C}_{i,t}$ is defined in (II.11). The following proposition gives an explicit expression of $\widehat{\mathfrak{C}}_{i,[t,t+\mathrm{d}t]}$.

Proposition IV.6. *For $1 \leq i \leq H$ and $t, u \geq 0$,*

$$\begin{aligned}\widehat{\mathfrak{C}}_{i,[t,t+u)} = \pi_\star \sum_{j=1}^{K_i} \frac{z_{ij,0}}{\mu_{ij}} \int_0^{+\infty} \mathbf{1}_{\underline{s}(v)\leq\overline{s}(v)} \\ \times \left(e^{\mu_{ij}\overline{s}(v)} - e^{\mu_{ij}\underline{s}(v)}\right) \mathrm{d}F_i(v),\end{aligned} \tag{IV.6}$$

where F_i in (IV.2)*, and where $\underline{s} : v \mapsto t \vee v$ and $\overline{s} : v \mapsto (t+u) \wedge (v + \frac{1}{\gamma_{i,v}})$.*

Proof. Let $1 \leq i \leq H$ and $t, u \geq 0$. From the explicit expression of the claims between t and $t+u$ given in (II.11), we have:

$$\begin{aligned}\widehat{\mathfrak{C}}_{i,[t,t+u)} &= \mathbb{E}[\mathfrak{C}_{i,[t,t+u)} \mid \mathcal{F}_t^{\mathcal{W}}] \\ &= \mathbb{E}\Bigg[\sum_{j=1}^{K_i} z_{ij,0} \int_{t\vee\tau_{ij}}^{(t+u)\wedge(\tau_{ij}+\delta_{ij})} \pi_{ij} \\ &\qquad \times e^{\left(\mu_{ij}-\frac{1}{2}\sigma_{ij}^2\right)s+\sigma_{ij}B_{ij,s}} \mathrm{d}s \Bigg| \mathcal{F}_t^{\mathcal{W}}\Bigg] \\ &= \sum_{j=1}^{K_i} z_{ij,0} \mathbb{E}\Bigg[\int_{\underline{s}(\tau_{ij})}^{\overline{s}(\tau_{ij})} \pi_{ij} \\ &\qquad \times e^{\left(\mu_{ij}-\frac{1}{2}\sigma_{ij}^2\right)s+\sigma_{ij}B_{ij,s}} \mathrm{d}s \Bigg| \mathcal{F}_t^{\mathcal{W}}\Bigg] \\ &= \sum_{j=1}^{K_i} z_{ij,0} \mathbb{E}[\pi_{ij}] \\ &\qquad \times \mathbb{E}\Bigg[\int_{\underline{s}(\tau_{ij})}^{\overline{s}(\tau_{ij})} e^{\left(\mu_{ij}-\frac{1}{2}\sigma_{ij}^2\right)s+\sigma_{ij}B_{ij,s}} \mathrm{d}s \Bigg| \mathcal{F}_t^{\mathcal{W}}\Bigg]\end{aligned}$$

where the last equality comes from Assumption IV.4. Using the cdf F_i of τ_{ij} in (IV.2), and knowing that $\mathbb{E}[\pi_{ij}] = \pi_\star$ and that (B_{ij}) are independent and independent of $\mathcal{F}_t^{\mathcal{W}}$, we then have

$$\begin{aligned}&\mathbb{E}\Bigg[\int_{\underline{s}(\tau_{ij})}^{\overline{s}(\tau_{ij})} e^{\left(\mu_{ij}-\frac{1}{2}\sigma_{ij}^2\right)s+\sigma_{ij}B_{ij,s}} \mathrm{d}s \Bigg| \mathcal{F}_t^{\mathcal{W}}\Bigg] \\ &\quad = \int_0^{+\infty} \int_{\underline{s}(v)}^{\overline{s}(v)} \mathbb{E}\left[e^{\left(\mu_{ij}-\frac{1}{2}\sigma_{ij}^2\right)s+\sigma_{ij}B_{ij,s}}\right] \mathrm{d}s \, \mathrm{d}F_i(v) \\ &= \int_0^{+\infty} \int_{\underline{s}(v)}^{\overline{s}(v)} e^{\mu_{ij}s} \, \mathrm{d}s \, \mathrm{d}F_i(v) \\ &= \int_0^{+\infty} \mathbf{1}_{\underline{s}(v)\leq\overline{s}(v)} \left(e^{\mu_{ij}\overline{s}(v)} - e^{\mu_{ij}\underline{s}(v)}\right) \mathrm{d}F_i(v),\end{aligned}$$

for all $1 \leq j \leq K_i$. With the definitions of $\underline{s}$ and $\overline{s}$ provided in the proposition, the result follows. □

Finally, we can easily have the expected output of the whole economy $\mathbb{E}[\mathbf{O}_t]$, where $\mathbf{O}_t$ defined in (II.7).

V. NUMERICAL ANALYSIS AND DISCUSSION

This section describes the data for parameter calibration and estimation, outlines the calibration methods and comments on the estimated parameters, and finally presents the simulation methodology and discusses the results.

A. The dataset

1. *Firm's data*

The dataset [30] contains $H = 2,929$ firms located in the Ile-de-France region. For each firm i, we know its sector of activity (using NAF classification as in [25]) as well as its annual revenue from 2010 to 2022. By using the methodology described in Section V B 1, we compute K_i for each firm i and $z_{0,ij}$, μ_{ij} as well as σ_{ij} for each subunit j of firm i. We can check that the number of subunits is distributed according to Zipf law (see Figure 15a and Figure 15b) over the entire dataset. Table I indicates the number of firms per number of subunits and per sector. The maximum firm size is $K = 12$.

2. *Cyber contagion data*

Each quarter, the *GuidePoint Security's Research and Intelligence Team* publishes a Ransomware and Cyber Threat Insights (see [19] for the year 2024). This report summarizes cyber crime incidents, mainly ransomware attacks, that occurred during the year. Data collected are obtained from publicly available

Number of subunits	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Agriculture, forestry and fishing	21	1	0	0	0	0	0	0	0	0	0	0
Mining and quarrying; Water supply; sewerage, and waste management	11	4	0	0	0	1	0	0	0	0	0	0
Manufacturing industry	249	51	20	7	2	2	1	1	3	1	2	0
Construction	433	62	30	11	7	4	3	5	2	1	0	1
Wholesale and retail trade; repair of motor vehicles and motorcycles	619	80	42	18	11	7	4	1	1	1	2	2
Transport and storage	141	32	8	3	8	2	0	1	1	0	0	1
Accommodation and food	141	14	1	0	2	0	0	1	0	1	0	0
Information and communication	67	5	5	4	2	0	3	0	1	0	0	0
Financial and insurance activities	57	2	0	2	1	1	0	0	1	0	0	0
Real estate activities	45	10	2	2	2	0	0	1	0	1	0	0
Professional, scientific, technical , and Administrative and support service	317	31	22	6	1	2	1	0	1	2	0	0
Primarily non-market services	47	12	1	1	0	1	0	2	1	0	0	0
Other service activities	115	8	7	7	2	0	1	1	0	0	0	0
Total	2263	312	138	61	38	20	13	13	11	7	4	4

TABLE I: Number of firms per size (number of subunits) and per sector

resources, including the sites and blogs of threat groups themselves. Figure 1b and Figure 14 summarize, respectively, the most impactful ransomware groups and the most impacted industries in 2024. For this numerical analysis, we calibrate the cyber episode on the **LockBit** ransomware which, according to [19], entered 2024 as the long-standing dominant ransomware group with the highest tempo by victim, but the group faced substantial disruption in the wake of February's international operation Cronos (see [16]).

The database gives the total infected per industry in 2024 for all the ransomware groups, the total infected per group day for each group. However, it does not give the number of infected per company size. We will build a proxy as follows. We first divide the number of attacks per sector by the total number of attacks to get the attack rate per sector in 2024 (see Table 13). By multiplying the latter by the total number of LockBit infections, we obtain the average number of attacks per sector. Then, based on the distribution of subsidiaries across sectors in Table I, we derive an estimate of the number of infected firms of varying sizes (see Figure 3). We are interested in what happens between May 1, 2024 and July 31, 2024, namely $T = 100$ days.

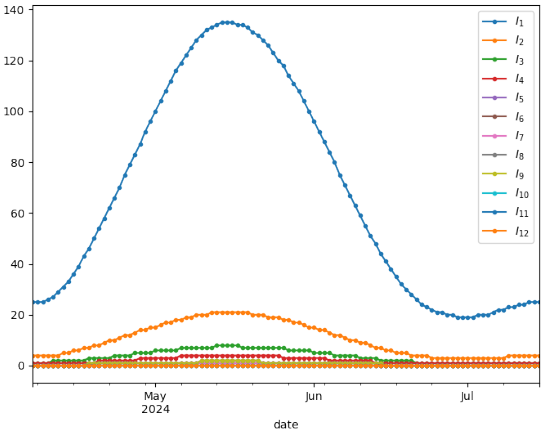


FIG. 3: Number of new firms infected per size each day in may-june-july 2024 (I_k)

B. Calibration procedure

1. Calibration of the firms parameters

For the simulations, the portfolio consists of H firms, with $H = 2,929$. Let us assume that we know their revenue $Z_{i,t}$ between year 1 and year $T^f \in \mathbb{N}^\star$. There are two possibilities: either we know the number of subsidiaries and their revenue over the period, or we have neither. Since the second option is more likely, we will build a proxy for the number of subsidiaries and their revenue. The empirical average of revenue in the database is written as

$$\overline{Z}_t = \frac{1}{H}\sum_{i=1}^{H} Z_{i,t}. \tag{V.1}$$

We are at date 1. We assume that companies with revenues lower than $\overline{Z}_1$ have 1 subsidiary and the others have more than one. This amounts to saying that for $i \in \{1,\dots,H\}$,

$$K_i = \left\lceil \frac{Z_{i,1}}{\overline{Z}_1} \right\rceil \in \mathbb{N}^*, \tag{V.2}$$

where $\lceil x \rceil$ is the smallest integer greater than or equal to x. Therefore, for all $j \in \{1,\dots,K_i\}$, we note

$$z_{ij,t} = \frac{Z_{i,t}}{K_i}. \tag{V.3}$$

Then the estimation of the Black Scholes parameters (II.4), $(\sigma_{ij})_{(i,j)\in\mathcal{I}}$ and $(\mu_{ij})_{(i,j)\in\mathcal{I}}$, is given by

$$\sigma_{ij}^{\text{year}} = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T-1}(r_{ij,t} - r_{ij,*})^2},$$

and

$$\mu_{ij}^{\text{year}} = r_{ij,*} + \frac{1}{2}(\sigma_{ij}^{\text{year}})^2,$$

where $r_{ij,t} := \log \frac{z_{ij,t+1}}{z_{ij,t}}$ for $1 \leq t \leq T$ and $r_{ij,*} := \frac{1}{T-1}\sum_{t=1}^{T-1} r_{ij,t}$. Finally, since cyber events are tracked on a daily frequency, we convert the annual parameters to daily parameters by assuming 365 days. We have

$$\sigma_{ij} = \frac{\sigma_{ij}^{\text{year}}}{\sqrt{365}} \quad \text{and} \quad \mu_{ij}^{\text{year}} = \frac{\mu_{ij}^{\text{year}}}{365}.$$

As shown on Figure 15b, the parameters of the distribution of the number of subunits (Zipf's law introduced in Equation (II.3)) are $\mathfrak{a} = 1.759$ and $q = 0.784$. The calibration of the Geometric Brownian motion dynamics of the subsidiaries (specifically parameters (μ_{ij}) and (σ_{ij})) follows the straightforward approach described in Section V B 1. We use the firms' revenue dataset to calibrate the firms/subunit characteristics. In Table II, we summarize the average drift, the volatility, and the initial revenue, for each size.

2. *Calibration of the contagion parameters*

To calibrate the SIR parameters, we follow the Optimization based on Forward Model Simulation approach proposed by [6], which is detailed below. We require the numbers of susceptibles and recovered individuals in the beginning of the episode, and –most importantly– the daily counts of new infections throughout the epidemic.

From Section V B 1, we determine $K := \max_{i\in\{1,\dots,H\}} K_i$ the maximum size of firms. Given that the cyber episode spans 1 May-31 July 2024, we denote $U = 100$ the number of days of the cyber attack, we have $\{t_0, t_1, \dots, t_U\}$ the set of observation dates. For calibration, we have from the dataset the daily detected infections, $(i_{k,t_u})_{1\leq k\leq K, 0\leq u\leq U}$. We reasonably assume that the initial removed are zero i.e. $(r_{k,t_0} = 0)_{1\leq k\leq K}$. We do not have the initial number of susceptible population previously noted $(s_{k,t_0})_{1\leq k\leq K}$. We could consider that they are model parameters.

But knowing that firm's size of the insurance portfolio follows a Zipf distribution whose parameter is already known, we can simply consider that firm's size of the entire database (denoted $h_{k,t_0} = s_{k,t_0} + i_{k,t_0} + r_{k,t_0}$ for each $k = 1\dots K$) also follows a Zipf law with parameters $\mathfrak{a} = 1.759$ and $q = 0.784$ (see Equation (II.3)). Then, given that $h_\star = \sum_{k=1}^{K} h_{k,t_0}$, determining the total number of firms $h_\star$ is enough because for all $k = 1\dots K$,

$$h_{k,t_0} = \left\lfloor h_\star \frac{\frac{1}{k^{\mathfrak{a}}}}{\sum_{j=1}^{K}\frac{1}{j^{\mathfrak{a}}}} \right\rfloor \tag{V.4}$$

where $\lfloor x \rfloor$ is the greatest integer less than or equal to x. We then consider $h_\star$ as a model parameter.

For the SIR model, the main challenge consists in calibrating the parameters in Ψ which are all stochastic. Recall that for $\varphi \in \Psi$, from (III.10),

$$\tilde{\varphi}_t = \varphi_0 e^{-\kappa_\varphi t} + \mu_\varphi\left(1 - e^{-\kappa_\varphi t}\right) + \Sigma_\varphi \int_0^t e^{-\kappa_\varphi (t-s)}\sqrt{\tilde{\varphi}_s}\, d\mathcal{W}_{\varphi,s}.$$

We note

$$\Theta := \{\varphi_0, \mu_\varphi, \kappa_\varphi, \Sigma_\varphi; \forall \varphi \in \Psi\} \in (\mathbb{R}_+)^{4\times|\Psi|}, \tag{V.5}$$

recalling that for each parameter $\varphi \in \Psi$, κ_φ is the speed of mean reversion, μ_φ is the long-term mean level, Σ_φ is the volatility factor, and φ_0 is the initial condition. Then, for a given $\Theta \in (\mathbb{R}_+)^{4\times|\Psi|}$, we simulate $M \in \mathbb{N}^*$ trajectories on $\{0, 1, \dots, U\}$ i.e. $(\varphi^m_{t_u})_{1\leq m\leq M, 0\leq u\leq U}$ for all $\varphi \in \Psi$. We now calibrate the SIR model parameters. We have at this stage $4 \times (2K + K^2) + 1$ parameters to determine, that is 673 because $K = 12$. This is large relative to the limited data at hand. In order to simplify and speed up the calibration, we make the following additional assumptions.

K_i	1	2	3	4	5	6	7	8	9	10	11	12
z_{ij}^0 (in €million)	4.70	5.99	6.99	6.69	7.30	8.98	7.49	5.40	6.70	6.61	10.03	9.82
$\sigma_{ij} \times 10^3$	12.69	12.30	10.91	9.70	16.16	8.16	19.03	26.31	8.64	12.71	4.92	4.79
$\mu_{ij} \times 10^3$	0.19	0.37	0.28	0.26	0.56	0.24	1.12	1.45	0.36	0.28	0.33	0.27

TABLE II: subunits' characteristics

1. As stated in Theorem III.4, $(b_{j,k})$ is binomial with in-firm attack rate a. Therefore we only need to determine $4 \times (2K+1)$ parameters.

2. All the SIR parameters (which are CIR process) have the same mean reversion speed ($\kappa_\varphi = \kappa_a$) and the volatility coefficient ($\Sigma_\varphi = \Sigma_a$) and the long-term mean level is equal to the initial value ($\mu_\varphi = \varphi_0$). Then we need to determine $2K+4$ parameters.

3. As state in [12], for each size, when the out-firm reproduction number (β_k/γ_k) is independent of the firm size and the infections are parallel, the recovery rate γ_k equals to the inverse of the expected recovery time $\sum_{j=1}^{k} \frac{1}{j}$. We obtain

$$\gamma_k = \frac{\gamma_1}{\sum_{j=1}^{k} \frac{1}{j}} \quad \text{and} \quad \beta_k = \frac{\beta_1}{\sum_{j=1}^{k} \frac{1}{j}}, \tag{V.6}$$

where the last equality assumes that the decay rate of immunity is independent of the firm size. Therefore, it is enough to determine γ_1 and β_1. We therefore need to calibrate six parameters: κ_a, Σ_a, a_0 $\gamma_{1,0}$, $\beta_{1,0}$, and $h_\star$. We now write

$$\Theta := \{\kappa_a, \Sigma_a, a_0, \gamma_{1,0}, \beta_{1,0}, h_\star\}. \tag{V.7}$$

For $1 \leq m \leq M$, we run the SIR model using the following Euler scheme that produces $(S_k^{m,sim}, I_k^{m,sim}, R_k^{m,sim})_{1 \leq k \leq K}$. We write for $0 \leq t \leq U-1$ and $1 \leq k \leq K$,

$$\begin{cases} S_{k,t_{u+1}}^{m,sim} - S_{k,t_u}^{m,sim} = -kY_{t_u}^{m,sim} S_{k,t_u}^{m,sim} \\ \qquad\qquad + Y_{t_u}^{m,sim} \sum_{j=k+1}^{K} jS_{j,t_u}^{m,sim} \cdot b_{jj-k,t_u}^{m}, \\ I_{k,t_{u+1}}^{m,sim} - I_{k,t_u}^{m,sim} = -\gamma_{k,t}^{m} I_{k,t_u} + Y_{t_u}^{m,sim} \sum_{j=k}^{K} jS_{j,t_u}^{m,sim} \cdot b_{jk,t_u}^{m}, \\ R_{k,t_{u+1}}^{m,sim} - R_{k,t_u}^{m,sim} = \gamma_{k,t_u}^{m} I_{k,t_u}^{m,sim}, \end{cases} \tag{V.8}$$

where

$$Y_{t_u}^{m,sim} = \frac{1}{N_0} \sum_{k=1}^{K} \beta_{k,t_u}^{m} \cdot kI_{k,t_u}^{m,sim},$$

and $I_{k,t_0}^{m,sim} = I_{k,t_0}$, $R_{k,t_0}^{m,sim} = 0$ and $S_{k,t_0}^{m,sim} = S_{k,t_0}$. The optimal parameters Θ are the ones that minimize the mean square error J_2 defined as

$$J_2(\Theta) := \frac{1}{M} \sum_{m=1}^{M} J_2^m(\Theta), \tag{V.9}$$

with $J_2^m(\Theta) := \sum_{k=1}^{K} \sum_{t=0}^{U} \left| I_{k,t} - I_{k,t}^{m,sim} \right|^2$. The procedure is summarized in Algorithm 1.

Algorithm 1 SIR Calibration

```
procedure SIR Calibration($K, U, (I_{k,t_u}), \mathfrak{a}, q, M$)
    Input: Maximum firm size $K$, horizon of the cyber-episode $U$, number of infected $(i_{k,t_u})_{1 \leq k \leq K, 0 \leq u \leq U}$, Zipf law parameters $\mathfrak{a}, q$.
    for all $\Theta = \{\kappa_a, \Sigma_a, a_0, \gamma_{1,0}, \beta_{1,0}, h_\star\}$ do
        Calculate the number of firms per size $(h_{k,t_0})$ using (V.4), then the initial number of susceptible using $(s_{k,t_0} = h_{k,t_0} - i_{k,t_0})$
        Calculate the total population $N_0 = \sum_{l=1}^{K} k(S_{k,t_0} + I_{k,t_0})$.
        Convert numbers of susceptible into susceptible rates $S_{k,t_0} = \frac{s_{k,t_0}}{\sum_{l=1}^{K} h_{k,t_0}}$, and numbers of infected into infection rates $I_{k,t_u} = \frac{i_{k,t_u}}{\sum_{l=1}^{K} h_{k,t_0}}$.
        Simulate $M$ trajectories $(\varphi_t^m)_{1 \leq m \leq M, 0 \leq t \leq U}$ for each $\varphi \in \Psi$ on $\{0, 1, \ldots, U\}$.
        for all $(\varphi_{t_u}^m)_{\varphi \in \Psi, 0 \leq u \leq U}$ do
            Simulate a forward SIR model with parameters $(\varphi_{t_u}^m)$ and initial condition $I_{k,t_0}^{m,sim} = I_{k,t_0}$, $R_{k,t_0}^{m,sim} = 0$, $I_{k,t_0}^{m,sim} = I_{k,t_0}$, as well as $N_0$, and obtain daily time series $(S_k^{m,sim}, I_k^{m,sim}, R_k^{m,sim})_{1 \leq k \leq K}$.
            Compute the objective function $J_2^m(\Theta)$.
        end for
        Compute the objective function $J_2(\Theta)$.
    end for
    Determine the minimum and the minimizer $\Theta$ of $J_2$.
    Output: the parameters $\Theta = (\kappa_a, \Sigma_a, a_0, \gamma_{1,0}, \beta_{1,0}, h_\star)$.
end procedure
```

Table III summarizes the 6 calibrated parameters.

The total number of firms $h_\star$ to calculate the initial population in each group summarized in Table IV.

Initial recovery rate	$\gamma_{1,0}$	0.6782
Initial "out–firm" infection rate	$\beta_{1,0}$	0.5471
Initial in-firm attack rate	a_0	0.3466
Volatility	Σ_φ	0.0151
Mean reversion speed	μ_φ	0.4474
Total population of firms	$h_\star$	14,210

TABLE III: The SIR model parameters

C. Estimation procedure of revenues and claims, simulations, and discussion

We would like to calculate the Aggregate Exceedance Probability at each time u, with $u \in \{0, 1, \ldots, T\}$, as introduced in (IV.5) i.e. for $x \geq 0$,

$$\widehat{\mathrm{AEP}}_{[u,u+1)}(x) = \mathbb{P}\left(\sum_{i \in \mathcal{I}^\star_{[t,t+u)}} \widehat{\mathfrak{C}}_{i,[u,u+1)} > x\right),$$

where the set of infected $\mathcal{I}^\star_{[t,t+u)}$ is defined in (II.12).

For a set of SIR parameters Θ, we simulate M trajectories $(\Psi^m)_{1\leq m\leq M}$ ($M = 10,000$ in this numerical analysis) of the parameters Ψ using an exact simulation method of CIR as in [1]. For each trajectory m of parameters, we simulate the cyber contagion using the Euler scheme for the SIR described in (V.8), with initial condition $S^{m,sim}_{k,t_0} = S_{k,t_0}$, $I^{m,sim}_{k,t_0} = I_{k,t_0}$, and $R^{m,sim}_{k,t_0} = 0$ for $1 \leq k \leq K$. We obtain for each scenario $1 \leq m \leq M$, at each date $0 \leq u \leq T$, and for each group $1 \leq k \leq K$, the number of susceptible $S^{m,sim}_{k,u}$, of infected $I^{m,sim}_{k,u}$, and of recovered $R^{m,sim}_{k,u}$. By using $I^{m,sim}_{k,u}$, $\gamma^{m,sim}_{k,u}$ and N_0, we calculate the force of epidemics $Y^{m,sim}_u$.

The next step is to simulate the arrival of cyber attack on each firm (and on each subunit). We fix the scenario $1 \leq m \leq M$. For each firm $1 \leq i \leq H$, we simulate $(N^{m,sim}_{ij,u})_{1\leq j\leq K_i, 0\leq u\leq T}$ as follows:

1. Primary (or out-firm) infections: for each subunit $1 \leq j \leq K_i$, we simulate the first jump of the Cox process $(N^{0,m,sim}_{ij,u})_{0\leq u\leq T}$ with intensity $(Y^{m,sim}_u)_{0\leq u\leq T}$.
2. Determine the time of the very first cyber attack in the firm i, $\tau^{m,sim}_i$ using (IV.1) with the convention $\tau^{m,sim}_i = T + 1$ if $\sum_{j=1}^{K_i} N^{0,m,sim}_{ij,T} = 0$. This means that firm i is not suffering from a primary infection. If $K_i = 1$, then $\tau^{m,sim}_{i,j} = \tau^{m,sim}_i$.
3. If $K_i \geq 2$ and $\tau^{m,sim}_i \leq T$ (that is firm i has many subunits and at least one is externally infected), then secondary (or in-firm) infections can occur: for each $1 \leq j \leq K_i$, if $N^{0,m,sim}_{ij,\tau^{m,sim}_i} = 0$ (that is subunit j is not yet infected), simulate a random variable $U^{m,sim}_{ij}$ with success parameter $a^{m,sim}_{\tau^{m,sim}_i}$
   - if $U^{m,sim}_{ij} = 1$, set $\tau^{m,sim}_{i,j} = \tau^{m,sim}_i$,
   - else, set $\tau^{m,sim}_{i,j} = T + 1$ (subunit j is not suffering from either a primary infection or a secondary infection).
4. If $K_i \geq 2$ and $\tau^{m,sim}_i = T + 1$, $\tau^{m,sim}_{i,j} = +\infty$ for all $1 \leq j \leq K_i$.

Simulating also M realizations of $(\pi^{m,sim}_{ij})_{1\leq m\leq M}$ of the random variables π_{ij} for $(i,j) \in \mathcal{I}$, we compute the Monte Carlo approximation of the total expected revenue at each date of the cyber episode using the last equality of (II.7).

$$\mathbb{E}[\mathbf{O}_u] \approx \sum_{i=1}^{H}\sum_{j=1}^{K_i} z_{ij,0}e^{\mu_{ij}u}\left(1 - \frac{1}{M}\sum_{m=1}^{M}\pi^{m,sim}_{ij}\mathbf{1}_{\tau^{m,sim}_{i,j} \leq u < \tau^{m,sim}_{i,j} + \frac{1}{\gamma^{m,sim}_{i,\tau^{m,sim}_{i,j}}}}\right), \tag{V.10}$$

for each $0 \leq u \leq T$. The simulated set of infected firms in each period $[u, u+1)$ with $0 \leq u \leq T-1$ (defined in (II.12)), is $\mathcal{I}^{\star,m,sim}_{[u,u+1)} = \left\{ i \text{ if } \sum_{j=1}^{K_i} \mathbf{1}_{u \leq \tau^{m,sim}_{i,j} < u+1} \geq 1 \right\}$.

We then compute the Monte Carlo approximation of the expected claim for each firm, in order to obtain in a second step the proxy of AEP defined in (IV.5). The total claims of firm i for each scenario m and at each date u is

$$\mathfrak{C}^{m,sim}_{i,[u,u+1)} = \sum_{j=1}^{K_i} z_{ij,0}\pi^{m,sim}_{ij}\int_{u\vee\tau^{m,sim}_{i,j}}^{(u+1)\wedge\left(\tau^{m,sim}_{i,j} + \frac{1}{\gamma^{m,sim}_{i,\tau^{m,sim}_{i,j}}}\right)} e^{\mu_{ij}s}\mathrm{d}s, \tag{V.11}$$

for all $i \in \mathcal{I}^{\star,m,sim}_{[u,u+1)}$. We finally obtain the empirical $\widetilde{\mathrm{AEP}}$ at each date u by

$$\widetilde{\mathrm{AEP}}^M_u(x) := \frac{1}{M}\sum_{m=1}^{M}\mathbf{1}\left\{\sum_{i\in\mathcal{I}^{\star,m,sim}_{[u,u+1)}} \mathfrak{C}^{m,sim}_{i,[u,u+1)} > x\right\} \tag{V.12}$$

where $\mathbf{1}_{(X>x)}$ is the indicator function that equals 1 if $X > x$, and 0 otherwise.

Algorithm 2 details the procedure to compute the $\widehat{\mathrm{AEP}}$.

1. *The contagion model*

After calibrations, we perform $M = 10,000$ simulations of the SIR model. In order to compare our

	H_1	H_2	H_3	H_4	H_5	H_6	H_7	H_8	H_9	H_{10}	H_{11}	H_{12}
Size	11144	1646	538	244	132	80	52	36	26	20	15	12

TABLE IV: The initial population per group

Algorithm 2 Calculate AEP

```
procedure AEP($x, T, M, \alpha^\pi, \beta^\pi, (\varphi_0, \kappa_\varphi, \Sigma_\varphi), (S_{k,t_0}, I_{k,t_0}, R_{k,t_0})$)
    Input: quantile $x$, horizon $T$, number of simulations $M$, the set of parameters of the SIR $(\varphi_0, \mu_\varphi, \kappa_\varphi, \Sigma_\varphi)_{\varphi\in\Psi}$, the initial values of the SIR $(S_{k,t_0}, I_{k,t_0}, R_{k,t_0})_{1\leq k\leq K}$, parameters of severity $(\alpha^\pi, \beta^\pi)$, and the parameters of firms $(z_{ij,0}, \mu_{ij}, \sigma_{ij})_{(i,j)\in\mathcal{I}}$.
    for all Scenario $1 \leq m \leq M$ do
        From $(\varphi_0, \mu_\varphi, \kappa_\varphi, \Sigma_\varphi)_{\varphi\in\Psi}$, simulate a trajectory of the set of parameters $\Psi$.
        Simulate a SIR model with parameter $\Psi$ and initial condition $(I_{k,t_0}, I_{k,t_0}, R_{k,t_0})_{1\leq k\leq K}$.
        Simulate the first infection times.
        Simulate a trajectory of severity with $(\alpha^\pi, \beta^\pi)$.
        for all Date $1 \leq u \leq T$ do
            Calculate the set of infected firms at time $u$.
            Calculate the total claims at $u$ using (V.11).
        end for
        Calculate the AEP at $u$ from (V.12).
    end for
    Output: The approached aggregate exceedance probability AEP.
end procedure
```

result with the observed LockBit attacks on Figure 3, we quantify the total number of infected subunits. Specifically, we take into account the dynamics of the total subunits infected, the peak as defined in (III.12), and the date that it is reached.

Figure 4 provides (in blue) the 99% confidence interval of the peak in the first $T = 100$ days, computed on $M = 10,000$ trajectories. Both the observed trajectory and the mean simulated trajectories, start from the same initial condition (32 firms, or equivalently 49 infected subunits), and they reach their peak at exactly the same date. This is compared with the Lockbit trajectory (in black). Our results overestimate by 17.15% the height of the peak. The latter is in fact the maximum of the averaged trajectory. Now, we consider the average of the peaks instead of the peak of the average. In Figure 5a, we plot the histogram of the peak. We obtain a peak that is very close to the observed one: the model gives 317 subunits infected when the observed is 273. Our model overestimates the peak by 16.12%.

Another interesting quantity is the date the peak is reached. We can see on Figure 5b that the data say that the peak is reached after 38 days while our model also predicts 38 days, with a maximum absolute deviation of 12 days.

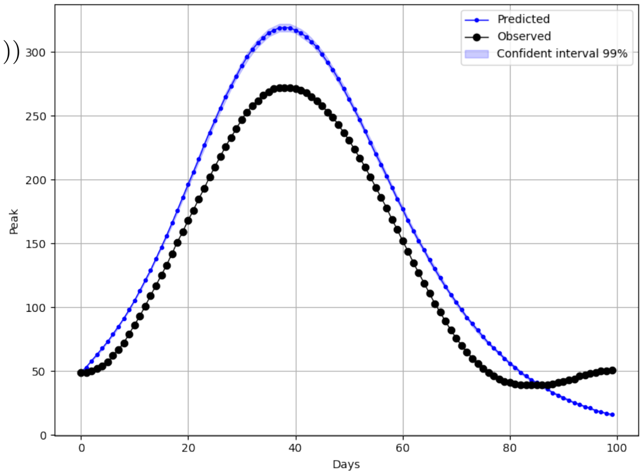


FIG. 4: The dynamics of the total number of infected subunits with confident interval

As in the function $\mathfrak{c}_2$ introduced by [22], the peak (and the date it is reached) can be used to model the limited capacity of the insurance company to respond to an incident. If the number of policyholders needing help is too large, the insurer's assistance teams can become overloaded. Furthermore, it is preferable to slightly overestimate both the height and the date at which it is reached rather than underestimate them, thereby yielding conservative risk measures.

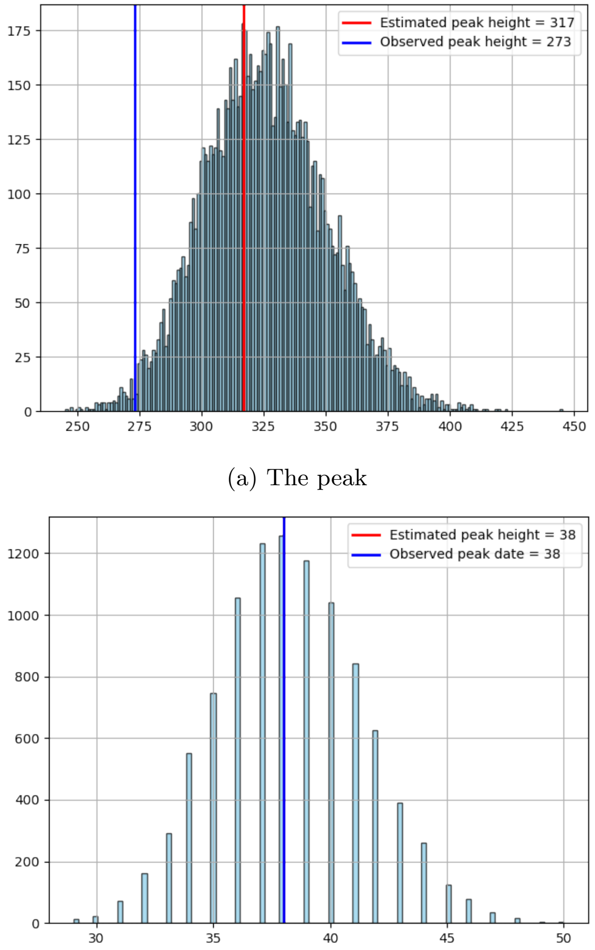
We are now looking at the evolution of the number of susceptible, infected, and recovered individuals. In Figure 16, we plot the evolution of new susceptibles, infected, and recovered.

Regarding the susceptibles (see Figure 16a), infected (see Figure 16b), and recovered (see Figure 16c) populations, the dynamics are the same but different groups reached their peak's at different date. As above, we have, in Table V, the date of the peak per firm size.

For each group, the epidemic dynamics exhibit a consistent temporal ordering. For firms of size 1, the peaks are closest or equals to the peak of epidemics (i.e. the maximum of the total number of infected subunits

$k =$	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
New susceptibles height date	41	37	33	28	22	16	11	0	0	0	0	0
Infected height date	38	40	39	35	32	28	26	0	0	0	0	0
Recovered height date	38	40	39	35	32	28	26	0	0	0	0	0

TABLE V: The date of the peak's height per size



(a) The peak

(b) The date when the peak is reached

FIG. 5: Histograms

for all the size). This is unsurprising, given that this group accounts for 78% of companies. For firms of size 2 to 8, the peak in the number of infected and recovered individuals occur at the same time. This is consistent because the recovery rate is particularly high ($\gamma_{1,0} = 0.678$ in Table III) which, according to [12], is equal to the inverse of the expected recovery time. Regarding the peak in the susceptible population for each size, it occurs before the peak of the total infected population. The peak decreases as size increases, indicating a collapse of the epidemics in the larger organizational units first. This is partly due to the structure of the model, in which large firms feed into small ones. Moreover, since firm's size follows Zipf's law, there are much less firm with big size than small size. In our example, there are precisely 1,000 times more size 1 firms than size 12 firms. Thus, large companies very quickly reach their peak (right from the start of the epidemic here).

Additional analyzes of the contagion model—especially regarding parameter sensitivity—can be conducted, drawing inspiration from [12].

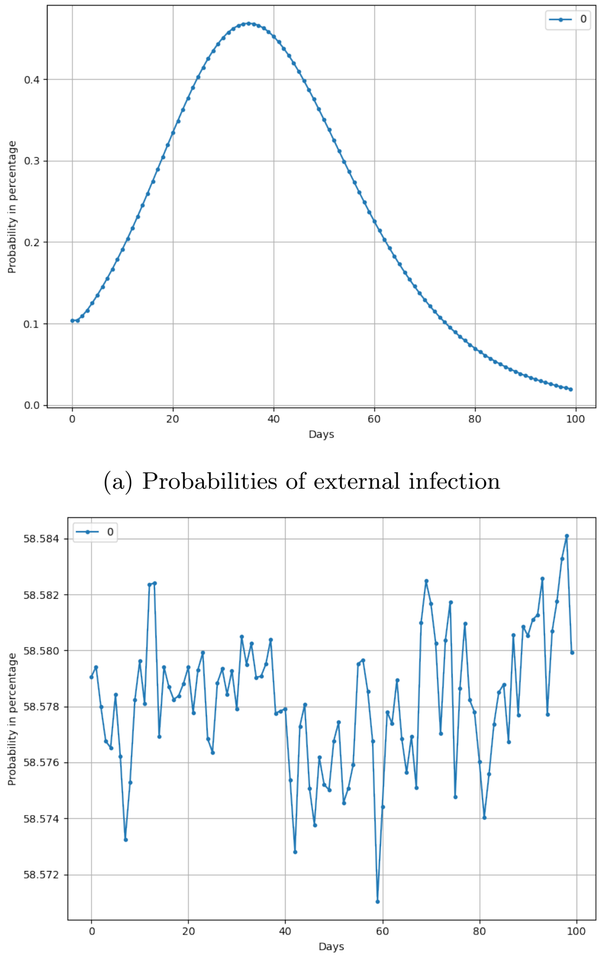
We now turn to the probabilities of internal and external contagion. Figure 6a shows, for each day of the epidemic, the probability that a subunit's infection originates outside the firm. This probability is low, though not zero, and it mirrors the epidemic dynamics, with a peak around day 38. By contrast, the probability that a subunit's infection comes from another subunit within the same firm , in Figure 6a, is high (> 58%) and, as assumed, independent of the epidemic's progression.

In summary, the likelihood that a firm becomes infected is small; however, once infection occurs, there is a significant chance that multiple subunits will be affected.

2. *Impact of the contagion model on a given firm*

This section aims to examine the broader impact of contagion on the economy (see (II.7)) or its representative firms (i.e. the arithmetic mean of firms per size $1 \leq k \leq K$, $\overline{Z}_t^{(k)} := \frac{\sum_{i=1}^{H} \mathbf{1}_{\{K_i=k\}} Z_{i,t}}{\sum_{i=1}^{H} \mathbf{1}_{\{K_i=k\}}}$). It is important to recall that contagion effects stem from the organizational structure of the firms within the portfolio – specifically, the distribution of firm sizes and the presence of subsidiary networks. While current data constraints preclude the estimation of firm-specific or sectoral cost differences, the model is designed with the flexibility to incorporate these factors should the data become available.

As introduced in Theorem II.2, when a subunit undergoes a cyber attack, its revenue negatively jumps by $\pi \sim \mathcal{B}(\alpha^\pi, \beta^\pi)$. For the following simulations, we assume



(a) Probabilities of external infection

(b) Probability of secondary infection

FIG. 6: Probabilities of infection

$\alpha^\pi = 50$ and $\beta^\pi = 10$ giving a mean loss of 0.833 with a standard deviation of 0.048. We start by calculating K_i for each firm i and $z_{0,ij}$, μ_{ij}, σ_{ij} for each subunit j of firm i. Then we run $M = 10,000$ scenarios corresponding to the simulation of the first arrival times of the cyber event τ_{ij} and of the severity π_{ij}.

We plot on Figure 7 the probability density function of the first infection time for different firm size i.e. $(\tau_i)_{1\leq i\leq 12}$. For reasons of scale and because we are interested in the first 100 days of the epidemic, we set the probability of $\tau_i \geq 100$ in Table VI.

From Table VI, we can state the larger the firm, the lower the probability that there will be no infections after 100 days . From Figure 7, we see that the distribution mode – the point at which the probability density attains its maximum – is reached well before the peak of the epidemic (which only occurs on day 38). This mode

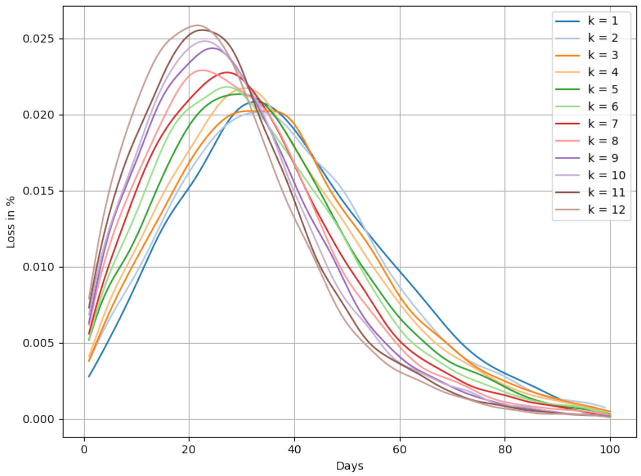


FIG. 7: PDF of the first infection time

naturally decreases with the size of the company. This means that the larger the size, the greater the chance that the firm will be infected before the peak is reached. Therefore, our model reproduces the stylized fact of cyber-episodes mentioned in [5, 27].

For each group $k = 1, \ldots, 12$, we plot on Figure 8 the average daily firm's claim as a percentage of the daily revenue.

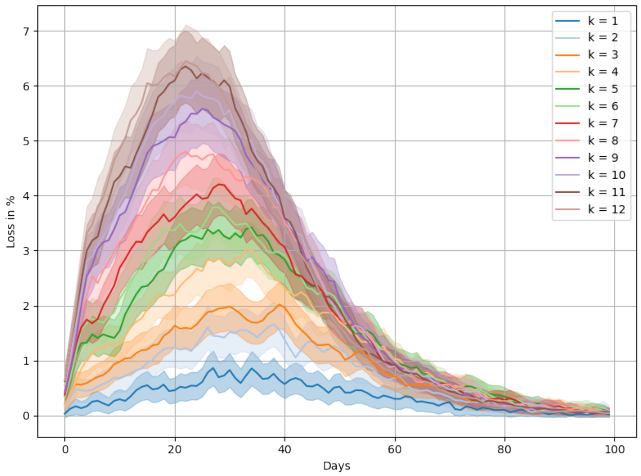


FIG. 8: Firm's revenue loss

As expected, for all sizes, the potential loss of firm is naturally correlated with the date of initial infection. Furthermore, the losses increase with the number of subunits, clearly implying that size is an aggravating factor. This is due to internal contamination between subsidiaries, so the probability is close to 60%.

We have seen that the average daily loss, whose the maximum is 6%, appears to be low. However, this masks heterogeneity: instead of averaging over 10,000

$k =$	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Recovered height date	0.794	0.623	0.500	0.394	0.310	0.249	0.194	0.150	0.129	0.095	0.069	0.060

TABLE VI: Probability of no infection ($\tau_i > 100$)

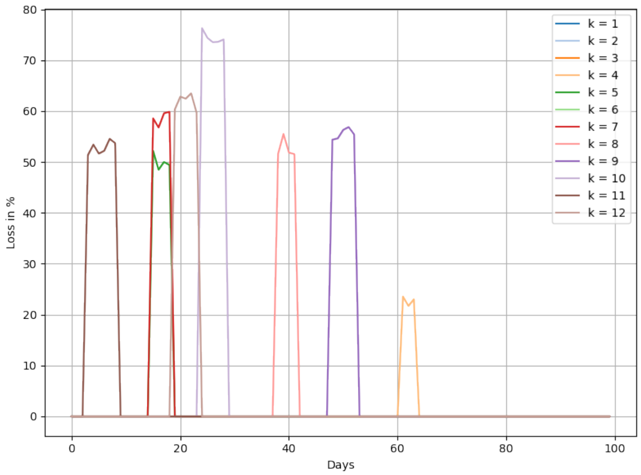


FIG. 9: Firm's revenue loss due to one attack

simulations, let us consider a single contagion scenario. We then obtain, as in Figure 9, that even if attacks are sporadic, the claim relative to the daily revenue is considerable and sometimes reaches 75%. Hence, the need for insurance. And, of course, the duration of the infection increases with the size of the company (recalling the random variable $\delta_{ij} = \frac{1}{\gamma_{i,\tau_{ij}}}$ defined in (IV.3) where γ is defined in (V.6)).

3. *Impact on the contagion model on an insurance portfolio*

In this section, we want to calculate the Aggregate Exceedance Probability (AEP) as described in Algorithm 2. Here, we focus (among other reasons to speed up simulations) on companies with a size greater than or equal to 2 ($K_i \geq 2$). This therefore concerns 621 firms whose total revenue on the first day of the cyber-episode is €38.38 millions.

a. Daily probability density function (pdf) of all the portfolio We plot both the probability density function of the severity of the portfolio's daily loss $\widehat{\mathfrak{C}}$ (and then the daily AEP curve on different dates): 7 days after the beginning of the epidemic, 14 days before the peak height, the date of peak height (day n°38), 14 days after the peak height, and 100 days after the beginning of the epidemic.

In Figure 10 and Table VII, we have the pdf of daily total severity i.e. $\sum_{i=1}^{621} \mathfrak{C}_{i,[u,u+1)}$. It is established that the closer we are to the peak of the cyber event, the wider and flatter the total claims distribution is (with a slight advantage for day 24, when the distribution mode of the firm's first infection time is highest). Moreover, the support of the distribution increases when we are close to the peak of the cyber contagion. Also, around the latter, the mean and the mode of the distribution are greater compared to on all other days. It can also be noted that the mode of the distribution 14 days after the peak is small than that 14 days before. To summarize, the cyber event induces a deformation in the distribution of total losses or compensations: as close we are to the first infection time, the probability density function becomes increasingly flattened, reflecting a broader dispersion around the mode.

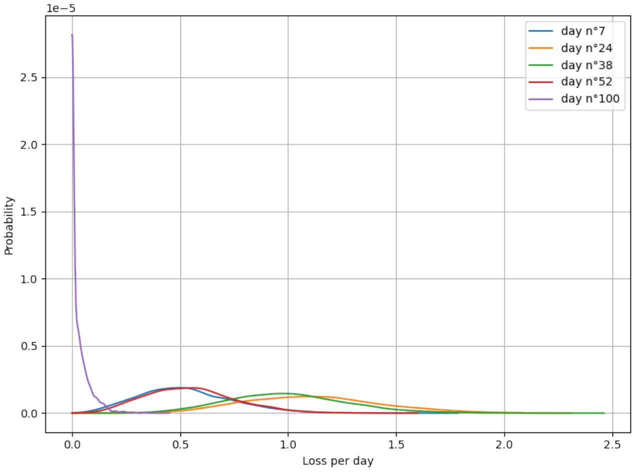


FIG. 10: PDF of the daily severity

Day	**7**	**24**	**38**	**52**	**100**
Lower bound	0	0	0	0	0
Upper bound	1.488	2.275	2.244	1.531	0.388
Mean	0.744	1.138	1.122	0.586	0.194
Mode	0.472	1.013	0.892	0.464	0.001

TABLE VII: Characteristics of the distribution's support (in million of €)

The losses in abscises is in millions euros and should be compared to the daily total revenue of a company in the portfolio, i.e. €38.38 million at day 0. In other words, we could expect a daily claim of up to 5.92% of €38.38 million. Let us clarify this by studying the tails of distributions using the AEP.

b. Aggregate Exceedance Probability (AEP) of the whole portfolio We calculate here the probability of the total daily claims from €1 million (2.6% of the total revenue on day 0) to €2.5 millions (6%) using the approached expression of AEP given in (V.12).

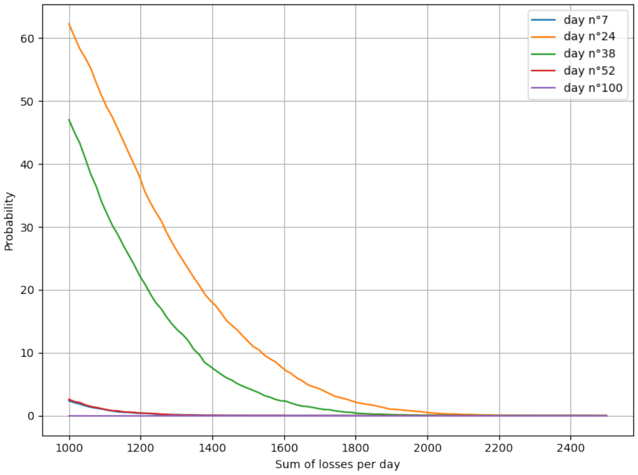


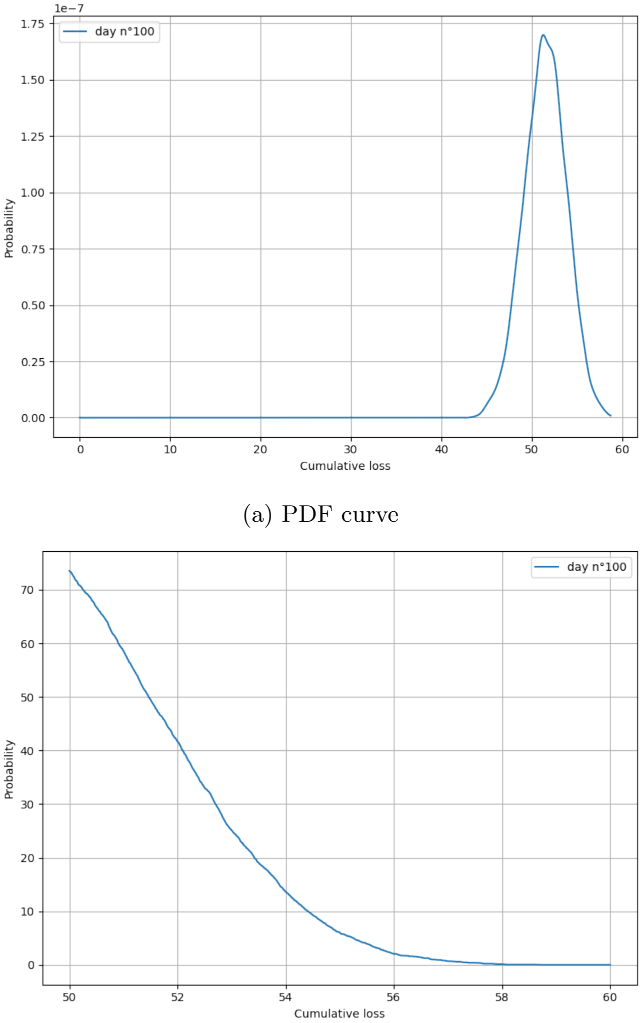
FIG. 11: AEP curve (the sum of losses are in €million)

We obtain the AEP curve in Figure 11 by realizing $M = 10,000$ scenarios of the epidemic i.e. of the daily total claims. As anticipated from the shape of the probability density function, the AEP curve reveals that the range of cumulative daily losses expands with the severity of the cyber event. As expected, at day 24 when the mode of first arrival time is reached, the probability that the sum of daily claims exceeds €1 million reaches 60%, while it remains close to zero in the initial and terminal phases of the epidemic.

To conclude, we analyze the total loss incurred by the whole portfolio of 621 firms, with size greater or equal, over the 100-day duration of the cyber-episode.

The pdf curve in Figure 12a suggests that the total claims after 100-day may range approximately up to €60.44 million, with a median greater than €51.47 million and a mode equals to €51.43 million. Moreover, the AEP curve in Figure 12b represents the right tail of the total claims after 100 days. It indicates that this amount will be higher than €52 million with a probability around 50%. Given a total revenue of €38.38 million, This would mean that there is a 50% chance that total compensation after 100 days will represent up to two days' revenue for all companies in the portfolio.

We can push the analyses further by calculating the AEP and pdf by activity's sector, by company size, and according to many other factors. Ultimately, all depends on how the insurance company organizes its portfolio. Furthermore, the frequency of aggregation can be adjusted to examine how losses evolve on a weekly or monthly basis.



(a) PDF curve

(b) AEP curve

FIG. 12: Financial impact after 100 days of the epidemic

VI. CONCLUSION

We started from stylized facts in economics and cybersecurity. We then assume that firms in an economy are composed of a number of independent subunits and that in a cyberattack episode, the environment modulates the contagion, transmission, and recovery rates – rendering them stochastic. We thus propose a stochastic multigroup SIR model coupled with a granular model of firm growth to describe the propagation of cyberattacks into a cyber-insurance portfolio. We use it to quantify the impact of such an event both on firms' revenue dynamics and on an insurance portfolio. The model shows that the infection of a subunit most likely originates from another subunit within the same firm, not from outside. Moreover, larger firms experience more internal transmission, leading to greater losses and higher insurance claims. This extends the previous work of [22, 23] and provides a more comprehensive framework for assessment of the impact of a cyber event on an economy or an insurance portfolio, under multiple scenario configurations.

A possible extension will be when dealing with a long-lasting malware epidemic, a recovered firm can become susceptible/infected again after a certain period of time. Therefore, using a SIRS model instead of SIR model. We finally assume that the severity is independent and identically distributed. However, it may well depend on on the dynamics of the epidemic or even on the firm's investments in cybersecurity.

[1] Alfonsi, A. (2015). Simulation of the CIR process. In Affine Diffusions and Related Processes: Simulation, Theory and Applications, pages 67–92. Springer.

[2] Allen, E. (2016). Environmental variability and mean-reverting processes. Discrete Contin. Dyn. Syst. Ser. B, 21(7):2073–2089.

[3] AXA Group (2024). Future risks report 2024. Accessed: 2025-09-29.

[4] Axtell, R. L. (2001). Zipf distribution of us firm sizes. science, 293(5536):1818–1820.

[5] Baksy, A., Caratelli, D., and Olson, L. M. (2025). Cyberattacks and firm size: The vulnerability of mid-size firms. Office of Financial Research, The OFR Blog.

[6] Balan, R. (2023). Lecture 9: Full calibration of SIR models. Lecture notes (math 420), Department of Mathematics, University of Maryland. Version: February 23, 2023.

[7] Bessy-Roland, Y., Boumezoued, A., and Hillairet, C. (2021). Multivariate hawkes process for cyber insurance. Annals of Actuarial Science, 15(1):14–39.

[8] Boumezoued, A., Cherkaoui, Y., and Hillairet, C. (2025). Cyber risk frequency modelling using hawkes processes: Calibration on attack and vulnerability data. https://hal.science/hal-05305048v1.

[9] Bouzalmat, I., El Idrissi, M., Settati, A., and Lahrouz, A. (2023). Stochastic sirs epidemic model with perturbation on immunity decay rate. Journal of Applied Mathematics and Computing, 69(6):4499–4524.

[10] Caraballo, T., Colucci, R., et al. (2017). A comparison between random and stochastic modeling for a SIR model. Communications on Pure and Applied Analysis, 16(1):151–162.

[11] Cox, J. C., Ingersoll, J. E., Ross, S. A., et al. (1985). A theory of the term structure of interest rates. Econometrica, 53(2):385–407.

[12] Doenges, P., Götz, T., Kruchinina, N., Krüger, T., Niedzielewski, K., Priesemann, V., and Schäfer, M. (2024). SIR model for households. SIAM Journal on Applied Mathematics, 84(4):1460–1481.

[13] Edwards, B., Hofmeyr, S., and Forrest, S. (2016). Hype and heavy tails: A closer look at data breaches. Journal of Cybersecurity, 2(1):3–14.

[14] Eling, M. and Loperfido, N. (2017). Data breaches: Goodness of fit, pricing, and risk measurement. Insurance: mathematics and economics, 75:126–136.

[15] European Central Bank (2025). Cyber resilience stress testing: A macroprudential approach. Macroprudential Bulletin, Issue 22.

[16] Europol (2024). Law enforcement disrupt world's biggest ransomware operation. Accessed: 2025-11-27.

[17] Farkas, S., Lopez, O., and Thomas, M. (2021). Cyber claim analysis using generalized pareto regression trees with applications to insurance. Insurance: Mathematics and Economics, 98:92–105.

[18] Grossi, P., Kunreuther, H., and Patel, C. C. (2005). Catastrophe modeling: a new approach to managing risk, volume 25. Springer Science & Business Media.

[19] GuidePoint Security (2025). GRIT 2025 ransomware & cyber threat report. Technical report, GuidePoint Security. Accessed: 2025-08-15.

[20] Guo, H., Li, M. Y., and Shuai, Z. (2006). Global stability of the endemic equilibrium of multigroup SIR epidemic models. Canadian applied mathematics quarterly, 14(3):259–284.

[21] Herskovic, B., Kelly, B., Lustig, H., and Van Nieuwerburgh, S. (2020). Firm volatility in granular networks. Journal of Political Economy, 128(11):4097–4162.

[22] Hillairet, C. and Lopez, O. (2021). Propagation of cyber incidents in an insurance portfolio: counting processes combined with compartmental epidemiological models. Scandinavian Actuarial Journal, 2021(8):671–694.

[23] Hillairet, C., Lopez, O., d'Oultremont, L., and Spoorenberg, B. (2022). Cyber contagion: impact of the network structure on the losses of an insurance portfolio. Insurance: Mathematics and Economics.

[24] Homer, D. and Li, M. (2017). Notes on using property catastrophe model results. In Casualty Actuarial Society E-Forum, volume 2, pages 1–15.

[25] Insee (2025). French classification of activities. Accessed on 27 November 2025.

[26] Ji, C., Jiang, D., and Shi, N. (2011). Multigroup SIR epidemic model with stochastic perturbation. Physica A: Statistical Mechanics and its Applications,

390(10):1747–1762.

[27] Kamiya, S., Kang, J.-K., Kim, J., Milidonis, A., and Stulz, R. M. (2021). Risk management, firm reputation, and the impact of successful cyberattacks on target firms. Journal of Financial Economics, 139(3):719–749.

[28] Lin, Y., Jiang, D., et al. (2013). Long-time behaviour of a perturbed SIR model by white noise. Discrete Contin. Dyn. Syst. Ser. B, 18(7):1873–1887.

[29] Moran, J., Secchi, A., and Bouchaud, J.-P. (2024). Revisiting granular models of firm growth. arXiv preprint arXiv:2404.15226.

[30] Pappers (2025). Pappers : Toute l'information gratuite sur les entreprises en france. Consulté le 26 novembre 2025.

[31] Peng, C., Xu, M., Xu, S., and Hu, T. (2017). Modeling and predicting extreme cyber attack rates via marked point processes. Journal of Applied Statistics, 44(14):2534–2563.

[32] Stanley, M. H., Amaral, L. A., Buldyrev, S. V., Havlin, S., Leschhorn, H., Maass, P., Salinger, M. A., and Stanley, H. E. (1996). Scaling behaviour in the growth of companies. Nature, 379(6568):804–806.

[33] The Geneva Association (2023). Cyber accumulation risk: Threat landscape and risk management. Accessed: 2025-09-29.

[34] Van den Driessche, P. and Watmough, J. (2002). Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Mathematical biosciences, 180(1-2):29–48.

[35] Wyart, M. and Bouchaud, J.-P. (2003). Statistical models for company growth. Physica A: Statistical Mechanics and its Applications, 326(1-2):241–255.

[36] Zeller, G. and Scherer, M. (2022). A comprehensive model for cyber risk based on marked point processes and its application to insurance. European Actuarial Journal, 12(1):33–85.

[37] Zhou, Y. and Zhang, W. (2016). Threshold of a stochastic SIR epidemic model with lévy jumps. Physica A: Statistical Mechanics and Its Applications, 446:204–216.

FUNDING

This research is part of the CyFi (Cyber Financialization) project, in cooperation with the cyber risk quantification tech startup *Citalid* and with the financial support of *Bpifrance.*

Appendix A: Another version of the SIR

The dynamical system governing the dynamics of the susceptible, infected and recovered firms of size k reads as

$$\begin{aligned}
\mathrm{d}S_{k,t} &= Y_t\left(-kS_{k,t} + \alpha_k R_{k,t} + \sum_{j=k+1}^{K} jS_{j,t}\cdot b_{j,j-k}\right)\mathrm{d}t \\
&\quad - \sigma_k I_{k,t}S_{k,t}\mathrm{d}\mathcal{W}_{k,t}, \\
\mathrm{d}I_{k,t} &= \left(-\gamma_k I_{k,t} + Y_t\sum_{j=k}^{K} jS_j(t)\cdot b_{j,k}\right)\mathrm{d}t + \sigma_k I_{k,t}S_{k,t}\mathrm{d}\mathcal{W}_{k,t}, \\
\mathrm{d}R_{k,t} &= \left[\gamma_k I_{k,t} - \alpha_k R_{k,t}\right]\mathrm{d}t,
\end{aligned}$$

where $Y_t = \frac{1}{N_0}\sum_{k=1}^{K}\beta_k\cdot kI_{k,t}$ and for each $1 \le k \le K$, $(\mathcal{W}_{k,t})_{t\in\mathbb{R}_+}$ is a Brownian motion and $\sigma_k > 0$. Moreover, we assume that $(\mathcal{W}_1,\dots,\mathcal{W}_K)$ are independent.

The noise term is inspired by [28] or [37]. This amounts to considering for each class of size k, a unique noise driver independent to the other. However, we could consider (1) that the noises $(\mathcal{W}_1,\dots,\mathcal{W}_K)$ are correlated, (2) and even that there are three noises for each class of size k: one for susceptible, one for infected, and one for recovered.

Remark A.1. We could, instead AEP, adopt the quantities introduced in [22, 23]. We first introduce D_{ij} the time at which the subunit j of policyholder $i \in \{1,\dots,H\}$ becomes recovered, and L_{ij} the duration of the assistance required by the victim. Since T_{ij} is the infection time of subunit j, therefore we introduce $\eta_{ij} := \mathbf{1}_{T_{ij}\le D_{ij}}$ which indicates if the subunit j managed to become immune before infection.

We limit ourselves here to the cost function $\mathfrak{P}_3$ defined as follows

$$\mathfrak{P}_3 := \int_0^{t_d}\phi\left(\frac{\mathcal{J}_t}{K_i}\right)\mathrm{d}t,$$

which is a way to define saturation of the response, and where ϕ is a positive function and

$$\mathcal{J}_t^i := \sum_{j=1}^{K_i}\eta_{ij}\mathbf{1}_{T_{ij}\le t} - \sum_{j=1}^{K_i}\eta_{ij}\mathbf{1}_{T_{ij}+L_{ij}\le t}$$

describing the number of firms for which the crisis is still ongoing.

Subsequently, we may assume L_{ij} to be Exponential, D_{ij} to be Exponential, Pareto-type or Weibull-type distribution, and define T_{ij} via the associated hazard rate function λ_{ij}.

Appendix B: Figures and tables

Industry	Victims	Rate
Manufacturing	610	20.68%
Technology	390	13.22%
Retail/Wholesale	335	11.36%
Healthcare	325	11.02%
Consulting	275	9.32%
Construction	235	7.97%
Education	205	6.95%
Legal	200	6.78%
Government	190	6.44%
Banking/Finance	185	6.27%

FIG. 13: Most impacted industries in 2024

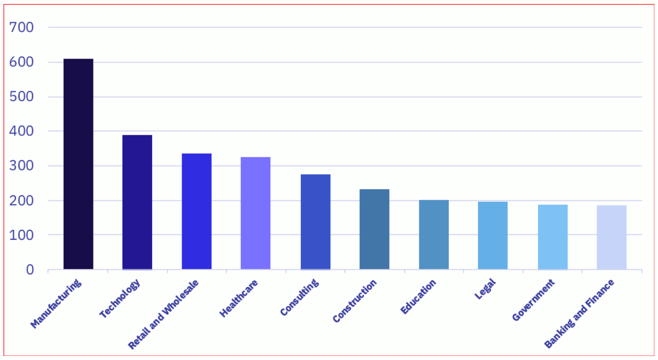
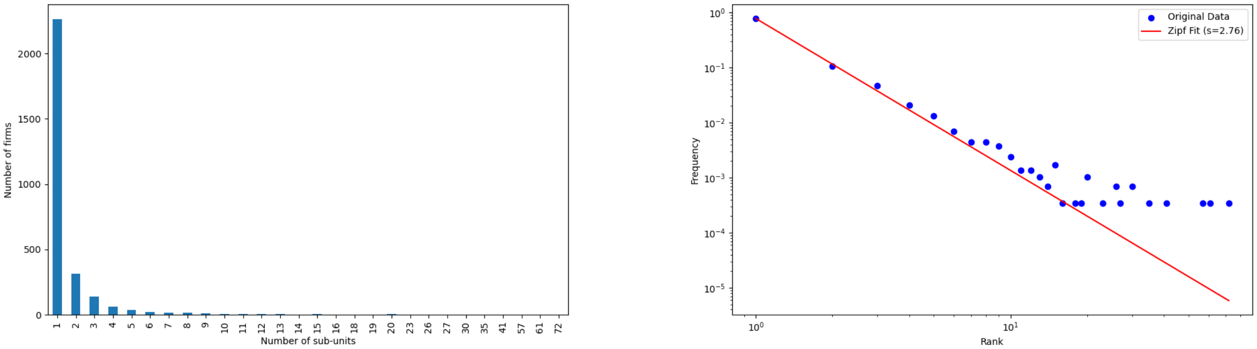


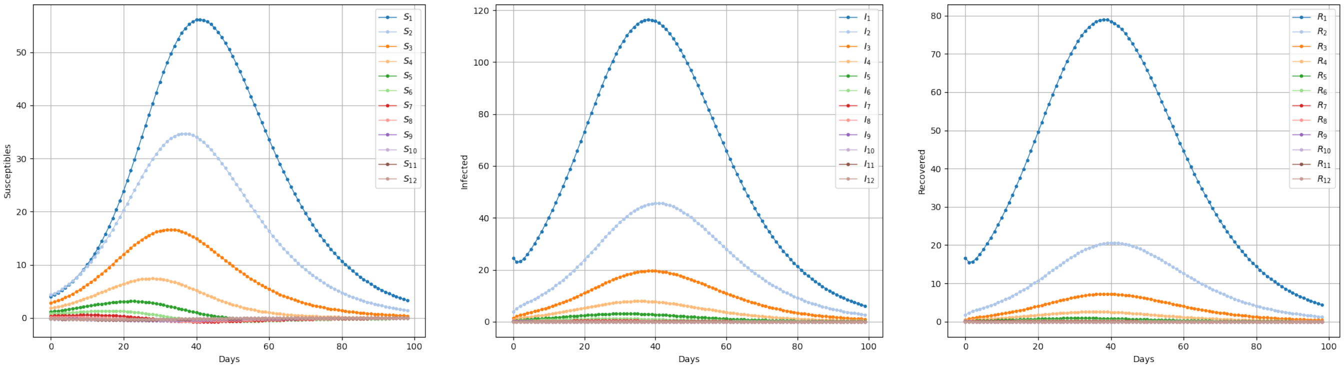
FIG. 14: Industries impacted (x-axis: sector, y-axis: victims)



(a) Histogram of the number of subunits

(b) Fit of the Zipf's law on the number of subunits

FIG. 15: Number of subunits



(a) New susceptible

(b) Infected

(c) New recovered

FIG. 16: Evolution of the number of each group as percentage of the initial population in the group (see Table IV)