IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 8 March 2023, accepted 28 April 2023, date of publication 2 May 2023, date of current version 10 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3272572

RESEARCH ARTICLE

The Price Tag of Cyber Risk: A Signal-Processing Approach

YUYING LI[1] AND ROGEMAR MAMON[1,2], (Member, IEEE)
[1]Department of Statistical and Actuarial Sciences, The University of Western Ontario, London, ON N6A 5B7, Canada
[2]Division of Physical Sciences and Mathematics, University of the Philippines Visayas, Miagao, Iloilo 5023, Philippines

Corresponding author: Rogemar Mamon (rmamon@stats.uwo.ca)

The work of Rogemar Mamon was supported by the Natural Sciences and Engineering Research Council of Canada through Discovery Grant RGPIN-2017-04235.

ABSTRACT The cyber risk insurance market is rapidly developing in consideration of the potentially huge losses attributed to cyberattacks. This requires the insurance business to have a valuation and risk management framework that will enable cyber insurance policy issuers to fulfil their future obligations. We present such a framework for cyber risk modelling, wherein the cyberattacks' occurrences as well as their inter-arrival and duration are captured by a regime-switching Markov model (RSMM). In this customised RSMM, the transition probabilities of the Markov chain are governed by another hidden Markov chain representing the various states of the cyber security environment. A self-calibrating mechanism is provided via filtering and a cyber kill chain is built based on the stages of the cyberattack. With the aid of change of reference probability measures and the EM algorithm, the estimators for the transition matrix are derived. Our main point of interest is the random losses from cyberattacks, which are assumed to follow a doubly-truncated Pareto distribution. The Vasiček model is utilised to describe the interest rate process for the discounting of losses. The premium for a cyber security insurance contract is calculated with the use of a simulated data set based on two pricing principles. Our methodology featuring dynamic parameter estimation and flexible adjustments in modelling various risk factors widens the available tools for pricing and cyber risk management.

INDEX TERMS Cyber insurance, HMM filters, premium calculation, regime-switching Markov model.

I. INTRODUCTION

The need for cyber risk insurance is now appreciated more than ever in this digital age by virtually all businesses relying heavily on e-commerce mode and information technology systems. Cyber risk refers to any risk of financial losses and costs incurred from reputational damage borne by a business organisation due to breaches in its computer networks. The damaging consequences include ransomed or stolen information, interruption of business operations, corrupted computer systems, and serious professional impacts (e.g., identity theft), amongst others. As per the document maintained by the National Protection and Programs Directorate under the Department of Homeland Security [52], the USA is estimated to have direct losses of approximately 2 to 3 billion dollars per year whilst the total indirect costs could reach as high as 40 billion dollars per year in the USA. The demand for cyber risk protection has increased recently giving impetus to a growing business line for insurers. Insurance companies develop cyber risk insurance products with coverage that is not provided by traditional types of policies.

Such products not only meet the requirement of the customers but also reduce legal costs triggered by legal disputes. Precise underwriting tools and detailed coverage description help resolve disputes between the insureds and insurers on what should be covered. The coverage is usually split into two categories. The first-party coverage is for losses and damage to the business of the insureds, while the third-party coverage or liability coverage is for losses of the insureds' customers or clients incurred as a result of a cyber event.

The associate editor coordinating the review of this manuscript and approving it for publication was Varuna de Silva.

There is enormous potential in the cyber risk insurance business. In the Betterley Report [9], the annual gross written premium for this segment of the insurance market increased from $2.75 billion in 2015 to $3.25 billion in 2016. According to Best's column [61], the top cyber writers in 2018 are the Chubb with 16% of the market share ($325.8 million) and AXA US with 12.6% of the market share ($255.9 million). In terms of policies in force identified in [61], Hartford ranks first with 510,000 policies. Despite the potential for earning considerable profits, companies are prudent about their total exposures and underwriting remains difficult. Biener et al. [10] discussed three major insurability problems of cyber risk. These are (i) the absence of information to aid the determination of independence and predictability of losses; (ii) information asymmetry and adverse selection as companies that experienced serious cyberattacks tend to be more willing to invest in cyber risk insurance; and (iii) existing policies only cover small losses but plausible extreme scenarios are not protected, which limits the development of the cyber risk insurance market. For example, the Data Breach Liability for small businesses offered by the CNA Financial Corporation, one of the top cyber liability insurance providers, has limits ranging from $100,000 to $2,000,000; see [23]. Apparently, the cyber insurance market is still at its early stages and standardized terminology and product regulations need further development.

Research in this area faces many challenges. This could be due to the scarcity of quality data that model validation entails, and the non-disclosure of the cost involved in data breaches. Sustained efforts are paramount in adjusting modelling approaches to be adaptable to an environment that is heavily dependent on fast-changing technology. Such efforts, as pointed out in Eling and Schnell [24], include modifications attuned to laws and regulations governing various aspects of cyber security risk.

Several researchers investigated the modelling of cyber risks using stochastic methods. Others focused on modelling the extreme losses or severities of cyberattacks. Wheatley et al. [67] found that the extremely heavy tailed truncated-Pareto distribution is an appropriate choice to model the recent data set covering 2007-2015 concerning the sizes of personal data breaches per incident. Jung [42] found that the data series on breach-loss maxima are stationary and serially correlated; the data series follow the Fréchet type of generalised extreme value distribution. The data source in [42] is Cowbell Cyber Inc, which is one of the largest private databases for data breach risk.

Certain studies on modelling the occurrences of cyberattacks were conducted in the past. Bessy-Roland et al. [8] proposed multivariate Hawkes processes, with specific kernel choices, aimed to capture the clustering and autocorrelation of the times of cyber events depending on their characteristics (e.g., type, target and location). In Fang et al. [33], the sparsity of enterprise-level data breaches is dealt with by leveraging the inter-entity or inter-enterprise dependence between multiple time series. Certain investigations centre on the dependence between the occurrences and the severities of cyber events. The computational complexity emerging from the correlation structure gives impetus to the utilisation of copulas, which are well-suited in capturing non-linear dependencies and in generating potential marginal distribution. For instance, a t-copula is an appropriate tool in examining extreme events as proposed by Böhme and Kataria [11]. Following Mukhopadhyay et al. [51], the joint distribution of the number of failures (frequency) and the loss given default (severity) were modelled by normal copulas and the derivation of the overall loss distribution was also shown. Xu et al. [69] modelled the dependence between the incidents' inter-arrival times and the breach sizes by the Gumbel copula and demonstrated as well that the ARMA-GARCH model could describe adequately the hacking breach sizes. A novel frequency-severity model for hacking breach risks of an individual company was proposed by Sun et al. [60] in which the breach frequency is modelled by a hurdle Poisson model and the breach severity is modelled by a non-parametric generalised Pareto distribution. The incorporation of network's features into a stochastic model is an enriching method for cyber risk modelling. An innovative approach of Xu and Lei [68] utilised epidemic models to characterise cyberattacks and facilitated the derivation of the dynamic upper bounds of the infection probabilities by applying Markov models. The premium principles were applied and demonstrated in [68] via simulation. Jevtić and Lanchier [41] presented a structural model of aggregate cyber loss distribution for small and medium-sized businesses under the assumption of a tree-based local area network (LAN) topology. Other relevant examples could be found in [5], [15], [31], and [32].

Considering the prime importance of digital advancements as the backbone of today's economic progress and way of life, technical groundwork tackling cyber risk issues appear to be gaining more traction. In Böhme and Kataria [11], cyber risk is modelled in two steps. The beta-binomial distribution is used to model the aggregate risk within a single company's network and the one-factor latent risk model is proposed to model the risks in multiple firms with similar characteristics at the global level. In [11], it was also discovered that cyber insurance is best suited for risks with high internal and low global correlations. A high internal correlation stimulates the need of cyber insurance for institutions whilst a low global correlation affects the insurers' decision in setting the premium. A related research work by Böhme and Schwartz [12] proposed a comprehensive framework in probing cyber risk's inherent properties such as interdependent security, correlated risks, and information asymmetries and in showing which parameters could provide guidance in the creation of future models with greater adaptability and improved functionality. Eling and Wirfs [26] identified "cyber risks of daily life" and "extreme cyber risks" by employing the peaks-over-threshold method from the extreme value theory with their analysis based on actual cost data. Their model produced consistent risk estimates, depending on country, industry, size, and other variables. Taking advantage of the emerging interests and growing developments in machine learning, applications of deep-learning techniques have permeated the field of cybersecurity. For example, Zhang et al. [70] made accurate high-dimensional point predictions via deep learning and the multivariate cyber risks and predicting the high quantiles using the extreme value theory.

Apart from the aforesaid methodologies, Husák et al. [38] found that Markov models function well in the presence of unobservable states and transitions. In contrast to other discrete modelling techniques such as attack graphs and Bayesian networks approaches, Markov models do not require possessing complete information to detect intrusion and predict attacks. This finding widens the applications of the hidden Markov models (HMMs) that include the detection and prediction of cyberattacks on computer networks; see [4] and [16]. This research also considers the utility of the HMM to model cyberattack occurrences. To estimate the model parameters, we rely on the Expectation-Maximization (EM) algorithm due to its robustness and ease of implementation. The EM algorithm is a numerical optimisation routine aiming at maximising the (log) likelihood of a batch of observations [14].

The EM-inspired methods are classified into two major categories: finite-memory approximations of the required smoothing computations [44] and finite-memory approximations of the data log-likelihood itself [56]. To numerically maximise the likelihood function, it is common to find the maximum likelihood parameter estimates (MLEs) in conjunction with the Kalman filtering. The Kalman filter is a special version of the HMM filter with continuous state space of latent variables and normally-distributed latent and observed variables. Moreover, the Kalman filter is an efficient recursive filter in the estimation of the internal state of a linear dynamic system from a series of noisy measurements. Research progress has been continually made in generalising the Kalman filter within the aspects of robustness to measurement outliers, accuracy of state estimations, and applicability to nonlinear systems (e.g., [35], [36]). For example, Gao et al. [34] proposed a novel Cubature Kalman Filter (CKF) approach for a tightly-coupled GNSS/INS (Global Navigation Satellite System/Inertial Navigation System) integration, which can be applied to vehicle positioning. The CKF put forward controls the interferences of both kinematic and observation modelling errors on state estimation. For this paper, we shall construct the EM algorithm based on the adaptive filter-based scheme introduced by [28]. By using the change of measure technique, we can derive filters under an ideal measure and obtain the real-world quantities through the Bayes' theorem for conditional expectations. Elliot and Hyndman [29] demonstrated the advantage of the filter-based algorithm over smoother-based EM algorithms. The filter-based algorithm will be at least twice as fast because it only needs a forward pass. Additionally, the filter-based algorithm can be easily implemented in parallel on a multiprocessor system. There may also be specific computational advantages for different models, such as the constant coefficient model in our case, where the filter-based algorithm can be modified to use the steady-state properties of the Kalman filter.

Recent research in the EM algorithm for HMMs has focused on developing more efficient and accurate algorithms for estimating model parameters (e.g. [1], [46]) as well as on applying HMMs to new and diverse applications and synthesising HMMs with new techniques for learning and inference such as deep learning and reinforcement learning (e.g. [45], [49]). Steady developments in the EM algorithm for HMMs have opened up more avenues for research and innovation in a wide range of fields.

In this paper, we consider the pricing of cyber risk insurance for a single company, focusing on policies that cover data breaches. We start with the modelling of the dynamics of cyberattacks based on the cyber kill chain (CKC). As outlined by Lockheed Martins Corp - one of the largest companies whose lines of business encompass aerospace, military support, security, and technology - the CKC defines seven stages of a cyberattack. In Fig. 1, we concentrate on the three stages or states of the CKC for the purpose of our cyber risk modelling and pricing. These stages are firewall working (stage 1), firewall fail (stage 2), and anti-phishing fail (stage 3). In stage 2, for example, when the firewall is unable to block malicious emails from spammers but the company's IT employees have mechanisms (e.g., phishing awareness training) to identify successfully spam emails, this could prevent unsuspecting email recipients from giving out their passwords via some webpage links in the email spam.

To model the transitions among the three states, Dionisi [21] applied the Markov chain, an idea that we generalise by considering a non-homogeneous regime-switching Markov model. More specifically, the transition probability of the Markov chain is stochastic and driven by another unobserved Markov chain that reflects the "state" of the cyber security environment. We derive the representation of the transition probabilities and the expected number of cyber attacks in a given amount of time. This model offers greater flexibility for the transition probability when fitting data that exhibit a wide range of characteristics. Compared to the discrete-state space of an HMM in discrete time introduced in Chapter 2 of Aggoun et al. [28], our proposed modelling approach is more appropriately suited to capture the transition patterns of the three states. In the discrete HMM of [28], the discrete finite-state stochastic process's states are the observed states and they follow a finite-range discrete distribution with probabilities driven by the hidden Markov chain. This fails, however, to explain the direct impact of the previously observed state compared to our formulation of the regime-switching Markov model (RSMM).

The Privacy Rights Clearinghouse [54] defined the categories of data breaches. On the basis of this definition,

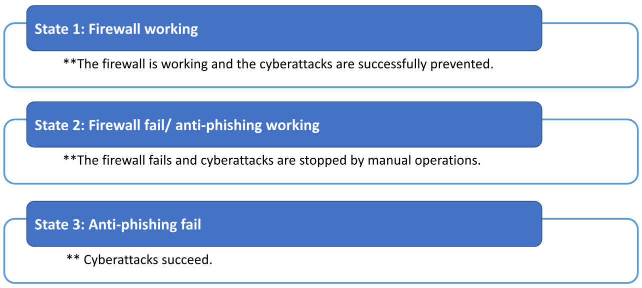


FIGURE 1. **A diagram depicting the cyber kill chain.**

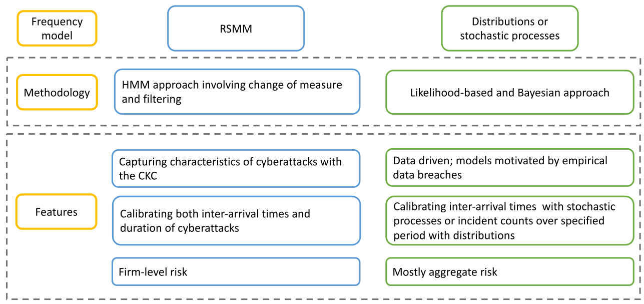


FIGURE 2. **A comparison between proposed frequency models and current frequency models.**

different types of data breaches could be catalogued by the transition from state 1 or 2 to state 3. For instance, a transition from state 1 to state 3 could occur when a portable device such as a laptop is lost. The data stored in the laptop may be leaked when password hacking by unscrupulous individuals succeeds. A transition from state 2 to state 3 could occur when the company's system is infected by malware and the employee opens an.exe attachment leading to the spread of computer virus infection. The losses are associated with the transition from state 1 or 2 to state 3. We employ the Monte-Carlo simulation to generate the transitions between the CKC states and to obtain the premiums based on two valuation principles used in practice for traditional insurance contracts. The relevant severities or breach sizes follow a doubly truncated Pareto distribution as advanced in Wheatley et al. [67] in which the models were based on the number of recorded data. The losses are deduced from breach sizes via the proportionality or functional-form assumption. The breach size refers to the number of data records lost while the loss size is the dollar amount of the loss incurred. In summary, the estimation of transition probabilities, determination of the number of attacks, and the calculation of premiums constitute a complete sequence of valuation steps.

Our work considers two main types of cyberattacks: hacking and insider threats. We recast the cyberattack event as an attacking/phishing process that could be described by stochastic models. Our proposed model could be calibrated not only to data on incident arrivals but also to incident duration. In contrast, most established frequencies or counting processes for cyber events capture the cyber incident counts over a specified period or inter-arrival times of cyberattacks. Modelling examples include established frequencies or counting process for cyber events following a negative binomial [22], hurdle Poisson model [60], Hawkes process [8], and autoregressive conditional mean model [69]. In addition, our starting point is to model the firm-level risk rather than aggregating risks, which is the common way in literature such as those in [22], [25], and [32]. The aforementioned discussion is summarised in Fig. 2. A complete pricing framework is also given to facilitate insurers with a cyber risk evaluation.

The paper is organised as follows. Section II introduces a regime-switching Markov modelling framework for the occurrences of data breaches. In Section III illustrates the construction of the total-loss process and the premium-calculation principles. The applicability and validation of our approach are demonstrated in Section IV through numerical implementation using simulated data. Lastly, Section V concludes.

II. REGIME-SWITCHING MARKOV MODEL

In this section, we outline our development of a regime-switching Markov model customised for the modelling of the CKC.

The regime-switching Markov model is constructed in Subsection II-A. Subsection II-B outlines the change of measure technique as a preliminary for the recovery of parameters from data. In Subsections ch: RSMM filter and II-D, the steps are detailed to obtain the optimal recursive estimations of parameters with online filters and the Expectation-Maximum (EM) algorithm. The long-run proportion of the number of attacks is derived in Subsection II-E.

A. DESCRIPTION OF THE REGIME-SWITCHING MARKOV MODEL

Adhering to the convention in matrix algebra, all vectors will be denoted by bold letters in lowercase while all matrices will be denoted by bold English or Greek letters in uppercase. Suppose $\mathbf{z}_k$ is a homogeneous discrete-time Markov chain with finite states. Assume that the initial state $\mathbf{z}_0$ is known. As in Elliott et al. [28], the state space of $\mathbf{z}_k$ is taken as the set of unit vectors $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$, where $\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)^\top \in \mathbb{R}^n$ and $\top$ denotes a vector or matrix transpose. The semi-martingale representation of $\mathbf{z}_k$ is

$$\mathbf{z}_{k+1} = \Pi \mathbf{z}_k + \mathbf{v}_{k+1}. \tag{1}$$

In equation (1), $\Pi = (\pi_{ij})$ is a transition matrix, $\mathbf{v}_{k+1}$ is a martingale increment with $E[\mathbf{v}_{k+1} | \mathcal{F}_k^{\mathbf{z}}] = 0$, and $\mathcal{F}_k^{\mathbf{z}}$ is the complete filtration generated by $\mathbf{z}_1, \mathbf{z}_2, \dots \mathbf{z}_k$. The above conditional expected value is computed under the real-world probability measure P.

The CKC's state process $\mathbf{y}_k$ takes values in $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_m\}$, where $\mathbf{f}_1 = (0, \dots, 0, 1, 0, \dots, 0)^\top \in \mathbb{R}^m$. In our case, $m = 3$ and the state $\mathbf{f}_i$ signifies that the CKC is in state i. Assume that $\mathbf{y}_k$ evolves as a Markov chain with transition matrix $\mathbf{B}(\mathbf{z}_k) = (b_{ij}(\mathbf{z}_k)) \in \mathbb{R}^{m\times m}$, where

$$b_{ji}(\mathbf{z}_k)\big|_{\mathbf{z}_k=\mathbf{e}_l} := P\left(\mathbf{y}_{k+1} = \mathbf{f}_j|\mathbf{y}_k = \mathbf{f}_i, \mathbf{z}_k = \mathbf{e}_l\right). \quad (2)$$

This implies that

$$\mathbf{y}_{k+1} = \mathbf{B}(\mathbf{z}_k)\mathbf{y}_k + \mathbf{w}_{k+1}, \quad (3)$$

where $\mathbf{w}_{k+1}$ is a martingale increment with $E[\mathbf{w}_{k+1}|\mathcal{F}_k] = \mathbf{0}$; $\mathcal{F}_k = \mathcal{F}^{\mathbf{z}}_k \bigvee \mathcal{F}^{\mathbf{y}}_k$ and $\mathcal{F}^{\mathbf{y}}_k$ is the complete filtration generated by $\{\mathbf{y}_k\}$.

Remark 1: Although the state equation (3) is similar to the form of the discrete HMM proposed in chapter 2 of Elliott et al. [28], our formulation here differs in two respects:

(i) the transition matrix $\mathbf{B}$ is time-dependent and hence, more general; and

(ii) the dynamics of $\mathbf{y}_{k+1}$ depends directly on $\mathbf{y}_k$ and not $\mathbf{z}_k$.

The theoretical difference leads to various modelling applications. For instance, a model for coin tossing in [62] is an illustration of a discrete HMM in pp. 15–56 of [28]. A sequence of coin-tossing outcomes is observed but it is modelled by two different and biased coins corresponding to two underlying Markov states. Given the choices of the coins, the tosses' outcomes are independent. This characteristic of the observed series does not fit with our RSMM as ours suggest correlated observed series given hidden states.

Our idea is to rewrite model (3) so that certain established results of homogeneous HMM with a discrete range could be adopted and extended into our case. We introduce $\mathbf{C} = (c_{ij}(\mathbf{y}_k)) \in \mathbb{R}^{m\times n}$, where

$$\begin{aligned} c_{ji}(\mathbf{y}_k)\big|_{\mathbf{y}_k=\mathbf{f}_l} &:= P\left(\mathbf{y}_{k+1} = \mathbf{f}_j|\mathbf{y}_k = \mathbf{f}_l, \mathbf{z}_k = \mathbf{e}_i\right) \\ &= b_{jl}(\mathbf{z}_k)\big|_{\mathbf{z}_k=\mathbf{e}_i}. \end{aligned} \quad (4)$$

Thus, invoking (2), (3) is equivalent to

$$\mathbf{y}_{k+1} = \mathbf{C}(\mathbf{y}_k)\mathbf{z}_k + \mathbf{w}_{k+1}. \quad (5)$$

Equation (5) is a one-step delay model and a reasonable model as $\mathbf{y}_{k+1}$ may not react to $\mathbf{z}$ immediately.

B. CHANGE OF REFERENCE PROBABILITY MEASURE

The rationale for the measure change is that, under the new measure $\overline{P}$ to be defined later, the sequence of observations $\mathbf{y}_k$ is transformed into a sequence of independent and identically distributed (IID) random variables each having a uniform distribution. So, a probability $\dfrac{1}{m}$ is assigned to each element $\mathbf{f}_i$, $1 \le i \le m$, in its range space. The transition matrix Π remains the same under $\overline{P}$; a proof tailored to our application is included as a lemma in Appendix A.

Define $\overline{\lambda}_l$ and $\overline{\Lambda}_k$ as

$$\overline{\lambda}_l := \prod_{i=1}^{m}\left(mc_l^{(i)}\right)^{y_l^{(i)}}, \quad (6)$$

$$\overline{\Lambda}_k := \prod_{l=1}^{k}\overline{\lambda}_l, \quad k \ge 1, \quad \overline{\Lambda}_0 = 1, \quad (7)$$

where $y_l^{(i)} = \langle \mathbf{y}_l, \mathbf{f}_i\rangle$ and $c_l^{(i)} = \langle C(\mathbf{y}_{l-1})\mathbf{z}_{l-1}, \mathbf{e}_i\rangle$. Then, $\overline{\Lambda}_k$ in (7) is referred to as the Radon-Nikod$\widehat{\text{y}}$m derivative of P with respect to $\overline{P}$, which is written as

$$\left.\frac{\mathrm{d}P}{\mathrm{d}\overline{P}}\right|_{\mathcal{F}_k} = \overline{\Lambda}_k.$$

We aim to estimate $\mathbf{z}$, given the observations under the real probability measure P. All calculations will be done though under $\overline{P}$ to take advantage of the IDD assumption, making the evaluation of conditional expectations more manageable. In other words, defining a new measure $\overline{P}$ is similar to constructing an idealised statistical setting under which calculations are performed with great ease because random variables are IID. The calculation results are then related back to the real-world setting (measure P) with the aid of the Bayes' theorem. To explain how this works, let us begin by letting $\widehat{\mathbf{z}}_k$ be the conditional expectation of $\mathbf{z}$ given $\mathcal{F}^{\mathbf{y}}_k$ under P. That is,

$$\begin{aligned} \widehat{\mathbf{z}}_k &:= E\left[\mathbf{z}_k|\mathcal{F}^{\mathbf{y}}_k\right] = \left(\widehat{z}_k^{(1)}, \widehat{z}_k^{(2)}, \ldots, \widehat{z}_k^{(n)}\right)^\top \in \mathbb{R}^n, \\ \widehat{z}_k^{(i)} &:= P\left(\mathbf{z}_k = \mathbf{e}_i|\mathcal{F}^{\mathbf{y}}_k\right) = E\left[\langle \mathbf{z}_k, \mathbf{e}_i\rangle|\mathcal{F}^{\mathbf{y}}_k\right]. \end{aligned}$$

By the Bayes' theorem for conditional expectation,

$$\widehat{\mathbf{z}}_k = E\left[\mathbf{z}_k|\mathcal{F}^{\mathbf{y}}_k\right] = \frac{\overline{E}\left[\overline{\Lambda}_k\mathbf{z}_k|\mathcal{F}^{\mathbf{y}}_k\right]}{\overline{E}\left[\overline{\Lambda}_k|\mathcal{F}^{\mathbf{y}}_k\right]},$$

which shows that the optimal estimate under P is expressed in terms of the calculations under $\overline{P}$.

Write $\mathbf{p}_k := \overline{E}\left[\overline{\Lambda}_k\mathbf{z}_k|\mathcal{F}^{\mathbf{y}}_k\right]$. This gives

$$\begin{aligned} \overline{E}\left[\overline{\Lambda}_k|\mathcal{F}^{\mathbf{y}}_k\right] &= \overline{E}\left[\overline{\Lambda}_k\left(\sum_{i=1}^{n}\langle \mathbf{z}_k, \mathbf{e}_i\rangle\right)\middle|\mathcal{F}^{\mathbf{y}}_k\right] \\ &= \sum_{i=1}^{n}\overline{E}\left[\langle\overline{\Lambda}_k\mathbf{z}_k, \mathbf{e}_i\rangle|\mathcal{F}^{\mathbf{y}}_k\right] \\ &= \sum_{i=1}^{n}\left\langle\overline{E}\left[\overline{\Lambda}_k\mathbf{z}_k, \mathbf{e}_i|\mathcal{F}_k\right]\right\rangle = \sum_{i=1}^{n}\langle \mathbf{p}_k, \mathbf{e}_i\rangle, \end{aligned}$$

where for the middle expression in the first equality above, we make use of the fact that $\sum_{i=1}^{n}\langle \mathbf{z}_k, \mathbf{e}_i\rangle = 1$. Therefore,

$$\widehat{\mathbf{z}}_k = \frac{\mathbf{p}_k}{\sum_{i=1}^{n}\langle \mathbf{p}_k, \mathbf{e}_i\rangle}.$$

C. COMPUTATION OF ONLINE FILTERS

As a prelude to the construction of online or recursive filters, define the vector $\mathbf{d}_k = (d_k^{(1)}, d_k^{(2)}, \ldots, d_k^{(n)})^\top$ by

$$d_k^{(j)} = m\prod_{i=1}^{m}\left(c_{ij}(\mathbf{y}_{k-1})\right)^{y_k^{(i)}}, \quad 1 \le j \le n.$$

Let G_k be any scalar $\mathcal{F}^{\mathbf{y}}_k$-adapted process; G_0 is $\mathcal{F}^{\mathbf{y}}_0$-measurable. The best estimate for G_k is defined as $E[G_k|\mathcal{F}^{\mathbf{y}}_k]$. Again, by the Bayes' theorem,

$$E[G_k|\mathcal{F}^{\mathbf{y}}_k] = \frac{\overline{E}[G_k\overline{\Lambda}_k|\mathcal{F}^{\mathbf{y}}_k]}{\overline{E}[\overline{\Lambda}_k|\mathcal{F}^{\mathbf{y}}_k]} = \frac{\overline{E}[G_k\overline{\Lambda}_k|\mathcal{F}^{\mathbf{y}}_k]}{\sum_{i=1}^{n}\langle \mathbf{p}_k, \mathbf{e}_i\rangle}. \quad (8)$$

The filter for G_k is $\gamma(G_k) := \overline{E}[G_k\overline{\Lambda}_k|\mathcal{F}^{\mathbf{y}}_k]$.

Remark 2: It has to be noted that $\langle \overline{E}[G_k \mathbf{z}_k \overline{\Lambda}_k | \mathcal{F}_k^{\mathbf{y}}], \mathbf{1}\rangle = \gamma(G_k\langle \mathbf{z}_k, \mathbf{1}\rangle) = \gamma(G_k)$, where $\mathbf{1}$ is a vector of 1's.

In terms of the filters, therefore, (8) becomes

$$E[G_k | \mathcal{F}_k^{\mathbf{y}}] = \frac{\gamma(G_k)}{\gamma(1)} = \frac{\langle \gamma(G_k \mathbf{z}_k), \mathbf{1}\rangle}{\sum_{i=1}^{n}\langle \mathbf{p}_k, \mathbf{e}_i\rangle}. \tag{9}$$

These filters will aid in obtaining an online parameter estimation scheme.

Before we delve into the calculation of optimal model parameters, we define (for $r, j = 1, 2, \ldots, n$ and $s, i = 1, 2, \ldots, m$) the following quantities:

$$\mathcal{J}_k^{j,r} = \sum_{l=1}^{k}\langle \mathbf{z}_{l-1}, \mathbf{e}_r\rangle\langle \mathbf{z}_l, \mathbf{e}_j\rangle, \tag{10}$$

$$\mathcal{O}_k^{r} = \sum_{l=1}^{k}\langle \mathbf{z}_{l-1}, \mathbf{e}_r\rangle, \tag{11}$$

$$\mathcal{T}_k^{s,r}(\mathbf{y}_k, \mathbf{f}_i) = \sum_{l=1}^{k}\langle \mathbf{z}_{l-1}, \mathbf{e}_r\rangle\langle \mathbf{y}_l, \mathbf{f}_s\rangle\langle \mathbf{y}_{l-1}, \mathbf{f}_i\rangle, \tag{12}$$

$$\mathcal{T}_k^{r}(\mathbf{f}_i) = \sum_{l=1}^{k}\langle \mathbf{z}_{l-1}, \mathbf{e}_r\rangle\langle \mathbf{y}_{l-1}, \mathbf{f}_i\rangle. \tag{13}$$

In equations (10)-(13), $\mathcal{J}_k^{j,r}$ is the number of jumps from $\mathbf{e}_r$ to state $\mathbf{e}_j$ in time k; $\mathcal{O}_k^r$ is the amount of time that the Markov chain $\mathbf{z}$ spent in state $\mathbf{e}_r$ up to k; $\mathcal{T}_k^{s,r}(\mathbf{y}_k, \mathbf{f}_i)$ counts the number of times up to k that $\mathbf{y}$ is in state $\mathbf{f}_s$ given that previously the Markov chain $\mathbf{z}$ was in state $\mathbf{e}_r$ and $\mathbf{y}$ was in state $\mathbf{f}_i$; $\mathcal{T}_k^r(\mathbf{f}_i)$ counts the number of times up to k for which the Markov chain $\mathbf{z}$ visited state $\mathbf{e}_r$ and $\mathbf{y}$ entered state $\mathbf{f}_i$. From (9), the filtered estimates of $\mathcal{J}_k^{j,r}$, $\mathcal{O}_k^r$, $\mathcal{T}_k^{s,r}(\mathbf{y}_k, \mathbf{f}_i)$ and $\mathcal{T}_k^r(\mathbf{f}_i)$ are given by

$$\widehat{\mathcal{J}}_k^{j,r} = \frac{\left\langle \gamma(\mathcal{J}_k^{j,r}\mathbf{z}_k), \mathbf{1}\right\rangle}{\sum_{i=1}^{n}\langle \mathbf{p}_k, \mathbf{e}_i\rangle},$$

$$\widehat{\mathcal{O}}_k^{r} = \frac{\left\langle \gamma(\mathcal{O}_k^{r}), \mathbf{1}\right\rangle}{\sum_{i=1}^{n}\langle \mathbf{p}_k, \mathbf{e}_i\rangle},$$

$$\widehat{\mathcal{T}}_k^{s,r}(\mathbf{y}_k, \mathbf{f}_i) = \frac{\left\langle \gamma\left(\mathcal{T}_k^{s,r}(\mathbf{y}_k, \mathbf{f}_i)\mathbf{z}_k\right), \mathbf{1}\right\rangle}{\sum_{i=1}^{n}\langle \mathbf{p}_k, \mathbf{e}_i\rangle},$$

$$\widehat{\mathcal{T}}_k^{r}(\mathbf{f}_i) = \frac{\left\langle \gamma\left(\mathcal{T}_k^{r}(\mathbf{f}_i)\mathbf{z}_k\right), \mathbf{1}\right\rangle}{\sum_{i=1}^{n}\langle \mathbf{p}_k, \mathbf{e}_i\rangle}.$$

In turn, the recursive relations of $\mathbf{p}_k$, $\mathcal{J}_k^{j,r}$, $\mathcal{O}_k^r$, $\mathcal{T}_k^{s,r}(\mathbf{y}_k, \mathbf{f}_i)$ and $\mathcal{T}_k^r(\mathbf{f}_i)$ are:

$$\mathbf{p}_k = \Pi\ \mathrm{diag}(\mathbf{d}_k)\mathbf{p}_{k-1}, \tag{14}$$

$$\gamma(\mathcal{J}_k^{j,r}\mathbf{z}_k) = \Pi\ \mathrm{diag}(\mathbf{d}_k)\gamma(\mathcal{J}_{k-1}^{j,r}\mathbf{z}_{k-1}) + d_k^{(r)}\langle \mathbf{p}_{k-1}, \mathbf{e}_r\rangle \pi_{jr}\mathbf{e}_j, \tag{15}$$

$$\gamma(\mathcal{O}_k^{r}\mathbf{z}_k) = \Pi\ \mathrm{diag}(\mathbf{d}_k)\gamma(\mathcal{O}_{k-1}^{r}\mathbf{z}_{k-1}) + d_k^{(r)}\langle \mathbf{p}_{k-1}, \mathbf{e}_r\rangle \boldsymbol{\pi}_r, \tag{16}$$

$$\gamma\left(\mathcal{T}_k^{s,r}(\mathbf{y}_k, \mathbf{f}_i)\mathbf{z}_k\right) = \Pi\ \mathrm{diag}(\mathbf{d}_k)\gamma\left(\mathcal{T}_{k-1}^{s,r}(\mathbf{y}_{k-1}, \mathbf{f}_i)\mathbf{z}_{k-1}\right) + m\langle \mathbf{p}_{k-1}, \mathbf{e}_r\rangle\langle \mathbf{y}_k, \mathbf{f}_s\rangle\langle \mathbf{y}_{k-1}, \mathbf{f}_i\rangle c_{sr}(\mathbf{f}_i)\boldsymbol{\pi}_r, \tag{17}$$

$$\gamma\left(\mathcal{T}_k^{r}(\mathbf{f}_i)\mathbf{z}_k\right) = \Pi\ \mathrm{diag}(\mathbf{d}_k)\gamma\left(\mathcal{T}_{k-1}^{r}(\mathbf{f}_i)\mathbf{z}_{k-1}\right) + d_k^{(r)}\langle \mathbf{p}_{k-1}, \mathbf{e}_r\rangle\langle \mathbf{y}_{k-1}, \mathbf{f}_i\rangle \boldsymbol{\pi}_r. \tag{18}$$

The proofs of (14), (15) and (16) can be found in Mamon et al. [48] whilst the proofs of (17) and (18) are detailed in Appendix B.

D. PARAMETER ESTIMATION OF THE REGIME-SWITCHING MODEL

The estimation of the model parameters is based on a sequence of measure changes along with the Expectation-Maximum (EM) algorithm, which can be found in section 2.7 of [28]. The EM algorithm is introduced below; see Elliott and Krishnamurthy [27] for a detailed exposition.

Let $\mathcal{Y} \subset \mathcal{F}$ and $\left\{P^{\theta}, \theta \in \Theta\right\}$ be a family of probability measures on a measurable space $(\Omega, \mathcal{F})$, which is absolutely continuous with respect to a fixed probability measure P^0; and Θ is a parameter space. The likelihood function entailed in estimating θ on the basis of information contained in $\mathcal{Y}$ is

$$\mathcal{L}(\theta) = E^{\theta}\left[\frac{\mathrm{d}P^{\theta}}{\mathrm{d}P^{0}}\bigg|\mathcal{Y}\right]$$

and the maximum likelihood estimator (MLE) of θ is

$$\widehat{\theta} \in \underset{\theta\in\Theta}{\mathrm{argmax}}\ \mathcal{L}(\theta).$$

We seek an estimator of θ that maximises the conditional expectation of the density. Nonetheless, the MLE cannot be calculated directly in general especially for a complicated density. The course of action is to resort to numerical or iterative methods such as the EM algorithm that approximates the true parameter estimates.

The algorithm's first step is to set $l = 0$ and choose $\widehat{\theta}_0$. The second step, also referred to as the E-step, is to set $\theta^* = \widehat{\theta}_l$ and the posterior is

$$P^{\theta^*}(\mathbf{z}_k = \mathbf{e}_r|\mathcal{Y}) = \frac{P^{\theta^*}(\mathbf{y}_{k+1} = \mathbf{f}_s|\mathcal{Y})P(\mathbf{z}_k = \mathbf{e}_r)}{\sum_{l=1}^{n}P^{\theta^*}(\mathbf{y}_{k+1} = \mathbf{f}_s|\mathcal{Y})P(\mathbf{z}_k = \mathbf{e}_l)} = \frac{c_{sr}(\mathbf{f}_i)P(\mathbf{z}_k = \mathbf{e}_r)}{\sum_{l=1}^{n}c_{sr}(\mathbf{f}_i)P(\mathbf{z}_k = \mathbf{e}_l)},$$

where $P(\mathbf{z}_k = \mathbf{e}_l)$ are prior information of $\mathbf{z}_k$. Next, we compute

$$Q(\theta, \theta^*) = E^{\theta^*}\left[\frac{\mathrm{d}P^{\theta}}{\mathrm{d}P^{\theta^*}}\bigg|\mathcal{Y}\right].$$

The third step, also referred to as the M-step, is to determine $\widehat{\theta}_{l+1} \in \underset{\theta\in\Theta}{\mathrm{argmax}}\ Q(\theta, \theta^*)$. The last step is to replace l by $l+1$ and repeat the procedure from the second step until a stopping criterion is met. The estimated values $\{\widehat{\theta}_l, l \geq 0\}$ are nondecreasing as guaranteed by the Jensen's inequality and they converge to a likelihood's local maximum. Guided by [19], this convergence is facilitated in our initialisation stage. We use the *fminsearch* in package "pracma" [13] to find $c_{sr}(\mathbf{f}_i)$'s that minimises the likelihood

$$\mathcal{L}\left(\mathbf{y}_1, \ldots \mathbf{y}_k; c_{sr}(\mathbf{f}_i), \quad 1 \leq s \leq m, 1 \leq r \leq n\right)$$

$$= \prod_k \sum_{r=1}^{n} \sum_{s=1}^{m} c_{sr}(\mathbf{f}_i)^{\langle \mathbf{z}_{k-1}, \mathbf{e}_r \rangle \langle \mathbf{y}_k, \mathbf{f}_s \rangle \langle \mathbf{y}_{k-1}, \mathbf{f}_i \rangle}.$$

The search of the parameters will be carried through in the range of [0, 1], which yields "global" optima within that range.

The optimal parameter set $\widehat{\Theta} = \{\widehat{\pi}_{jr}, \widehat{c}_{sr}(\mathbf{f}_i), 1 \le j, r \le n, 1 \le s, i \le m\}$ maximizes the Q function, and through the EM algorithm, these optimal parameters are:

$$\widehat{\pi}_{jr} = \frac{\gamma(\mathcal{J}_k^{j,r})}{\gamma(\mathcal{O}_k^r)}, \tag{19}$$

$$\widehat{c}_{sr}(\mathbf{f}_i) = \frac{\gamma\left(\mathcal{T}_k^{s,r}(\mathbf{y}_k, \mathbf{f}_i)\right)}{\gamma\left(\mathcal{T}_k^{r}(\mathbf{f}_i)\right)}. \tag{20}$$

The respective proofs of (19) and (20) are presented in [48] and Appendix C. We also compute the variances of the estimators from the following Fisher information:

$$\mathcal{I}(\pi_{jr}) = \frac{\widehat{\mathcal{J}}_k^{j,r}}{\pi_{jr}^2}, \tag{21}$$

$$\mathcal{I}(c_{sr}) = \frac{\widehat{\mathcal{T}}_k^{s,r}}{c_{sr}^2} \tag{22}$$

The derivation of (21) and (22) can be found in [37] and Appendix D, respectively.

E. THE LONG-RUN PROPORTION OF THE ATTACKS

From the model specification in (1), the Markov chain $\mathbf{z}_k$ has the transition probability matrix $\mathbf{\Pi}$. We shall assume that $\mathbf{z}_k$ is irreducible with finite states. Thus, $\mathbf{z}_k$ is positive recurrent and has the long-run proportions $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)^\top$, which uniquely solve the equation

$$\boldsymbol{\alpha} = \boldsymbol{\alpha}\Pi \;\text{ with }\; \sum_{i=1}^{n} \alpha_i = 1.$$

The Markov chain $\mathbf{y}_k$ is also assumed irreducible with finite states when $\mathbf{z}$ is fixed. Suppose further that $\boldsymbol{\xi}_i = \left(\xi_i^{(1)}, \ldots, \xi_i^{(m)}\right)^\top$, for $i = 1, \ldots n$, is the long-run proportions of $\mathbf{y}_k$ given that $\mathbf{z}_{k-1} = \mathbf{e}_i$. That is,

$$\boldsymbol{\xi}_i = \boldsymbol{\xi}_i \mathbf{B}(\mathbf{e}_i) \;\text{ with }\; \sum_{j=1}^{m} \xi_i^{(j)} = 1,$$

where $b_{sr}(\mathbf{e}_i) = c_{si}(\mathbf{f}_r)$ as indicated in in (4). Note that $\boldsymbol{\alpha}$ and $\boldsymbol{\xi}_i$ are also stationary probabilities, that is, if $P(\mathbf{z}_0 = \mathbf{e}_i) = \alpha_i$, $P(\mathbf{z}_k = \mathbf{e}_i) = \alpha_i, k \ge 1$; similar properties hold for $\boldsymbol{\xi}_i$. By conditioning on $\mathbf{z}$, in the long run,

$$P\left(\mathbf{y}_k = \mathbf{f}_j\right) = \sum_{i=1}^{n} \xi_i^{(j)} \alpha_i.$$

Recall that a cyberattack occurs when $\mathbf{y}_k$ jumps from state 1 or 2 to state 3. Hence, the long-run proportion of cyberattacks is

$$\sum_{j=1}^{2} \sum_{i=1}^{3} \xi_i^{(j)} \alpha_i b_{j3}(\mathbf{e}_i). \tag{23}$$

Finally, the number of cyberattacks is equal to the value produced by (23) multiplied by the number of $\mathbf{y}_k$ within a certain time horizon.

III. THE TOTAL LOSS PROCESS AND PREMIUM CALCULATION

We now consider the total losses over the time period T and the principles underlying the calculation of premiums. We shall ensure that the subdivisions $\Delta t_k := t_k - t_{k-1}$ of $[0, T]$ will coincide with the time-unit subdivisions of $\mathbf{y}_k$. It is assumed that the optimal estimates of π_{jr}'s and $c_{sr}(\mathbf{f}_i)$'s are produced by our proposed filtering method in the previous Section upon application to a data set. The ensuing discussion is divided into three Subsections dealing with an interest rate model to discount the losses, the total loss process, and the computation of premiums.

A. THE INTEREST RATE MODEL

Let r_k be the interest rate at time k and independent of $\{\mathbf{y}_k, k = 1, 2, \ldots, T\}$. The r_k process is based on a continuous-time version of r_t possessing the Vašiček dynamics via the stochastic differential equation

$$\mathrm{d}r_t = \tau(a - r_t)\,\mathrm{d}t + \nu\,\mathrm{d}W_t, \tag{24}$$

where the parameters τ, a and $\nu > 0$ are constants and $\{W_t\}$ is a standard Brownian motion. The solution of (24) is

$$r_t = r_0 + a(1 - e^{-\tau t}) + \nu \int_0^t e^{-\tau(t-s)}\,\mathrm{d}W_t.$$

Clearly, r_k follows the normal distribution with

$$\mu_{r,k} := E[r_k] = r_0 e^{-\tau k} + a(1 - e^{-\tau k}) \text{ and}$$

$$\sigma_{r,k}^2 := \mathrm{Var}[r_k] = \frac{\nu^2}{2\tau}\left(1 - e^{-2\tau k}\right).$$

The interest rate model in (24) is discretised when incorporating it into the insurance valuation. As k goes to infinity, we obtain the respective long-term mean and variance

$$\mu_r := a \tag{25}$$

$$\sigma_r^2 := \frac{\nu^2}{2\tau}. \tag{26}$$

It is apparent that the constant interest rate situation is a special case of our generalised framework, which naturally embeds stochastic discounting. For the related annuity valuation that takes into account our stochastic interest-rate-modelling approach, see [71].

B. THE TOTAL LOSS PROCESS

The incurred loss $L_k^{(i)}$, for $i = 1, 2$, occurs when $\mathbf{y}_k$ goes from state i to state 3, i.e., $\mathbf{y}_k = \mathbf{f}_3$ and $\mathbf{y}_{k-1} = \mathbf{f}_i$. Suppose $L_k^{(i)}$'s are IID random variables with the same distribution as $L^{(i)}$. The total discounted loss during $[0, T]$, denoted by S_T, is

$$S_T = \sum_{k=1}^{T} \sum_{i=1}^{2} \langle \mathbf{y}_k, \mathbf{f}_3 \rangle \langle \mathbf{y}_{k-1}, \mathbf{f}_i \rangle L_k^{(i)} e^{-r_k k}, \tag{27}$$

where r_k is converted to the force of interest rate per time interval Δt_k. The distribution of losses may be estimated from real data. In general, the costs per cyberattack could be not determined exactly. But, since the data breach sizes are disclosed and they are useful information given their very close link to the losses. Such a link could also be described in a quantitative way. We denote the breach sizes by $A^{(i)}$. In this paper, the relationship between $A^{(i)}$ and $L^{(i)}$ is illustrated under two separate assumptions: (i) proportionality assumption, i.e., $L^{(i)}$ is proportional to $A^{(i)}$; and (ii) functional-form assumption, i.e., $L^{(i)}$ is derived from $A^{(i)}$ through the equation

$$\log\left(L^{(i)}\right) = 7.68 + 0.7568 \times \log\left(A^{(i)}\right). \tag{28}$$

Assumption (i) was supported by the empirical reports that provide average costs per data breach record, such as \$161 per record in [39]. Therefore, the losses with a known number of breached records can be estimated via average costs. The log-log model in Assumption (ii) was shown by Jacobs [40], and was widely applied to estimate costs in multiple models such as [22], [25], and [60]. Algarni and Malaiya summarised various models available to compute the costs of data breaches; see [3].

Proposed probability distributions to model the cyber-attack severities include the log-normal family of distributions [22], non-parametric generalised Pareto distribution [60], and mixed distributions [33]. The severity of large casualty losses for certain lines of business such as general liability, commercial auto, and workers' compensation is approximately Pareto-distributed. These results motivate the use of Pareto distribution in modelling cyber losses; see [58] and page 94 of [43]. In this paper, we assume that $A^{(i)}$ follows a doubly truncated Pareto (DTP) distribution as suggested in [67].

Thus, the distribution function of $A^{(i)}$ is

$$F_{A^{(i)}}(x) = \frac{1-(x/u_i)^{-\delta_i}}{1-(v_i/u_i)^{-\delta_i}},$$

where $0 < u_i < x \leq v_i, 0 < \delta < 1$ and $i = 1, 2$. The expectation and second moment of $A^{(i)}$ are

$$E\left[A^{(i)}\right] = \frac{u^\delta v - v^\delta u}{v^\delta - u^\delta} \quad \text{and} \quad E\left[\left(A^{(i)}\right)^2\right] = \frac{u^\delta v^2 - v^\delta u^2}{v^\delta - u^\delta},$$

respectively. Suppose that the initial states of $\mathbf{z}$ and $\mathbf{y}$ are assigned as the corresponding stationary probabilities. Then,

$$E[S_T] = \sum_{k=1}^{T}\sum_{j=1}^{2}\sum_{i=1}^{2}\xi_i^{(j)}\alpha_i b_{j3}(\mathbf{e}_i)E\left[L^{(j)}\right]E\left[e^{-r_k k}\right],$$

where the discount factor is calculated under the interest-rate setting in [47]. For simplicity, the interest-rate model is discretised and the rates' long-term mean and variance are used to approximate the expected total losses. Therefore,

$$\begin{aligned}
E[S_T] &= \sum_{k=1}^{T}\sum_{j=1}^{2}\sum_{i=1}^{2}\xi_i^{(j)}\alpha_i b_{j3}(\mathbf{e}_i)E\left[L^{(j)}\right]E\left[e^{-r_k k}\right] \\
&= \sum_{k=1}^{T}E\left[e^{-r_k k}\right]\sum_{j=1}^{2}\sum_{i=1}^{2}\xi_i^{(j)}\alpha_i b_{j3}(\mathbf{e}_i)E\left[L^{(j)}\right] \\
&= \sum_{k=1}^{T}\exp\left(-\mu_r k + \frac{1}{2}\sigma_r^2 k^2\right) \\
&\quad \times \sum_{j=1}^{2}\sum_{i=1}^{2}\xi_i^{(j)}\alpha_i b_{j3}(\mathbf{e}_i)E\left[L^{(j)}\right] \\
&\approx \int_0^T \exp\left(-\mu_r k + \frac{1}{2}\sigma_r^2 k^2\right)\mathrm{d}k \\
&\quad \times \sum_{j=1}^{2}\sum_{i=1}^{2}\xi_i^{(j)}\alpha_i b_{j3}(\mathbf{e}_i)E\left[L^{(j)}\right] \\
&\approx \frac{\sqrt{\pi\left(e^{T^2}-1\right)}}{\sqrt{2}\sigma_r}\exp\left(-\frac{\mu_r^2}{2\sigma_r^2}\right) \\
&\quad \times \sum_{j=1}^{2}\sum_{i=1}^{2}\xi_i^{(j)}\alpha_i b_{j3}(\mathbf{e}_i)E\left[L^{(j)}\right],
\end{aligned} \tag{29}$$

where r_k and σ_r are defined in (25) and (26) and the integral calculation in (29) can be found in Appendix E. If the horizon time T is large enough, we can approximate the sum of r_k's moment generating functions by an integral as we did in (29). For the special case of a constant interest rate over the time horizon $[0, T]$, i.e., $r_k = r$, the expected total losses becomes

$$E[S_T] = \frac{1-e^{-rT}}{e^r - 1}\sum_{j=1}^{2}\sum_{i=1}^{2}\xi_i^{(j)}\alpha_i b_{j3}(\mathbf{e}_i)E\left[L^{(j)}\right].$$

C. THE PREMIUM CALCULATION

From the standard-deviation premium principle, the premium $H(S_T)$ is given by

$$H(S_T) = E[S_T] + \lambda_r\sqrt{V[S_T]}, \tag{30}$$

where $\lambda_r > 0$ is the risk loading that represents the level of transaction costs. A more risk-averse insured, for instance, is willing to pay a premium with larger λ_r. Alternatively, the premium $H(S_T)$ based on the principle of equivalent utility is the solution of the equation

$$u(\omega) = E\left[u\left(\omega - S_T + H(S_T)\right)\right], \tag{31}$$

where u is an increasing concave utility wealth and ω is the initial wealth. We shall consider in our numerical implementation the utility function of the form

$$u(x) = 1 - e^{-\kappa x}, \quad x > 0.$$

The solution to (31) has the closed-form representation

$$H(S_T) = \frac{1}{\kappa}\log\left(E\left[e^{\kappa S_T}\right]\right), \tag{32}$$

where κ is the risk-aversion parameter. When κ approaches 0, the premium converges to $E[S_T]$. Given the above-mentioned utility function, the premium principle is called the exponential premium principle; for more details, see [20].

IV. NUMERICAL ILLUSTRATION

The procedure for implementing our pricing framework is as follows:

- Step 1: Simulate the data as an underwriting basis of cyberattack occurrences.
- Step 2: Implement the RSMM and obtain the transition probabilities.
- Step 3: Simulate the total losses with parameters estimated in Step 2 and calculate the premiums.

In this section, how the data simulated in step 1 is illustrated in Subsection IV-A and the estimation results of Step 2 are presented in Subsection IV-B. We discuss how the parameters of losses are set and the behaviours of simulated premiums under different principles and loss assumptions in Subsection IV-C. The semi-parametric approximation is displayed in Subsection IV-D. Finally in Subsection IV-E, we conduct a case study with higher chances of cyberattacks and compare the RSMM with other models in terms of AICs and BICs.

A. DATA SIMULATION

In the absence of a reliable data set, we use a simulated data set to demonstrate the practicalities of our online parameter estimation via HMM filtering. This is followed by determining the number of cyberattacks through simulation with the utilization of the estimated parameters. This leads to the final step of obtaining the premiums.

Firewalls are equipped with real-time cyber security monitors. They provide a record of cyber-attack stages in minutes. Reports encapsulated in the PRC data [60] indicated that majority of companies had only one incident and only 8 companies had more than two incidents from 01 January 2010 to 31 March 2019. It could be reasonably assumed that there are no multiple cyberattacks in one day for a single institution. Thus, the transitions between CKC states that lead to cyber-attack incidents on a minute-frequency basis over a 24-hour period will be recorded as the transitions in the daily frequency. For example, in the famous WannaCry ransomware incident [66], the attack was ongoing from 07:44 to 15:03 UTC on 12 May 2017. In this case, we shall record a CKC stage 3 on the 12th of May and a CKC stage 1 on the 13th of May. By changing the frequency of the data, the model complexity is reduced. Unfortunately, the publicly available data only specifies the date when the cyberattack was made known to the public. To apply our RSMM model, the starting and ending times of the cyberattacks are needed. Collecting reliable data from the firewalls directly, if possible, would be ideal. Due to limited data and resources, we illustrate our framework by simulated data. Suppose we have one-year data sets from a group of 200 institutions that share similar cyber risk characteristics such as data and organisational types. This data set will serve as an underwriting basis for the insured that could be classified into the same group. Additional details on how cyber risk insurance carriers assess the risk are given in [55]. Our simulated data set has $365 \times 200 = 73,000$ observations.

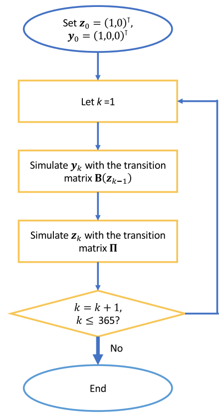


FIGURE 3. Simulation flow chart.

Following the simulation steps in Fig. 3, the data set for one company could be obtained. These steps could then be repeated 200 times to generate the full underwriting data. Note that for each company, we have paths of the CKC process with daily frequency for one year. The cyber security environment is assumed to be switching between good ($\mathbf{e}_1$) and bad ($\mathbf{e}_2$) states. Suppose the transition matrix of $\mathbf{z}$ is

$$\Pi = \begin{bmatrix} 0.995 & 0.010 \\ 0.005 & 0.990 \end{bmatrix}.$$

The entries of $\mathbf{B}(\mathbf{e}_1)$ and $\mathbf{B}(\mathbf{e}_2)$ are assumed to be

$$\mathbf{B}(\mathbf{e}_1) = \begin{bmatrix} 0.997 & 0.800 & 0.030 \\ 0.002 & 0.000 & 0.020 \\ 0.001 & 0.200 & 0.950 \end{bmatrix} \text{ and}$$

$$\mathbf{B}(\mathbf{e}_2) = \begin{bmatrix} 0.995 & 0.700 & 0.010 \\ 0.003 & 0.000 & 0.010 \\ 0.002 & 0.300 & 0.980 \end{bmatrix}.$$

The transition diagram reflecting the above transition matrix of the CKC is depicted in Fig. 4. The numbers in black and black are the transition probabilities when $\mathbf{y}$ is in states $\mathbf{e}_1$ and $\mathbf{e}_2$, respectively. For example, the probability of $\mathbf{y}_k$ going from state 1 to state 2 given $\mathbf{z}_k = \mathbf{e}_1$ is 0.002; and the probability of $\mathbf{y}_k$ going from state 2 to state 3 given $\mathbf{z}_k = \mathbf{e}_2$ is 0.2. With this set of parameters, more than half of the institutions end up with no incidents for the whole year, and around one third end up with one incident. In comparison with the PRC data, the setting of these parameters is reasonable.

B. APPLYING RSMM MODEL

We shall estimate the transition probabilities for each company and the average of the estimates is the final estimate. In other words, the data for each group of 365 observations are processed giving the various filters and hence, the model parameter estimates; each data point being processed constitutes one algorithm step. Fig. 5 displays the estimated results for the various transition probabilities after the completion of 365 algorithm steps.

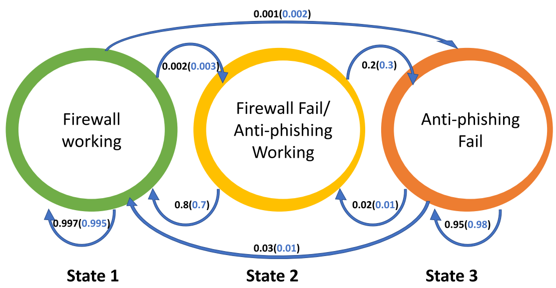


FIGURE 4. **A portrayal of the CKC's state transitions.**

As expected, the estimated values of π_{jr}'s and $c_{sr}(\mathbf{f}_1)$'s converge to their "true" values for a sufficiently large amount of time. However, $c_{sr}(\mathbf{f}_2)$'s and $c_{sr}(\mathbf{f}_3)$'s do not exhibit convergence. Checking Fig. 5(a), we find that the CKC chain visits state 2 or 3 only infrequently, which indicates there is not enough data for the model to update its parameters dynamically going towards the "true" values. We shall see parameter-estimate convergence when there are more cyberattack occurrences as illustrated in Subsection IV-E. Recall that there are 200 sample paths corresponding to 200 institutions, and each path has a one-year length of data points. The transition probability estimates of each institution are plotted in Fig. 6. The dashed lines represent the corresponding 95% confidence interval using the standard errors calculated from the parametric bootstrap [63]. In contrast, the estimates of $c_{sr}(\mathbf{f}_i)$'s are significantly affected by the states of $\mathbf{y}$'s. In particular, $c_{sr}(\mathbf{f}_2)$'s estimates still fluctuate widely throughout the entire period.

Our 'best' estimate of each transition probability is the average of the estimates from the 200 institutions, and they are recorded in the following matrix:

$$\widehat{\Pi} = \begin{bmatrix} 0.996 & 0.014 \\ 0.004 & 0.986 \end{bmatrix},$$
$$\widehat{\mathbf{B}}(\mathbf{e}_1) = \begin{bmatrix} 0.997 & 0.850 & 0.036 \\ 0.002 & 0.000 & 0.019 \\ 0.001 & 0.150 & 0.945 \end{bmatrix}$$
$$\text{and } \widehat{\mathbf{B}}(\mathbf{e}_2) = \begin{bmatrix} 0.993 & 0.721 & 0.034 \\ 0.004 & 0.000 & 0.026 \\ 0.003 & 0.279 & 0.940 \end{bmatrix}.$$

These estimates are further implemented in the premium calculation. The standard errors of these estimates are obtained with the parametric bootstrap and displayed below.

$$\mathrm{SE}\left(\widehat{\Pi}\right) = \begin{bmatrix} 0.00167 & 0.00610 \\ 0.00167 & 0.00610 \end{bmatrix},$$
$$\mathrm{SE}\left(\widehat{\mathbf{B}}(\mathbf{e}_1)\right) = \begin{bmatrix} 0.00298 & 0.09618 & 0.05899 \\ 0.00232 & 0.00000 & 0.04161 \\ 0.00182 & 0.09618 & 0.06664 \end{bmatrix}$$
$$\text{and } \mathrm{SE}\left(\widehat{\mathbf{B}}(\mathbf{e}_2)\right) = \begin{bmatrix} 0.01201 & 0.10842 & 0.05963 \\ 0.00752 & 0.00000 & 0.04935 \\ 0.00970 & 0.10842 & 0.07169 \end{bmatrix}.$$

We observe that the accuracy of final estimates could be improved despite some non-convergence for a single path. The SEs are apparently larger though for $b_{s2}(\mathbf{e}_r)$'s and $b_{s3}(\mathbf{e}_r)$'s as there are fewer transitions starting from state 2 or 3.

C. SIMULATIONS FOR PREMIUMS

The simulation of the breach sizes $A^{(i)}$'s is performed using the DTP distribution with the parameters $u_1 = u_2 = 1$, $v_1 = 2,202,078$, $v_2 = 11,818,259$, $\delta_1 = 0.0668$, and $\delta_2 = 0.0068$. The parameters are chosen based on the PRC dataset [54] from 2013 to 2017. The starting year corresponds to that of the data set used to demonstrate the functional-form assumption in (28) whilst the ending year is chosen based on the completeness of the PRC data. We compare the PRC data with the raw incident reports in one of its major sources, the U.S. Department of Health and Human Services Office for Civil Rights, and discover that the numbers of cyberattacks are not consistent from 2018. In particular, we select cyberattacks of medical-organisations type because data from medical organisations are sufficiently available and reliable. We then classify the cyberattacks into two subsets (subset 1 and subset 2) based on attack types whether hacking is involved or not with the goal of setting up parameters for $i = 1$ and $i = 2$, separately.

Next, we randomly choose 60 samples from each subset 1 and subset 2. The parameters u_i's, v_i's δ_i's are taken as the MLEs of the samples. The above process is replicated 10 times and the averages of the results in each iteration become our inputs in this experiment. Note that the above process only assists us in finding reasonable parameter values to conduct the simulation.

In practice, the parameters must be estimated based on the data of a group of companies that share similar traits with the company seeking cyber risk insurance such as the type, size and historical breach records. With the simulated number of records per attack, we shall obtain the corresponding losses under two loss assumptions in Subsection III-B: (i) proportionality and (ii) functional-form assumptions. In particular, we assign \$161 for the cost-per-record assumption (i), which is the global average cost in 2021 [39]. By applying the Euler discretisation scheme to (24) and simulating the interest rate process with $\tau = 55.8711, a = 0.0739$ and $\nu = 0.3452$, we get the discount factor in percentages. The simulated parameters are set with the reference to the 1-year U.S. T-bill yields in 2021. The simulated annual interest rates range from 0.006% to 0.162%.

The cyber-risk insurance premium for three months, six months and one year could be computed now with the generated discount factors and loss random variables. Suppose that there is no deductible limit in the insurance contract and the loss is paid on the day that it is incurred. As mentioned in the previous section, we apply two methods based on: (i) standard-deviation and (ii) exponential-premium principles following equations (30) and (32), respectively. In principle (i), we set $\lambda_r = 0, 0.1$; whilst in principle (ii), we have

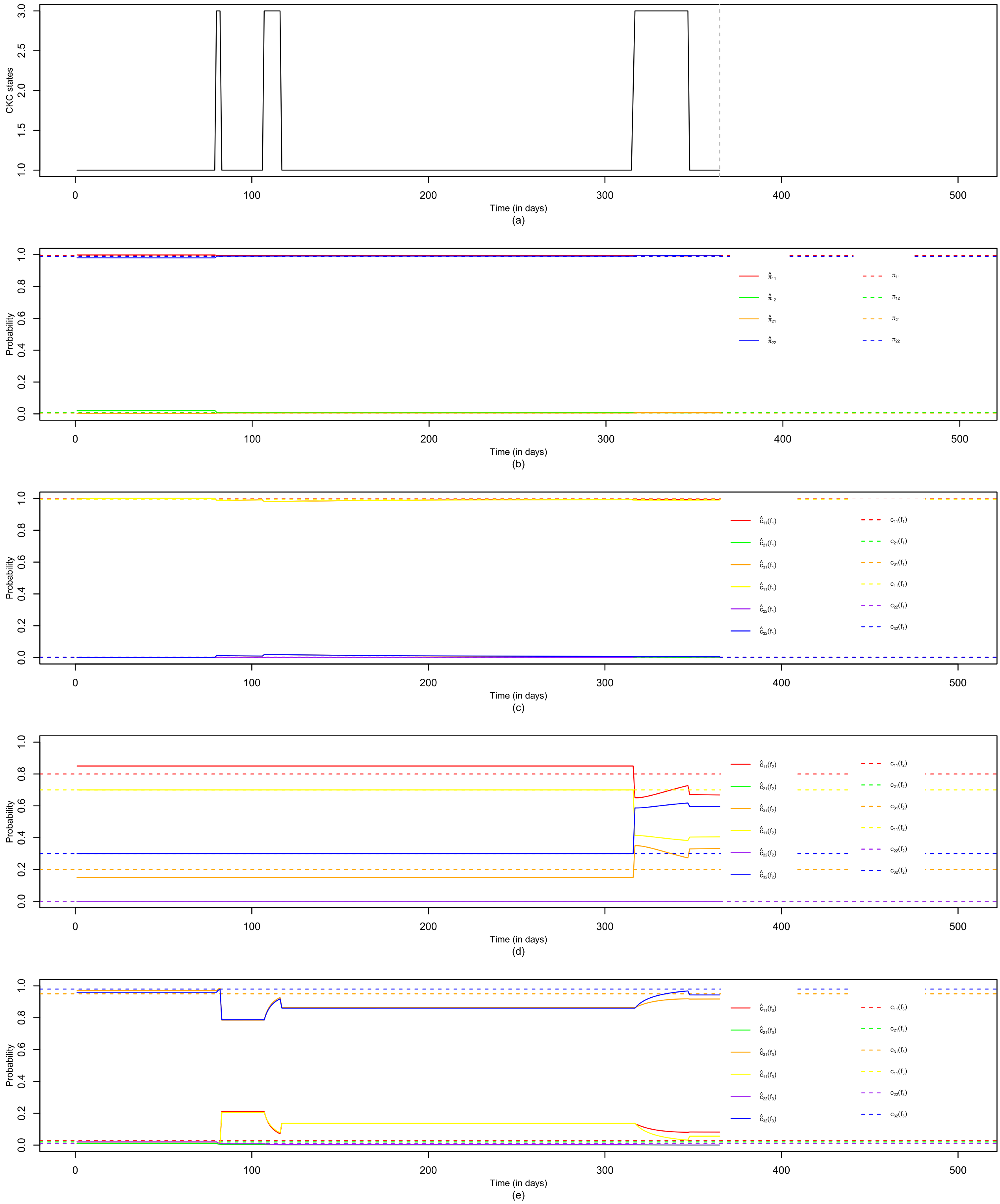


FIGURE 5. Evolution of the transition probability estimates for π_{jr} and $c_{sr}(\mathbf{f}_i)$ ($i = 1, 2, 3$) on a daily basis for 365 algorithm steps.

$\kappa = \frac{1}{1000}, \frac{1}{10{,}000}, \frac{1}{100{,}000}$. The same values of λ_r and κ are used in [25].

There are one million one-year scenarios generated by the simulation, and these scenarios are divided into 200 subgroups with equal sizes of 5,000. The expectation and the variance involved in (30) and (32) are both estimated from 5,000 scenarios in each subgroup. We display the means and the standard errors (SEs) of the premiums obtained per

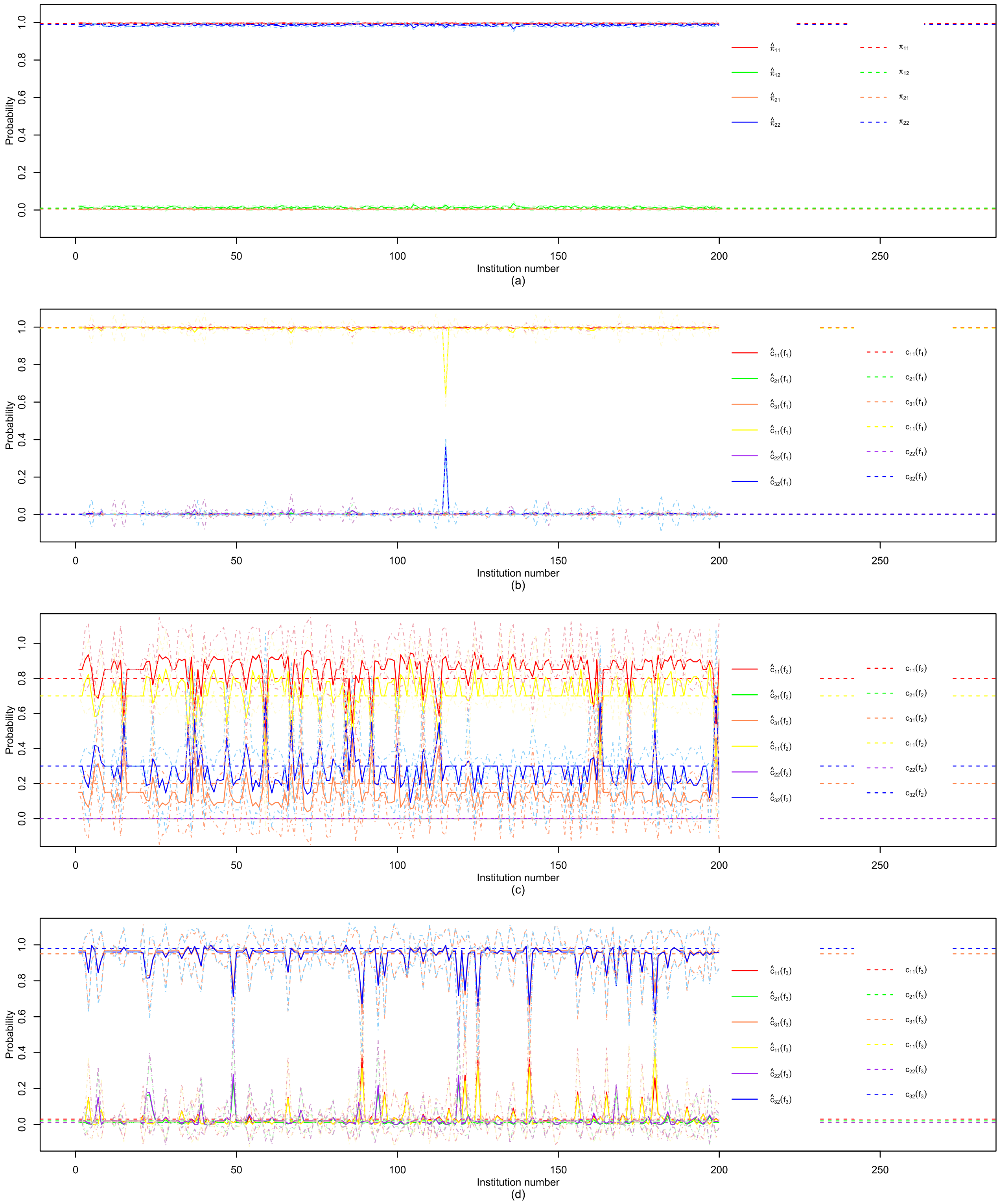


FIGURE 6. **Estimates of π_{jr} and $c_{sr}(\mathbf{f}_i)$ ($i = 1, 2, 3$) on a 12-hour interval (200 estimates for each parameter).**

subgroup in Table 1; the average premiums and their SEs based on 200 subgroups are shown for each combination of the loss assumption and the premium principle with terms of 3, 6, and 12 months. In particular, each 3-month scenario is extended to 6 and 12-month scenarios with the same random seed. From Table 1, we have the following findings:

TABLE 1. Premiums in millions under the (i) standard-deviation and (ii) utility-based premium principles.

Assumption		Principle (i) $\lambda_r = 0$	Principle (i) $\lambda_r = 0.1$	Principle (ii) $\kappa = 10^{-3}$	Principle (ii) $\kappa = 10^{-4}$	Principle (ii) $\kappa = 10^{-5}$
(i) Proportionality assumption	**Term: 3 months**					
	Mean	29.2679	44.5373	48.4185	30.4883	29.3854
	SE	2.0828	2.9765	4.8832	2.2236	2.0962
	Term: 6 months					
	Mean	61.0705	82.7197	99.1278	63.5178	61.3063
	SE	3.0999	4.0124	7.1209	3.3029	3.1191
	Term: one year					
	Mean	118.6995	148.6075	191.1689	123.3661	119.1490
	SE	4.2738	5.1951	10.4878	4.5608	4.3010
(ii) Functional-form assumption	**Term: 3 months**					
	Mean	10.5781	15.2208	11.7794	10.687	10.5890
	SE	0.6248	0.8422	0.7397	0.6348	0.6258
	Term: 6 months					
	Mean	22.0634	28.6289	24.4527	22.2814	22.0850
	SE	0.9358	1.1540	1.0981	0.9499	0.9372
	Term: one year					
	Mean	42.5934	51.5152	46.9889	42.9954	42.6332
	SE	1.2807	1.4923	1.4981	1.2994	1.2827

1) The premiums are proportional to the terms of the cyber-risk insurance given a loss assumption and a premium principle.
2) The means and the SEs increase as λ_r increases or as κ decreases under both assumptions.
3) For a fixed term and a given assumption, the premiums and the SEs under principle (i) with $\lambda_r = 0$ are close to those under principle (ii) with $\kappa = 10^{-5}$.
4) The means and SEs under the proportionality assumption are much larger than those under the functional-form assumption.
5) The SE of premiums could be as large as 10 million under the proportionality assumption.

The first finding could be explained by the formula for $E[S_T]$ whenever the role of interest rates is negligible and a sufficient number of scenarios are simulated. The second observation is straightforward. The third one supports the theoretical result that as κ goes to 0, the expected premium will converge to the expected total losses. Additionally, we verified what lead to the fourth finding. In the original literature of assumption (ii) [40], the number of records per cyberattack in the data that supported the model is at most 100,000 whilst the simulated severities could be up to 11 million. It seems that the log-log model should be updated and we should rely on the results from the proportionality-based model. As for the last one, the aim is to decrease the SE of the premium. The SE would be 21.4831 million when there are 200 subgroups with 1000 scenarios in each subgroup whilst the SE increases to 22.3257 million when there are 1000 subgroups with a size of 1000 scenarios. This suggests that increasing the size of the subgroups rather than the number of subgroups could decrease the SE. Therefore, including more institutions will help in premium determination more accurately.

Furthermore, we apply a pairwise t or F test to check whether the findings in the second and third observations are statistically significant. Below are our statistical test results:

1) The premiums and the SEs increase significantly as λ_r increases under both assumptions.
2) Under the proportionality assumption, the premiums and SEs of premiums decrease significantly when κ decreases but the SEs do not differ significantly as κ changes from 10^{-4} to 10^{-5} for all the three contract's terms.
3) Under the functional-form assumption, cases become more complicated. In a three-month policy, the premiums and SEs of premiums decrease significantly when κ decreases from 10^{-3} to 10^{-4} but not from 10^{-4} to 10^{-5}. In comparison, the premiums decrease significantly when κ decreases from 10^{-4} to 10^{-5} for other policy terms.
4) For a fixed term and a given assumption, a paired $t-$test demonstrates that the premiums calculated when $\kappa = 10^{-5}$ are significantly larger than the premiums calculated when $\lambda_r = 0$ based on a p-value of less than 10^{-22}. On the contrary, the variances of the premiums under these two cases are the same with a p-value of over 0.99 from the F test.

In summary, by adjusting the risk-averse parameters, we could achieve different levels of premiums. The changes in the means and their SEs are term-independent when λ_r or κ changes under the proportionality assumption; more specifically, the significance of the change is the same for all the three terms. We find, however, that the means are more sensitive to the change in κ than to the change in SEs.

Romanosky et al. [55] pointed out that quoted cyber-insurance premiums a few years ago are typically for policies with limits of \$100,000 and deductible of \$10,000. Noticeably, our simulated-based premiums are unrealistically high. To reflect the present business settings in our valuation, we calculate the premiums with a deductible of \$50,000 and a payment limit of \$500,000 in Table 2. The premiums remarkably drop to a reasonable range and the premiums under two

TABLE 2. Premiums in thousands with a deductible of \$50,000 and a limit of \$500,000 under the (i) standard-deviation and (ii) utility-based premium principles.

Assumption		Principle (i)		Principle (ii)		
		$\lambda_r = 0$	$\lambda_r = 0.1$	$\kappa = 10^{-3}$	$\kappa = 10^{-4}$	$\kappa = 10^{-5}$
(i) Proportionality assumption	**Term: 3 months**					
	Mean	65.4365	82.2964	81.4809	66.8753	65.5788
	SE	2.1300	2.3628	2.5467	2.1695	2.1339
	Term: 6 months					
	Mean	125.3017	146.9595	150.6062	127.6661	125.5364
	SE	2.9709	3.1427	3.3170	3.0079	2.9747
	Term: one year					
	Mean	218.5111	243.2706	249.4873	221.5818	218.8176
	SE	3.7154	3.7623	3.7553	3.7263	3.7166
(ii) Functional-form assumption	**Term: 3 months**					
	Mean	65.5714	82.4335	81.6185	67.0106	65.7137
	SE	2.1320	2.3650	2.5490	2.1716	2.1359
	Term: 6 months					
	Mean	125.8114	147.4723	151.1151	128.1764	126.0462
	SE	2.9645	3.1349	3.3075	3.0012	2.9682
	Term: one year					
	Mean	220.2729	245.0043	251.1486	223.3363	220.5787
	SE	3.7373	3.7819	3.7713	3.7476	3.7384

loss assumptions also become practically comparable. Still, if the coverage is set to a maximum of \$500,000, the product could be deemed insufficient to meet the needs of the client. Given the situation that more than half of the underwriting institutions experience a cyberattack, it is reasonable that the premiums could be raised at a level half of the payment limit. Of course, the premiums could be lowered if the underwriting group has lower cyber risk.

D. SEMI-PARAMETRIC APPROXIMATION OF TOTAL LOSSES

Although the simulation of the total discounted loss S_T is helpful when S_T does not have an explicit functional form, the simulation method requires considerable computing time and resources. To remedy this issue, the distribution of S_T is characterised by some accurate approximation. Given our discussion in Subsection IV-C, we only consider total losses obtained under the proportionality assumption. To get a rough idea of what distribution could approximate S_T, we plot the histograms of 1,000,000 simulated S_T's when calculating premiums for each policy term in Fig. 7(a)-(c). We notice that there is a large portion of S_T being zero and the range of S_T is wide. Therefore, we introduce a transformed total loss X_T^L, obtained by truncating S_T at zero and taking the logarithm of positive S_T's so that

$$X_T^L = \begin{cases} \text{undefined}, & S_T = 0 \\ \log(S_T), & S_T > 0. \end{cases}$$

The respective cumulative distribution functions (CDFs) of X_T^L and S_T are

$$F_{X_T^L}(x) = \frac{F_{S_T}(e^x) - F_{S_T}(0)}{1 - F_{S_T}(0)}, \quad x \in \mathbb{R},$$

$$F_{S_T}(x) = \begin{cases} F_{S_T}(0), & x = 0 \\ F_{S_T}(0) + \left(1 - F_{S_T}(0)\right) F_{X_T^L}(\log x), & x > 0 \end{cases}.$$

Empirically, $F_{S_T}(0)$ could be approximated by the proportions of zero-total losses. In our simulated study, $F_{S_T}(0) = 0.8608$, 0.7264, and 0.5130 for 3-month, 6-month, and one-year terms, respectively. The histograms of X_T^L's are plotted in Fig. 7(d)-(f). We see the apparent multi-mode patterns of the three histograms. For this reason, we use the mixture models due to their flexibility in capturing the distribution of X_T^L's. McLachlan et al. [50] provided recently a review of finite mixture models and noted that mixture models are increasingly utilised as a convenient, semi-parametric way to model unknown distributional shapes. For example, Park and Lord [53] modelled vehicle-crash occurrences by a two-component finite mixture of negative regression models. There are also research outputs devoted to the parameter estimation of mixture models. In this paper, we use a fast and stable algorithm proposed by Wang [64] to estimate the non-parametric mixing component. The algorithm codes are included in the R package "nspmix" [65].

The density of a mixture model has the form

$$f(x; H, \varsigma) = \int_\Omega f(x; \vartheta, \varsigma) \mathrm{d}H(\vartheta), \tag{33}$$

where ς is the structural parameter, $f(x; \vartheta, \varsigma)$, $x \in \mathcal{X}$, $\vartheta \in \Omega \subset \mathbb{R}$ is the component density, and $H(\vartheta)$ is the mixing distribution function. In particular, we restrict the non-parametric $H(\vartheta)$ as a discrete distribution function with finite mass points. Denote $\mathbb{I}_{\vartheta_j}$ as the indicator random variable at $\vartheta_j \in \Omega$ for $j = 1, 2, \ldots, M$. Let $H(\vartheta) = \sum_{j=1}^M w_j \mathbb{I}_{\vartheta_j}$, where $w_1, w_2, \ldots, w_M > 0$, and $\sum_{j=1}^M w_j = 1$. Now, the density (33) could be rewritten as

$$f(x; \boldsymbol{w}, \boldsymbol{\vartheta}, \varsigma) = \sum_{j=1}^M w_i f(x; \vartheta_j, \varsigma),$$

where $\boldsymbol{w} = (w_1, w_2, \ldots, w_M)^\top$ and $\boldsymbol{\vartheta} = (\vartheta_1, \vartheta_2, \ldots, \vartheta_M)^\top$. Given that X_T^L takes both positive and negative values, its

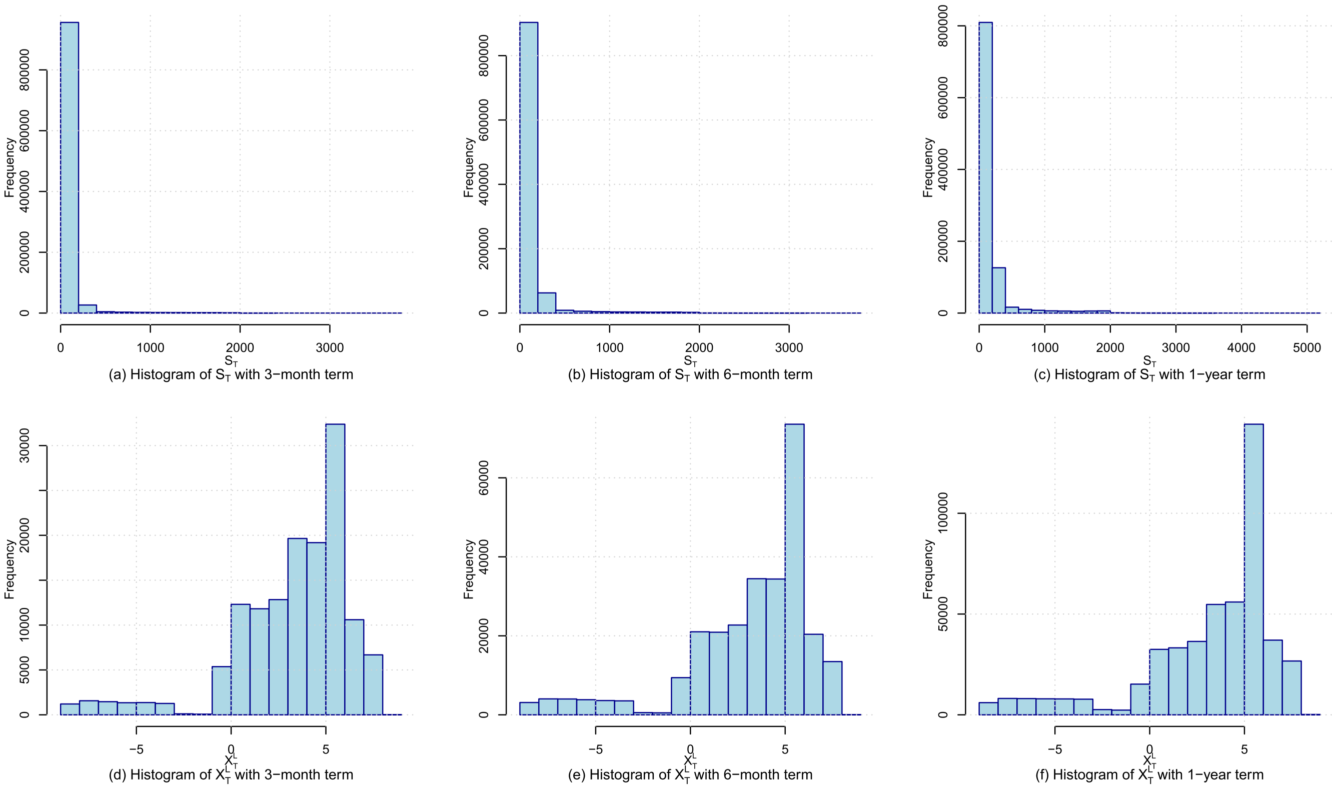


FIGURE 7. Histograms of S_T and X_T^L under different terms.

model candidate could be the normal density

$$f(x; \vartheta_j, \varsigma) = \frac{1}{\sqrt{2\pi}\,\varsigma} e^{-\frac{(x-\vartheta_j)^2}{2\varsigma^2}}, \quad x \in \mathbb{R}.$$

Wang [64] proposed the CNM algorithm to estimate the MLEs of $\boldsymbol{w}$, $\boldsymbol{\vartheta}$ and M with a fixed ς. In the CNM algorithm, CN stands for the constrained Newton method and M for the multiple support points being added at each iteration. As ς decreases, the number of normal components used for approximation increases. To minimise the possibility of overfitting, we implement the two most widely used model-selection tools: the Akaike Information Criterion (AIC) [2] and the Bayesian Information Criterion (BIC) [57] given by

$$\begin{aligned} \text{AIC} &= 2 \times \text{number of parameters} - 2\log\mathcal{L}, \\ \text{BIC} &= \text{number of parameters} \\ &\quad \times \log\ (\text{number of data}) - 2\log\mathcal{L}, \end{aligned}$$

where $\mathcal{L}$ is the maximum value of the likelihood function for a model under consideration. Notice that the effect of the penalty terms is substantially influenced by the number of data. In our case, the respective number of X_T^L's are 139,249, 273,610 and 486,955 for the 3, 6 and 12-month terms. Our target is to determine the proper ς value ranging from 0.1 to 2. The upper limit of ς is selected by calculating approximately the standard deviations of the data spanning the two bumps in the three histograms of X_T^L's. If the AIC and BIC are calculated using the full data sets, we only observe a monotonic trend over the range of ς, which is not useful to determine the ς. Therefore, we compute the AICs and BICs for 500 subgroups, each with 2,000 scenarios, regrouped from the data simulated in Subsection IV-C. We do not directly use the same 200 subgroups as we aim to control the size of the subgroups to better perform the AIC/BIC analysis and conduct statistical tests. After truncation and taking the logarithm, the numbers of data points of X_T^L's in each subgroup are roughly 278, 547 and 974 for the 3-, 6- and 12-month terms, respectively. In addition, we also conduct the Kolmogorov-Smirnov (KS) [17] and Anderson-Darling (AD) [59] goodness-of-fit tests for each subgroup of X_T^L's for a fixed ς. The null hypothesis of both tests is that the data follow a specified distribution, which is a normal mixture in our case. The KS test tends to be more sensitive near the centre of the distribution than at the tails whilst the AD test is a modification of the KS test and puts more weight to the tails. Additionally, the critical values of the KS test do not depend on the specific distribution being tested whilst the AD test relies on the specific distribution in calculating the critical values.

Next, we discuss how to choose ς for different terms of the insurance policy. For the 3-month X_T^L's, refer to Fig. 8. The medians of AICs are minimised at $\varsigma = 0.9$ and 1 whilst the medians of BICs decrease to the minimum at $\varsigma = 1.5$, which are indicated by the red dashed lines. However, we do

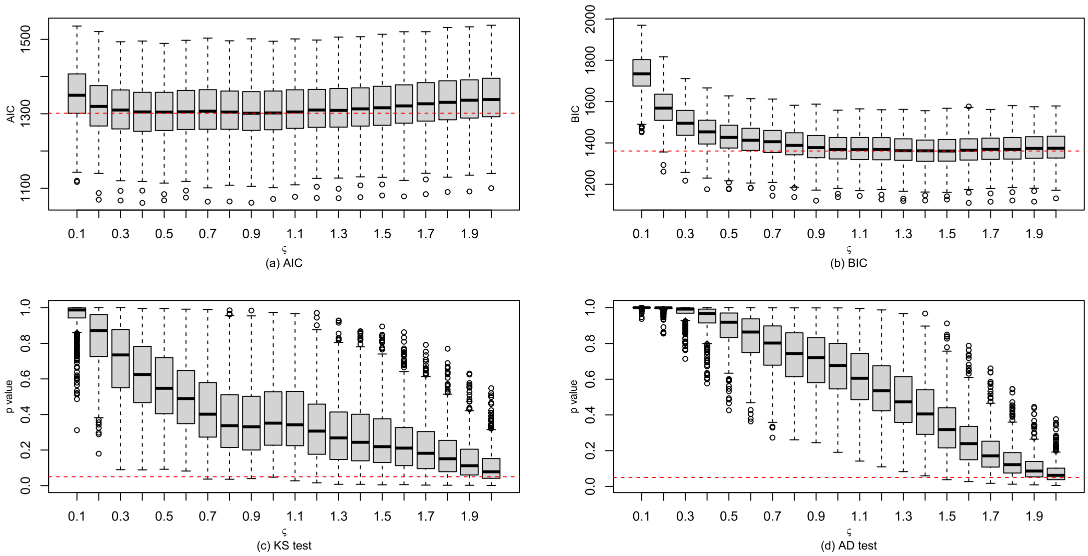


FIGURE 8. AICs, BICs and goodness-of-fit test results in a 3-month policy.

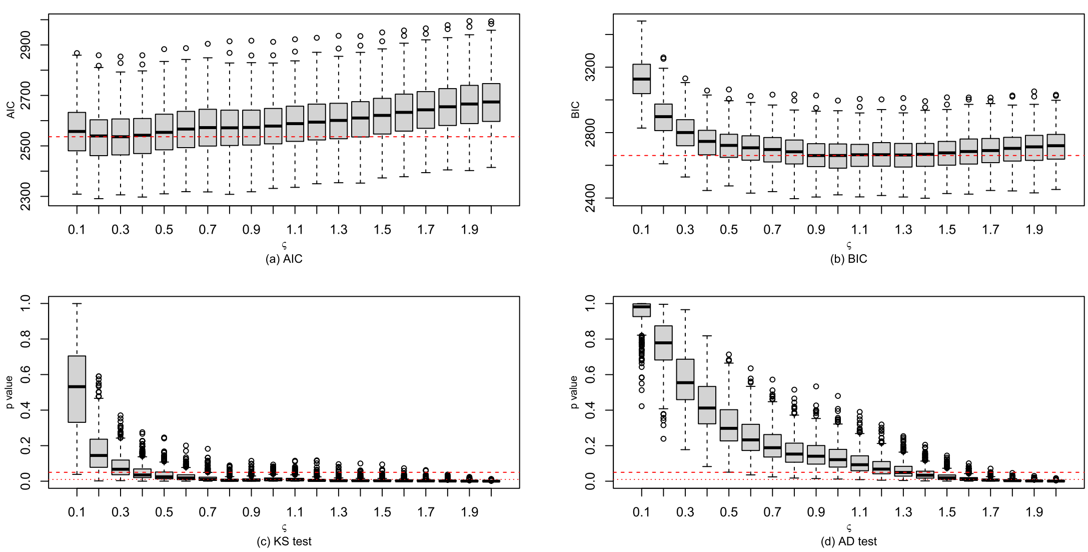


FIGURE 9. AICs, BICs and goodness-of-fit test results in a 6-month policy.

not observe significant differences in the BICs when $\varsigma \geq 0.9$. The p-values of the KS tests are greater than 5% when ς is equal to 1 or below 0.6 whilst the p-values of the AD tests remain above 5% when ς is below 1.2. Therefore, we let $\varsigma = 1$ when the policy term is 3 months. Inspecting Fig. 9, we take $\varsigma = 0.4$ for the 6-month X_T^L's. We find that the AICs and BICs are larger and the p-values are generally smaller than the quantities displayed in Fig. 8. This is mainly caused by the increase in the number of data points in each subgroup, from 278 to 547. There is no choice of ς that satisfies every criterion and retains the null hypothesis in every goodness-of-fit test. We may choose a ς according to the purpose of modelling. If we prefer a model with fewer parameters, we let $\varsigma = 1$ and end up with a mixture model with 9 components. The medians of the BICs are at the minimum level and the $p-$values of the AD tests barely exceed 1%. In contrast, if we pursue better fitting results under the KS tests, we could choose $\varsigma = 0.4$. In what follows, we present the fitting results for the case of $\varsigma = 0.4$. With a similar analysis that relies on Fig. 10 ς is set to 0.5 for the case of a one-year policy.

With given ς's, we obtained the MLEs of $\boldsymbol{w}$, $\boldsymbol{\vartheta}$ and M in all subgroups. The fitted parameters are presented in Table 3.

Moreover, we plot the fitted density (black curve), superimposing it on the histograms of X_T^L's in Fig. 11(a)-(c). The proportions $\boldsymbol{w}$ are represented by the vertical black lines that stick out from the hollow black points, corresponding to $\boldsymbol{\vartheta}$. The latter three plots pictorially present how the normal mixture models fit our data with the quantile-quantile (Q-Q)

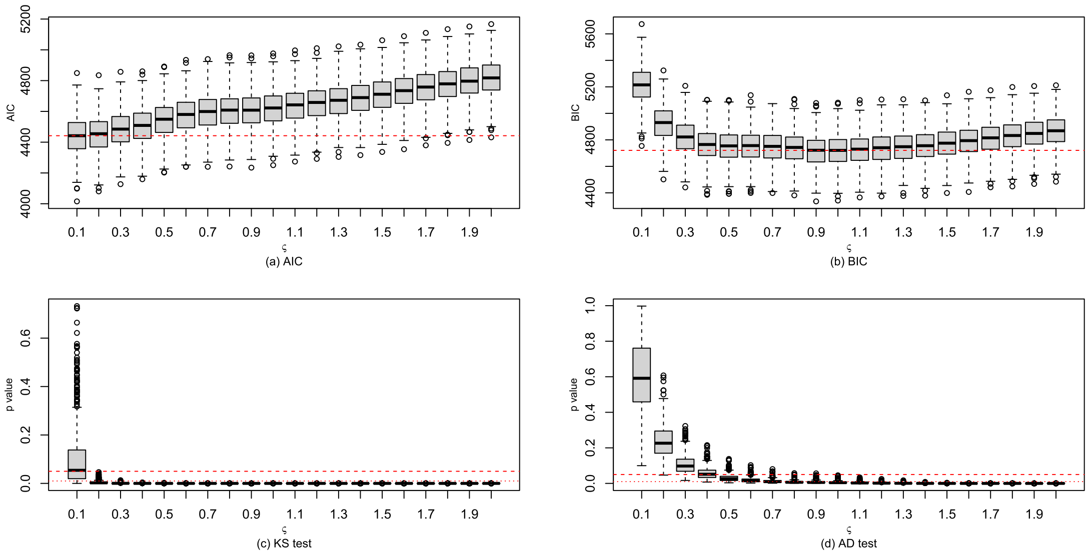


FIGURE 10. AICs, BICs and goodness-of-fit test results in a 1-year policy.

TABLE 3. Fitted parameter values of a mixture of normal models.

3-month, $\varsigma = 1$, M =8								
$\boldsymbol{w} =$	(0.007701,	0.024950 ,	0.027013 ,	0.154451 ,	0.067751 ,	0.274141,	0.006110 ,	0.437883)$^\top$
$\boldsymbol{\vartheta} =$	(-7.362348,	-7.190936,	-4.448352,	0.865405,	1.036816,	3.436577,	5.493516,	5.664927)$^\top$
6-month, $\varsigma = 0.4$, M =32								
$\boldsymbol{w} =$	(0.010572,	0.006447,	0.012117,	0.005449,	0.005368,	0.003720,	0.006505,	0.004657
	0.003016,	0.006596,	0.003944,	0.004667,	0.008888,	0.000047,	0.000931,	0.001384,
	0.000246,	0.015563,	0.070981,	0.027948,	0.042736,	0.006916,	0.062748,	0.001656,
	0.099276,	0.022333,	0.051529,	0.114686,	0.261713,	0.054250,	0.071946,	0.011165)$^\top$
$\boldsymbol{\vartheta} =$	(-8.220620,	-8.049192,	-7.192049,	-6.506335 ,	-6.334906,	-5.820621,	-5.649192,	-5.134906,
	-4.963478,	-4.449192,	-4.277764 ,	-3.592049 ,	-3.420621,	-2.220621,	-2.049192,	-1.706335,
	-1.534907,	0.007950,	0.179379,	1.036522,	1.207950,	1.893664,	2.065093,	2.922236,
	3.093664,	3.607950,	3.779379,	4.465093,	5.665093,	5.836521,	7.036521 ,	7.207950)$^\top$
1-year, $\varsigma = 0.5$, M =23								
$\boldsymbol{w} =$	(0.020510,	0.002302,	0.018419,	0.000238,	0.014456,	0.000800,	0.002794,	0.008365,
	0.008974,	0.019472,	0.000408,	0.003571,	0.003404,	0.066606,	0.019218,	0.048898,
	0.055246,	0.006660,	0.116027,	0.132604,	0.058653,	0.317055,	0.075320)$^\top$	
$\boldsymbol{\vartheta} =$	(-8.036385,	-7.861755,	-6.813974,	-6.639344,	-5.766192,	-5.591562,	-5.416932,	-4.718411,
	-4.543781,	-3.496000,	-3.321370,	-1.924328,	-1.749698,	0.171235,	0.345865,	1.219016,
	1.917537 ,	2.092167,	3.139948,	4.187729,	5.584771,	5.759401,	7.156443)$^\top$	

plots. With a large M, the black points are closely aligned to the dashed line, suggesting a very good fit. In conclusion, the distribution of the transformed total losses could be approximated by a semi-parametric mixture of normals model.

E. FURTHER COMPARISON

We also compare two cases in which the probabilities of transition from one state to another in the first case are lower than the corresponding probabilities in the second case. This comparison is pertinent in gauging the frequency of cyberattacks. Below are cases A and B encapsulated in the transition matrices governing the dynamics of the Markov chains.

Case A:

$$\Pi = \begin{bmatrix} 0.995 & 0.010 \\ 0.005 & 0.990 \end{bmatrix},$$

$$\mathbf{B}(\mathbf{e}_1) = \begin{bmatrix} 0.997 & 0.800 & 0.030 \\ 0.002 & 0.000 & 0.020 \\ 0.001 & 0.200 & 0.950 \end{bmatrix}$$

$$\text{and } \mathbf{B}(\mathbf{e}_2) = \begin{bmatrix} 0.995 & 0.700 & 0.010 \\ 0.003 & 0.000 & 0.010 \\ 0.002 & 0.300 & 0.980 \end{bmatrix}.$$

Case B:

$$\Pi = \begin{bmatrix} 0.95 & 0.10 \\ 0.05 & 0.90 \end{bmatrix},$$

$$\mathbf{B}(\mathbf{e}_1) = \begin{bmatrix} 0.94 & 0.70 & 0.05 \\ 0.04 & 0.00 & 0.05 \\ 0.02 & 0.30 & 0.90 \end{bmatrix}$$

$$\text{and } \mathbf{B}(\mathbf{e}_2) = \begin{bmatrix} 0.83 & 0.50 & 0.02 \\ 0.14 & 0.00 & 0.03 \\ 0.03 & 0.50 & 0.95 \end{bmatrix}.$$

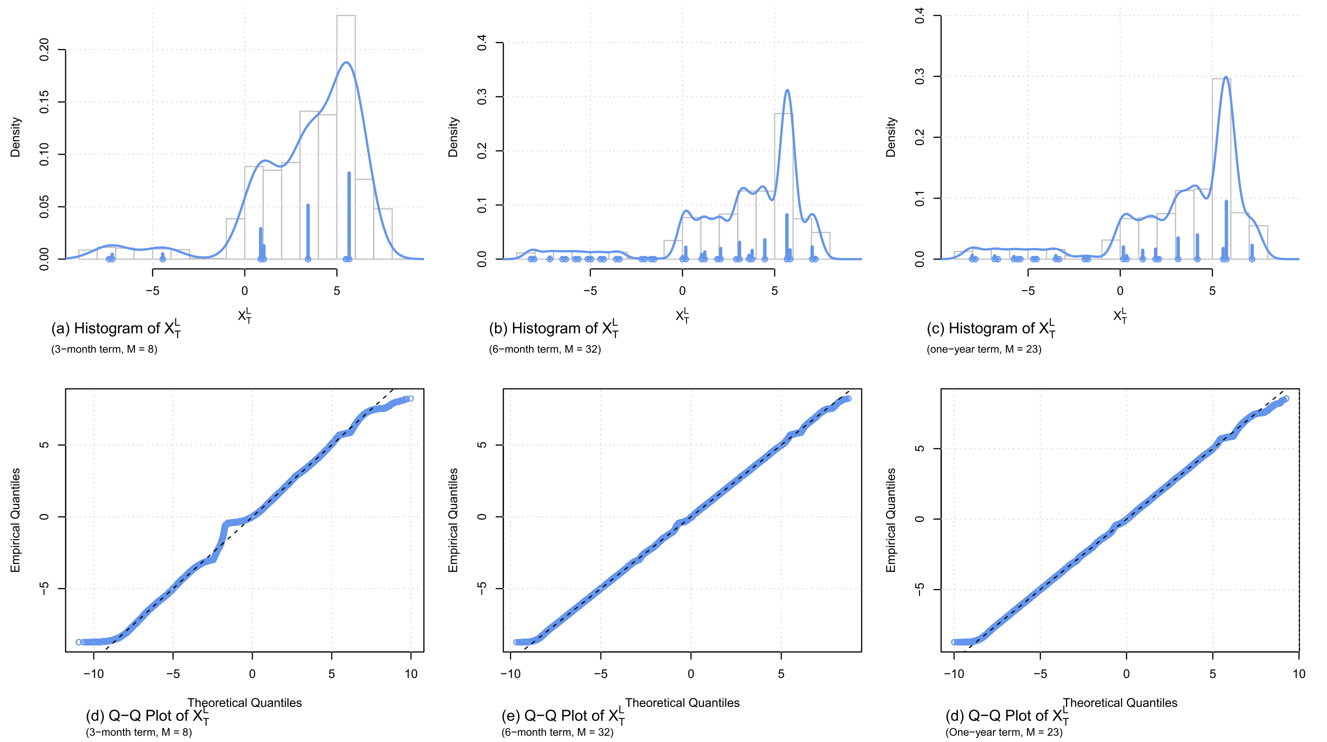


FIGURE 11. Fitted density and corresponding Q-Q plots.

In Case A, the number of successful attacks is 123 and in Case B, this number is 2351 over a one-year period for 200 institutions. The movements of the estimated parameters in Case A are plotted in Figs. 5 and 6. In contrast, Fig. 12 traces the daily evolution of the estimated parameters for a single institution, and Fig. 13 illustrates the one-year final estimates for each of the 200 institutions. The estimates and their SEs under Case B are as follows:

$$\widehat{\Pi} = \begin{bmatrix} 0.920 & 0.083 \\ 0.080 & 0.917 \end{bmatrix},$$

$$\widehat{\mathbf{B}}(\mathbf{e}_1) = \begin{bmatrix} 0.932 & 0.637 & 0.051 \\ 0.056 & 0.000 & 0.051 \\ 0.011 & 0.363 & 0.898 \end{bmatrix},$$

$$\widehat{\mathbf{B}}(\mathbf{e}_2) = \begin{bmatrix} 0.879 & 0.577 & 0.032 \\ 0.080 & 0.000 & 0.038 \\ 0.041 & 0.423 & 0.930 \end{bmatrix},$$

$$\mathrm{SE}\left(\widehat{\Pi}\right) = \begin{bmatrix} 0.01158 & 0.01154 \\ 0.01155 & 0.01148 \end{bmatrix},$$

$$\mathrm{SE}\left(\widehat{\mathbf{B}}(\mathbf{e}_1)\right) = \begin{bmatrix} 0.02390 & 0.10763 & 0.02380 \\ 0.02120 & 0.00000 & 0.02334 \\ 0.01028 & 0.10763 & 0.03354 \end{bmatrix} \text{ and}$$

$$\mathrm{SE}\left(\widehat{\mathbf{B}}(\mathbf{e}_2)\right) = \begin{bmatrix} 0.03462 & 0.10745 & 0.01690 \\ 0.02544 & 0.00000 & 0.02060 \\ 0.02104 & 0.10745 & 0.02782 \end{bmatrix}.$$

In Case B, the behaviour of $\widehat{c}_{sr}(\mathbf{f}_3)$ exhibits better convergence patterns in Fig. 12, especially in Fig. 12(e). There are also marked reductions in the SEs of $\widehat{c}_{sr}(\mathbf{f}_3)$'s. Indeed, Fig. 13(d) confirms as well that the estimated values are less volatile. However, we do not observe any improvement in the behaviour $\widehat{c}_{sr}(\mathbf{f}_2)$'s. We find that the percentage of being in state 2 only increases from 0.28% in Case A to 5.18% in Case B compared to the increase from 4.92% to 40.72% of being in state 3. This suggests that if the frequency of cyber-attacks is low, implementing the model on a data set over a larger group of institutions and a relatively longer time period or extending the time unit per observation must be considered. Otherwise, the estimations' accuracy could be affected. Also, if we have a high-powered computing machinery featuring lots of memory and storage as well as the capacity to complete complex calculations at a faster speed, the estimations that entail longer periods could definitely be carried out.

TABLE 4. Comparison of the RSMM with ACD models. Bolded numbers indicate criterion values corresponding to the best model.

Model	log-likelihood	AIC	BIC
RSMM	**-556.035**	**1156.070**	**1358.431**
ACD	-732.568	1471.136	1479.573
LACD1	732.536	1471.073	1479.510
LACD2	-734.589	1475.178	1483.614

In our framework, we proposed an RSMM to model the process of cyber-attack occurrences. Xu et al. [69] modelled the inter-arrival times of cyberattacks with the autoregressive conditional mean (ACD) model. We complete our model comparison, with the ACD as the benchmark, using the AIC

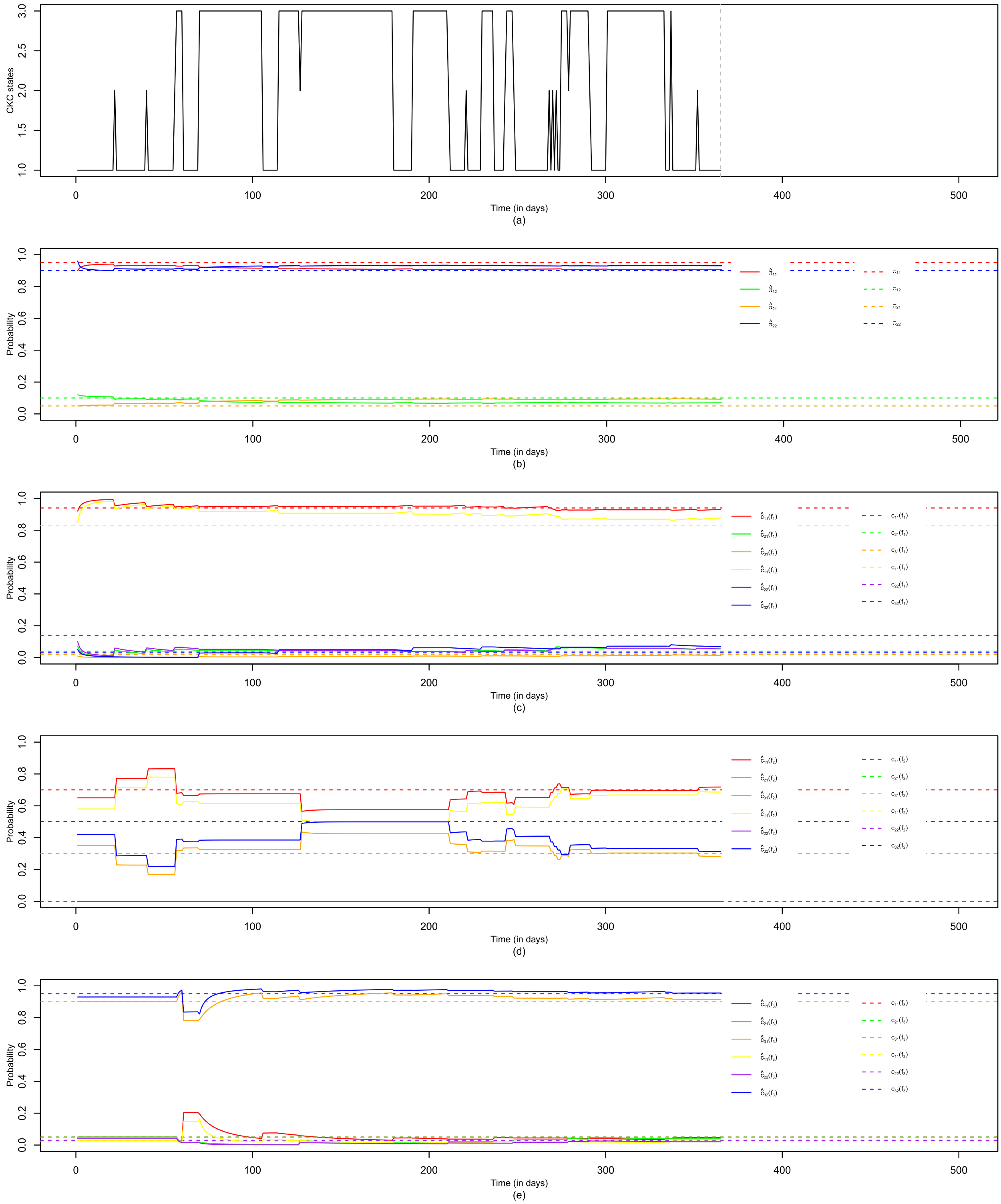


FIGURE 12. Evolution of the transition probability estimates for π_{jr} and $c_{sr}(\mathbf{f}_i)$ ($i = 1, 2, 3$) on a daily basis in Case B for 365 algorithm steps.

and BIC metrics

$$\begin{aligned} \text{AIC} &= 2 \times (n^2 + m^2 n) - 2 \log \mathcal{L}_a, \\ \text{BIC} &= (n^2 + m^2 n) \log T - 2 \log \mathcal{L}_a, \end{aligned}$$

where

$$\mathcal{L}_a = \prod_{k=1}^{T} \sum_{i=1}^{2} \langle \mathbf{y}_{k-1}, \mathbf{f}_i \rangle \langle \mathbf{y}_k, \mathbf{f}_3 \rangle \langle \mathbf{y}_k, \mathbf{B}\,(\mathbf{z}_{k-1})\,\mathbf{y}_{k-1} \rangle,$$

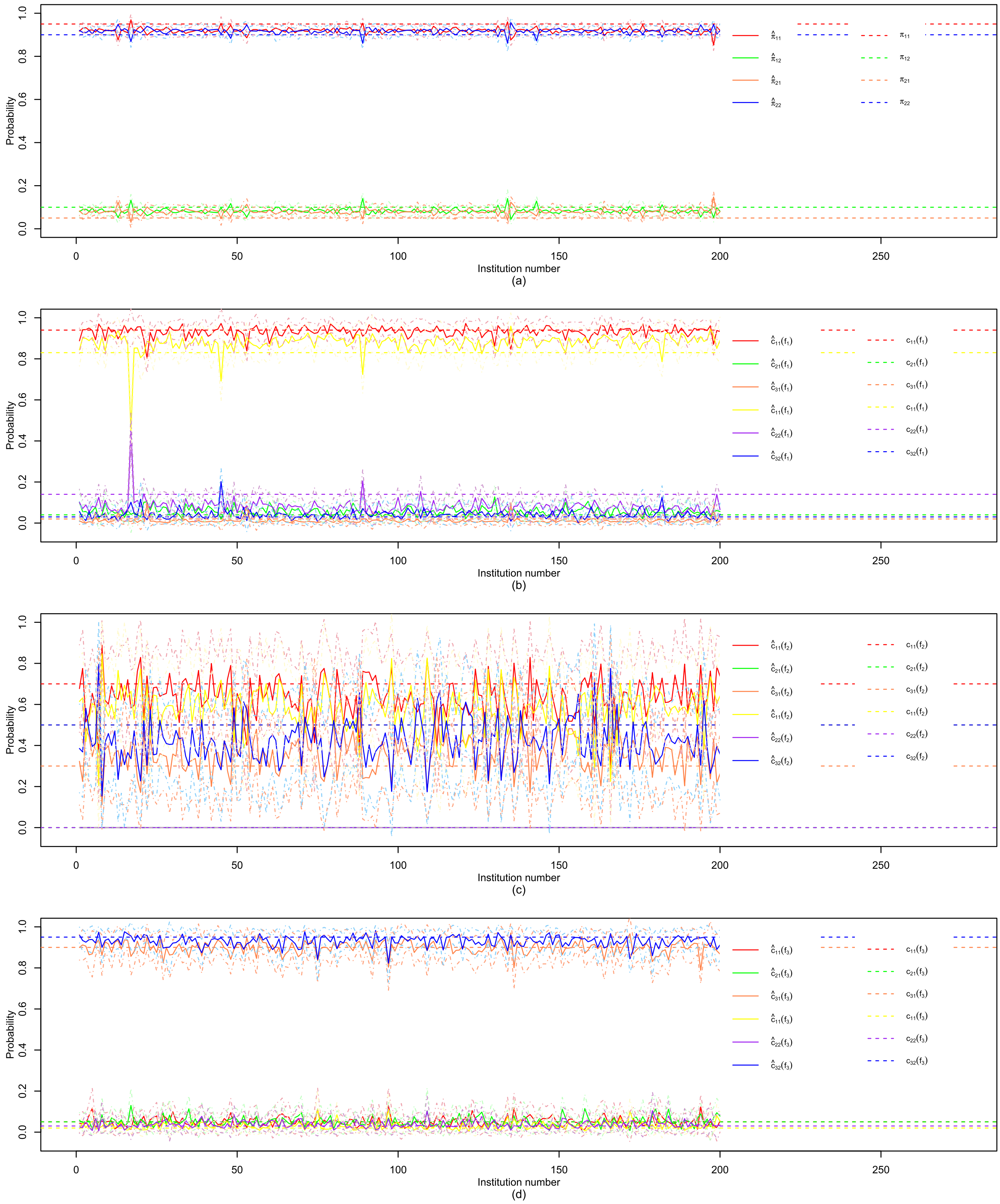


FIGURE 13. Estimates of π_{jr} and $c_{sr}(\mathbf{f}_i)$ $(i = 1, 2, 3)$ in Case B for each institution (based on 200 estimates for each parameter).

and $n^2 + m^2 n$ is the number of transition probabilities in the RSMM model. Note that to make a comparison with the ACD models, we calculate the likelihood function of the RSMM based on only successful cyberattacks. The results are displayed in Table 4. The models in the last three rows are the standard ACD, the type-I log-ACD and type-II ACD models, respectively. The computations are performed using the package "ACDm" in the statistical software R [7]. The details of the ACD models are presented in Appendix F. Clearly, the RSMM is superior to the ACD models in terms of the AIC and BIC. This result is based on RSMM's simulated data. It is worth noting that the RSMM could model not only the inter-arrival times and the durations of the cyberattacks but also the different types of attacks taken into account.

V. CONCLUSION

We developed a regime-switching hidden Markov model for the occurrences of cyberattacks specifically designed to value cyber-insurance contracts. Our model showed greater flexibility than the usual homogenous Markov chain in dealing with the transition probabilities' variations. Thus, our framework provides a more accurate depiction of cyber-attack data. Compared to the typical discrete HMM, our new modelling setup captures the correlation of the previous and current CKC states.

Within our proposed model setting, we demonstrated the price calculations of cyber-security insurance by simulation under the standard-deviation and exponential-premium principles. The discount factors were generated using the discretised version of the Vasiček interest rate model and the breach sizes of cyberattacks or severities are modelled by a DTP distribution. We considered the proportionality and functional-form assumptions to transform severities to dollar-amount losses. The EM-based model parameters get updated immediately whenever new information is available with each filtering-algorithm step, where the data-filtering window could be readily adjusted by the user depending on their purpose. The premiums obtained via the functional-form assumption are much smaller than those obtained via the proportionality assumption. We found that the log-log model in the functional-form assumption should be modified for a large amount of severities. Moreover, increasing the number of institutions aids in decreasing the SEs of premiums. We also examined how the premiums vary with the changes in the risk-averse parameters of the two premium principles for three different policy terms. The paired $t-$ and $F-$ tests signified that the means of premiums are more sensitive to changes in κ than SEs. Changes in the means and SEs are term-independent when the risk-averse parameters change under the proportionality assumption. For example, if we shrink κ, the premium changes are significant in all three policy terms.

In addition, we developed a semi-parametric approximation to the total losses in the absence of a closed-form solution. We established that finite-component normal mixture models could provide a good fit. Considering that the frequency of cyberattacks is low in practice, we implemented the regime-switching Markov model on a data set with fewer cyberattacks. Nonetheless, we included a case study with higher frequencies of cyberattacks to evidently demonstrate that the estimation of model parameters could be stabilised through the availability of a data set covering longer time intervals supported by the availability of high-powered computing resources. We also showed that our RSMM is a better model in capturing the cyberattack occurrences than the ACD models in terms of AICs and BICs.

Further research is warranted in assessing the performance of our proposed model and estimation approach using reliable data sets from the industrial and business sectors. The analysis of our modelling, filtering and pricing framework in supporting cyber insurance products with features customised to the needs of the clients is also an equally important pursuit; an example would be addressing a cyber threat that has a potential loss arising from cyber-related business interruptions, even when the cyber events originate from third-party IT service providers [18].

APPENDIX A
PROOF OF A MEASURE-CHANGE RELATED RESULT

Lemma 1: Under $\overline{P}$, $\{\mathbf{y}_k\}$, $k \in \mathbb{N}$, is a sequence of IID random variables. Each $\mathbf{y}_k$ is distributed as uniform with probability $\frac{1}{m}$ assigned to each vector $\mathbf{f}_i$, $1 \le i \le m$ in its range space.

Proof: Define λ_l and Λ_k by

$$\lambda_l = \prod_{i=1}^{m} \left(\frac{1}{m c_l^{(i)}} \right)^{y_l^{(i)}}, \quad \Lambda_k = \prod_{l=1}^{k} \lambda_l, \quad k \ge 1, \quad \Lambda_0 = 1. \tag{34}$$

The expression for Λ_k in (34) is the Radon-Nikodŷm derivative of $\overline{P}$ with respect to P, also written as $\left.\frac{\mathrm{d}\overline{P}}{\mathrm{d}P}\right|_{\mathcal{F}_k} = \Lambda_k$.

First, we note that

$$\begin{aligned} E\left[\lambda_{k+1} | \mathcal{F}_k\right] &= E\left[\prod_{i=1}^{m} \left(\frac{1}{m c_{k+1}^{(i)}} \right)^{y_{k+1}^{(i)}} \middle| \mathcal{F}_k \right] \\ &= E\left[\sum_{i=1}^{m} \frac{y_{k+1}^{(i)}}{m c_{k+1}^{(i)}} \middle| \mathcal{F}_k \right] \\ &= \frac{1}{m} \sum_{i=1}^{m} \frac{P\left(y_{k+1}^{(i)} = 1 | \mathcal{F}_k\right)}{P\left(\mathbf{y}_{k+1} = \mathbf{f}_i | \mathbf{y}_k, \mathbf{z}_k\right)} = 1, \end{aligned} \tag{35}$$

where the second equality holds since $y_{k+1}^{(i)}$ can only take the value 0 or 1. By the Bayes' theorem and (35),

$$\begin{aligned} &\overline{P}\left(\mathbf{y}_{k+1} = \mathbf{f}_i | \mathcal{F}_k\right) \\ &\quad = \overline{E}\left[\langle \mathbf{y}_{k+1}, \mathbf{f}_i \rangle | \mathcal{F}_k\right] = \frac{E\left[\Lambda_{k+1} \langle \mathbf{y}_{k+1}, \mathbf{f}_i \rangle | \mathcal{F}_k\right]}{E\left[\Lambda_{k+1} | \mathcal{F}_k\right]} \\ &\quad = \frac{\Lambda_k E\left[\lambda_{k+1} \langle \mathbf{y}_{k+1}, \mathbf{f}_i \rangle | \mathcal{F}_k\right]}{\Lambda_k E\left[\lambda_{k+1} | \mathcal{F}_k\right]} \\ &\quad = E\left[\prod_{j=1}^{m} \left(\frac{1}{m c_{k+1}^{(j)}} \right)^{y_{k+1}^{(j)}} \langle \mathbf{y}_{k+1}, \mathbf{f}_i \rangle \middle| \mathcal{F}_k \right] \\ &\quad = \frac{1}{m c_{k+1}^{(i)}} P\left(y_{k+1}^{(i)} = 1 | \mathcal{F}_k\right) = \frac{1}{m}. \end{aligned}$$

As a consequence, $\overline{P}\left(\mathbf{y}_{k+1} = \mathbf{f}_i\right) = \frac{1}{m}$ which is independent of the filtration $\mathcal{F}_k$.

APPENDIX B
PROOFS OF EQUATIONS (17) AND (18)

Proof of (17): From equations (6) and (12), we get

$$
\begin{aligned}
&\gamma\left(\mathcal{T}_k^{s,r}(\mathbf{y}_k,\mathbf{f}_i)\mathbf{z}_k\right)\\
&= \overline{E}[\Lambda_k \mathcal{T}_k^{s,r}(\mathbf{y}_k,\mathbf{f}_i)\mathbf{z}_k|\mathcal{F}_k^{\mathbf{y}}]\\
&= \overline{E}\Big[\Lambda_{k-1}\lambda_k\left(\mathcal{T}_{k-1}^{s,r}(\mathbf{y}_{k-1},\mathbf{f}_i)\right.\\
&\quad \left. +\ \langle \mathbf{z}_{k-1},\mathbf{e}_r\rangle\langle\mathbf{y}_k,\mathbf{f}_s\rangle\langle\mathbf{y}_{k-1},\mathbf{f}_i\rangle\right)\mathbf{z}_k\Big|\mathcal{F}_k^{\mathbf{y}}\Big]\\
&= \sum_{j=1}^{n}\overline{E}\left[\Lambda_{k-1}\mathcal{T}_{k-1}^{s,r}(\mathbf{y}_{k-1},\mathbf{f}_i)\langle\mathbf{z}_{k-1},\mathbf{e}_j\rangle\prod_{i=1}^{m}\left(mc_k^{(i)}\right)^{y_k^{(i)}}\Big|\mathcal{F}_k^{\mathbf{y}}\right]\boldsymbol{\pi}_j\\
&\quad +\overline{E}\left[\Lambda_{k-1}\langle\mathbf{z}_{k-1},\mathbf{e}_r\rangle\prod_{i=1}^{m}\left(mc_k^{(i)}\right)^{y_k^{(i)}}\Big|\mathcal{F}_k^{\mathbf{y}}\right]\\
&\quad \times\langle\mathbf{y}_k,\mathbf{f}_s\rangle\langle\mathbf{y}_{k-1},\mathbf{f}_i\rangle\boldsymbol{\pi}_r\\
&= \sum_{j=1}^{n}\left\langle\gamma\left(\mathcal{T}_{k-1}^{s,r}(\mathbf{y}_{l-1},\mathbf{f}_i)\mathbf{z}_{k-1}\right),\mathbf{e}_j\right\rangle d_k^{(j)}\boldsymbol{\pi}_j\\
&\quad + m\langle\mathbf{p}_{k-1},\mathbf{e}_r\rangle\langle\mathbf{y}_k,\mathbf{f}_s\rangle\langle\mathbf{y}_{k-1},\mathbf{f}_i\rangle c_{sr}(\mathbf{f}_i)\boldsymbol{\pi}_r\\
&= \Pi\ \mathrm{diag}(\mathbf{d}_k)\gamma\left(\mathcal{T}_{k-1}^{s,j}(\mathbf{y}_l,\mathbf{f}_i)\mathbf{z}_{k-1}\right)\\
&\quad + m\langle\mathbf{p}_{k-1},\mathbf{e}_r\rangle\langle\mathbf{y}_k,\mathbf{f}_s\rangle\langle\mathbf{y}_{k-1},\mathbf{f}_i\rangle c_{sr}(\mathbf{f}_i)\boldsymbol{\pi}_r.
\end{aligned}
$$

The justification of the result in (18) follows similar reasoning as above.

APPENDIX C
PROOF OF (20)

The idea of the proof is similar to that in Section 2.7 of [28]. To perform the measure change from $c_{sr}(\mathbf{f}_i)$ to $\widehat{c}_{sr}(\mathbf{f}_i)$, we define a new measure $P^{\widehat{c}_{sr}(\mathbf{f}_i)}$ via $\frac{\mathrm{d}P^{\widehat{c}_{sr}(\mathbf{f}_i)}}{\mathrm{d}P^{c_{sr}(\mathbf{f}_i)}}\Big|_{\mathcal{F}_k} = \Lambda_k^* = \prod_{l=1}^{k}\lambda_l^*$, where

$$
\lambda_l^* = \sum_{r=1}^{n}\sum_{s=1}^{m}\left(\frac{\widehat{c}_{sr}(\mathbf{f}_i)}{c_{sr}(\mathbf{f}_i)}\right)^{\langle\mathbf{z}_{k-1},\mathbf{e}_r\rangle\langle\mathbf{y}_k,\mathbf{f}_s\rangle\langle\mathbf{y}_{k-1},\mathbf{f}_i\rangle}.
$$

So,

$$
\begin{aligned}
\log\frac{\mathrm{d}P^{\widehat{c}_{sr}(\mathbf{f}_i)}}{\mathrm{d}P^{c_{sr}(\mathbf{f}_i)}} &= \sum_{l=1}^{k}\sum_{r=1}^{n}\sum_{s=1}^{m}\left[\log\left(\widehat{c}_{sr}(\mathbf{f}_i)\right)-\log\left(c_{sr}(\mathbf{f}_i)\right)\right]\\
&\quad\times\langle\mathbf{z}_{k-1},\mathbf{e}_r\rangle\langle\mathbf{y}_k,\mathbf{f}_s\rangle\langle\mathbf{y}_{k-1},\mathbf{f}_i\rangle\\
&= \sum_{r=1}^{n}\sum_{s=1}^{m}\mathcal{T}_k^{s,r}(\mathbf{y}_k,\mathbf{f}_i)\log\left(\widehat{c}_{sr}(\mathbf{f}_i)\right)+R, \quad (36)
\end{aligned}
$$

where R does not contain $\widehat{c}_{sr}(\mathbf{f}_i)$. Observe that $\sum_{s=1}^{m}\mathcal{T}_k^{s,r}(\mathbf{y}_k,\mathbf{f}_i) = \mathcal{T}_k^{r}(\mathbf{f}_i)$; hence,

$$
\sum_{s=1}^{m}\widehat{\mathcal{T}}_k^{s,r}(\mathbf{y}_k,\mathbf{f}_i) = \widehat{\mathcal{T}}_k^{r}(\mathbf{f}_i). \quad (37)
$$

The $\widehat{c}_{sr}(\mathbf{f}_i)$'s optimal estimate is the value that maximises the log-likelihood (36) subject to the constraint $\sum_{r=1}^{n}\sum_{s=1}^{m}\widehat{c}_{sr}(\mathbf{f}_i) = 1$.

Constructing the function $\mathcal{L}(\widehat{c}_{sr}(\mathbf{f}_i),\beta)$ involving the Lagrange multiplier β, we have

$$
\begin{aligned}
\mathcal{L}(\widehat{c}_{sr}(\mathbf{f}_i),\beta) &= \sum_{r=1}^{n}\sum_{s=1}^{m}\widehat{\mathcal{T}}_k^{s,r}(\mathbf{y}_k,\mathbf{f}_i)\log\left(\widehat{c}_{sr}(\mathbf{f}_i)\right)\\
&\quad+\beta\left(\sum_{s=1}^{m}\widehat{c}_{sr}(\mathbf{f}_i)-1\right)+R. \quad (38)
\end{aligned}
$$

Differentiating (38) with respect to $\widehat{c}_{sr}(\mathbf{f}_i)$ and β and then equating the derivatives to 0, we get

$$
\frac{1}{\widehat{c}_{sr}(\mathbf{f}_i)}\widehat{\mathcal{T}}_k^{s,r}(\mathbf{y}_k,\mathbf{f}_i)+\beta = 0 \quad (39)
$$

and

$$
\sum_{s=1}^{m}\widehat{c}_{sr}(\mathbf{f}_i) = 1. \quad (40)
$$

Equation (39) yields

$$
\widehat{c}_{sr}(\mathbf{f}_i) = -\frac{\widehat{\mathcal{T}}_k^{s,r}(\mathbf{y}_k,\mathbf{f}_i)}{\beta}. \quad (41)
$$

Summing (41) over s and applying (37), and (40), we have

$$
1 = -\frac{\widehat{\mathcal{T}}_k^{r}(\mathbf{f}_i)}{\beta}. \quad (42)
$$

Combining (39) and (42) leads to

$$
\widehat{c}_{sr}(\mathbf{f}_i) = \frac{\widehat{\mathcal{T}}_k^{s,r}(\mathbf{y}_k,\mathbf{f}_i)}{\widehat{\mathcal{T}}_k^{r}(\mathbf{f}_i)} = \frac{\gamma\left(\mathcal{T}_k^{s,r}(\mathbf{y}_k,\mathbf{f}_i)\right)}{\gamma\left(\mathcal{T}_k^{r}(\mathbf{f}_i)\right)},
$$

which is in agreement with equation (20).

APPENDIX D
JUSTIFICATION OF EQUATION (22)

We write the log-likelihood of c_{sr} as

$$
\mathcal{L}(c_{sr}) = \sum_{l=1}^{k}\left(\log(c_{sr})\right)\langle\mathbf{z}_{l-1},\mathbf{e}_r\rangle\langle\mathbf{y}_l,\mathbf{f}_s\rangle\langle\mathbf{y}_{l-1},\mathbf{f}_i\rangle.
$$

Thus, the Fisher information of c_{sr} is

$$
\mathcal{I}(c_{sr}) = -E\left[\frac{\mathrm{d}^2}{\mathrm{d}c_{sr}^2}\mathcal{L}(c_{sr})\Big|c_{sr}\right] = \frac{\widehat{\mathcal{T}}_k^{s,r}}{c_{sr}^2}.
$$

APPENDIX E
CALCULATION OF THE INTEGRAL IN EQUATION (29)

We first calculate $I_r := \int_0^T e^{x^2}\mathrm{d}x$ by considering $I_r^2 = \int_0^T\int_0^T e^{x^2+y^2}\mathrm{d}x\mathrm{d}y$. Polar integration is used with $x^2+y^2 = r^2$ and $\mathrm{d}x\mathrm{d}y = r\mathrm{d}r\mathrm{d}\theta$. Note that the symbols x, y, r, and θ are only utilised in this appendix and they have no relationship with those previously defined. Consequently,

$$
\begin{aligned}
I_r^2 &= \int_0^{\frac{\pi}{2}}\int_0^T re^{r^2}\mathrm{d}r\mathrm{d}\theta = \frac{1}{2}\int_0^{\frac{\pi}{2}}\left(e^{T^2}-1\right)\mathrm{d}\theta\\
&= \frac{\pi}{4}\left(e^{T^2}-1\right).
\end{aligned}
$$

The integral part of (29) becomes

$$\int_0^T \exp\left(-\mu_r k + \frac{1}{2}\sigma_r^2 k^2\right) \mathrm{d}k = \int_0^T \exp\left[-\frac{\mu_r^2}{2\sigma_r^2} + \frac{1}{2}\sigma_r^2\left(k - \frac{\mu_r}{\sigma_r^2}\right)^2\right] \mathrm{d}k.$$

Finally, letting $x^2 = \frac{1}{2}\sigma_r^2\left(k - \frac{\mu_r}{\sigma_r^2}\right)^2$ we obtain

$$\begin{aligned}\int_0^T \exp\left(-\mu_r k + \frac{1}{2}\sigma_r^2 k^2\right) \mathrm{d}k &= \exp\left(-\frac{\mu_r^2}{2\sigma_r^2}\right)\int_0^T e^{x^2}\frac{\sqrt{2}}{\sigma_r}\mathrm{d}x \\ &= \frac{\sqrt{2}}{\sigma_r}\exp\left(-\frac{\mu_r^2}{2\sigma_r^2}\right) I_r \\ &= \frac{\sqrt{\pi\left(e^{T^2}-1\right)}}{\sqrt{2}\sigma_r}\exp\left(-\frac{\mu_r^2}{2\sigma_r^2}\right).\end{aligned}$$

APPENDIX F
ACD MODELS

The ACD model was originally proposed to describe the evolution of the inter-arrival time, or duration between stock transactions [30]. Suppose the incidents happen at $t_1, t_2, \ldots, t_N$, where N is the number of incidents. Let $t_0 = 0$. The event duration is defined as $\zeta_i := t_i - t_{i-1}$, for $i = 1, 2, \ldots, N$. The basic idea of the conditional mean model is to standardise the durations by leveraging the historical information. That is,

$$\zeta_i = \Psi_i \epsilon_i,$$

where Ψ_i's are functions of the historical durations and represented by the historical information up to time t_{i-1}, i.e.,

$$\Psi_i = E\left[\zeta_i | \mathcal{F}_{i-1}^{\zeta}\right].$$

The ϵ_i's are IID errors with $E[\epsilon_i] = 1$. Below are the expressions for the three ACD models.

- Standard ACD model (ACD) [30]:

$$\Psi_i = \varepsilon + \sum_{j=1}^{q_1}\phi_j\zeta_{i-j} + \sum_{j=1}^{q_2}\varphi_j\Psi_{i-j},$$

  where $\varepsilon, \phi_j, \varphi_j \geq 0$, and q_1 and q_2 are positive integers for the possible value of the order of the autoregressive terms.
- Type-I log-ACD model (LACD1) [6]:

$$\log(\Psi_i) = \varepsilon + \sum_{j=1}^{q_1}\phi_j\log(\epsilon_{i-j}) + \sum_{j=1}^{q_2}\varphi_j\log(\Psi_{i-j}).$$

- Type-II log-ACD model (LACD2) [6]:

$$\log(\Psi_i) = \varepsilon + \sum_{j=1}^{q_1}\phi_j log(\zeta_{i-j}) + \sum_{j=1}^{q_2}\varphi_j\log(\Psi_{i-j}).$$

We let $q_1 = q_2 = 1$ as in Xu et al. [69]. The distribution of the standardised errors of ϵ_i's could be chosen from the generalised Gamma, Weibull, exponential, Burr, generalised F and q-Weibull distributions embedded in the R package "ACDm" [7]. In our case, we select exponential distribution with minimum mean squared errors.

REFERENCES

[1] D. Aiylam, "Parameter estimation in HMMs with guaranteed convergence," Ph.D. dissertation, Dept. EECS, MIT, Cambridge, MA, USA, 2018.
[2] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974. Accessed: Dec. 12, 2020, doi: 10.1109/TAC.1974.1100705.
[3] A. M. Algarni and Y. K. Malaiya, "A consolidated approach for estimation of data security breach costs," in *Proc. ICIM-Ei*, London, U.K., May 2016, pp. 26–39, doi: 10.1109/INFOMAN.2016.7477530.
[4] M. A. R. A. Amin, S. Shetty, L. Njilla, D. K. Tosh, and C. Kamhoua, "Hidden Markov model and cyber deception for the prevention of adversarial lateral movement," *IEEE Access*, vol. 9, pp. 49662–49682, 2021. Accessed: Feb. 28, 2023, doi: 10.1109/ACCESS.2021.3069105.
[5] Y. Antonio and S. W. Indratno, "Cyber insurance rate making based on Markov model for regular networks topology," *J. Phys., Conf. Ser.*, vol. 1752, no. 1, Feb. 2021, Art. no. 012002. Accessed: Jan. 4, 2022, doi: 10.1088/1742-6596/1752/1/012002.
[6] L. Bauwens and P. Giot, "The logarithmic ACD model: An application to the bid-ask quote process of three NYSE stocks," *Annales d'Économie Statistique*, vol. 60, no. 60, pp. 117–149, Oct./Dec. 2000. Accessed: Jan 6, 2022, doi: 10.2307/20076257.
[7] M. Belfrage. (Jul. 16, 2015). *ACDm: Tools for Autoregressive Conditional Duration Models, R Package Version 1.0.4.1.* [Online]. Available: https://cran.r-project.org/web/packages/ACDm/
[8] Y. Bessy-Roland, A. Boumezoued, and C. Hillairet, "Multivariate Hawkes process for cyber insurance," *Ann. Actuarial Sci.*, vol. 15, no. 1, pp. 14–39, Mar. 2021. Accessed: Jan. 4, 2022, doi: 10.1017/S1748499520000093.
[9] R. Betterley, "Cyber/privacy insurance market survey: A tough market for larger insureds, but smaller insureds finding eager insurers," Betterley Risk Consultants, Friendship, ME, USA, Betterley Rep., Jun. 2016. [Online]. Available: http://betterley.com/samples/cpims16_nt.pdf
[10] C. Biener, M. Eling, and J. H. Wirfs, "Insurability of cyber risk: An empirical analysis," *Geneva Papers Risk Insurance Issues Pract.*, vol. 40, no. 1, pp. 131–158, Jan. 2015. Accessed: Jan. 4, 2022, doi: 10.1057/gpp.2014.19.
[11] R. Böhme and G. Kataria, "Models and measures for correlation in cyber-insurance," presented at the *5th Workshop on the Economics of Information Security*, Cambridge, U.K., Jun. 26-28, 2006.
[12] R. Böhme and G. Schwartz, "Modeling cyber-insurance: Towards a unifying framework," Presented at the 9th Workshop Econ. Inf. Secur., Cambridge, MA, USA, Jun. 2010.
[13] H. Borchers. (Sep. 22, 2022). *Pracma: Practical Numerical Math Functions, R Package Version 2.4.2.* [Online]. Available: https://cran.r-project.org/web/packages/pracma/index.html
[14] O. Cappé, "Online EM algorithm for hidden Markov models," *J. Comput. Graph. Statist.*, vol. 20, no. 3, pp. 728–749, Jan. 2011. Accessed: Feb. 28, 2023, doi: 10.1198/jcgs.2011.09109.
[15] E. Cator, R. van de Bovenkamp, and P. Van Mieghem, "Susceptible-infected-susceptible epidemics on networks with general infection and cure times," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 87, no. 6, Jun. 2013, Art. no. 062816. Accessed: Dec. 17, 2020, doi: 10.1103/PhysRevE.87.062816.
[16] T. Chadza, K. G. Kyriakopoulos, and S. Lambotharan, "Analysis of hidden Markov model learning algorithms for the detection and prediction of multi-stage network attacks," *Future Gener. Comput. Syst.*, vol. 108, pp. 636–649, Jul. 2020. Accessed Feb. 28, 2023, doi: 10.1016/j.future.2020.03.014.
[17] I. M. Chakravarti, R. G. Laha, and J. Roy, "Planning of surveys and experiments," in *Handbook of Methods of Applied Statistics*, vol. 2. New York, NY, USA: Wiley, 1967, pp. 392–394.
[18] S. Chambers, "New cyber insurance product launches," Asia Shipping Media, Singapore, Tech. Rep., Apr. 2020. Accessed: Aug. 21, 2020. [Online]. Available: https://splash247.com/new-cyber-insurance-product-launches/

[19] P. Date, R. Mamon, and A. Tenyakov, "Filtering and forecasting commodity futures prices under an HMM framework," *Energy Econ.*, vol. 40, pp. 1001–1013, Nov. 2013. Accessed: Jan. 27, 2023, doi: 10.1016/j.eneco.2013.05.016.

[20] M. Denuit, "The exponential premium calculation principle revisited," *ASTIN Bull.*, vol. 29, no. 2, pp. 215–226, Nov. 1999. Accessed: Jan. 4, 2022, doi: 10.2143/AST.29.2.504612.

[21] S. Dionisi, "Determining the likelihood of a cybersecurity failure for use in cybersecurity," in *Cybersecurity: Impact on Insurance Business and Operations* (The Joint Risk Management Section of the SOA, the CAS and the CIA). Schaumburg, IL, USA: The SOA (Society of Actuaries), 2017, pp. 16–19. Accessed: Jan. 21, 2020. [Online]. Available: https://www.soa.org/globalassets/assets/files/static-pages/sections/joint-risk-mgmt/cyber-security-impact.pdf

[22] B. Edwards, S. Hofmeyr, and S. Forrest, "Hype and heavy tails: A closer look at data breaches," *J. Cybersecur.*, vol. 2, no. 1, pp. 3–14, Dec. 2016. Accessed: Jan. 22, 2021, doi: 10.1093/cybsec/tyw003.

[23] *Efficient, Powerful Data Breach Coverages*, CNA, Chicago, IL, USA. Accessed: Aug. 25, 2020. [Online]. Available: https://www.cna.com/web/wcm/connect/a6b2e08a-cfb4-494f-be31-c6b83f97be8e/SmallBizSSDataBreachInsured_CNA.pdf?MOD=AJPERES

[24] M. Eling and W. Schnell, "What do we know about cyber risk and cyber risk insurance?" *J. Risk Finance*, vol. 17, no. 5, pp. 474–491, Nov. 2016. Accessed: Dec. 29, 2020, doi: 10.1108/JRF-09-2016-0122.

[25] M. Eling and K. Jung, "Copula approaches for modeling cross-sectional dependence of data breach losses," *Insurance, Math. Econ.*, vol. 82, pp. 167–180, Sep. 2018. Accessed: Oct. 21, 2020, doi: 10.1016/j.insmatheco.2018.07.003.

[26] M. Eling and J. Wirfs, "What are the actual costs of cyber risk events?" *Eur. J. Oper. Res.*, vol. 272, no. 3, pp. 1109–1119, Feb. 2019. Accessed: Feb. 21, 2021, doi: 10.1016/j.ejor.2018.07.021.

[27] R. J. Elliott and V. Krishnamurthy, "New finite-dimensional filters for parameter estimation of discrete-time linear Gaussian models," *IEEE Trans. Autom. Control*, vol. 44, no. 5, pp. 938–951, May 1999. Accessed: Feb. 15, 2020, doi: 10.1109/9.763210.

[28] R. J. Elliott, L. Aggoun, and J. B. Moore, "Discrete states and discrete observations," in *Hidden Markov Models: Estimation and Control*, vol. 29, 2nd ed. New York, NY, USA: Springer, 2008, pp. 15–54.

[29] R. J. Elliott and C. B. Hyndman, "Parameter estimation in commodity markets: A filtering approach," *J. Econ. Dyn. Control*, vol. 31, no. 7, pp. 2350–2373, Jul. 2007. Accessed: Mar. 1, 2023, doi: 10.1016/j.jedc.2006.07.005.

[30] R. F. Engle and R. R. Jeffrey, "Autoregressive conditional duration: A new model for irregularly spaced transaction data," *Econometrica*, vol. 66, no. 5, pp. 1127–1162, Sep. 1998. Accessed: Jan. 11, 2022, doi: 10.2307/2999632.

[31] M. A. Fahrenwaldt, S. Weber, and K. Weske, "Pricing of cyber insurance contracts in a network model," *ASTIN Bull.*, vol. 48, no. 3, pp. 1175–1218, Sep. 2018. Accessed: Jan. 11, 2022, doi: 10.1017/asb.2018.23.

[32] S. Farkas, O. Lopez, and M. Thomas, "Cyber claim analysis using generalized Pareto regression trees with applications to insurance," *Insurance, Math. Econ.*, vol. 98, pp. 92–105, May 2021. Accessed: Jan. 11, 2022, doi: 10.1016/j.insmatheco.2021.02.009.

[33] Z. Fang, M. Xu, S. Xu, and T. Hu, "A framework for predicting data breach risk: Leveraging dependence to cope with sparsity," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 2186–2201, 2021. Accessed: Jan. 10, 2022, doi: 10.1109/TIFS.2021.3051804.

[34] B. Gao, G. Hu, Y. Zhong, and X. Zhu, "Cubature Kalman filter with both adaptability and robustness for tightly-coupled GNSS/INS integration," *IEEE Sensors J.*, vol. 21, no. 13, pp. 14997–15011, Jul. 2021. Accessed: Feb. 20, 2023, doi: 10.1109/JSEN.2021.3073963.

[35] B. Gao, S. Gao, G. Hu, Y. Zhong, and C. Gu, "Maximum likelihood principle and moving horizon estimation based adaptive unscented Kalman filter," *Aerosp. Sci. Technol.*, vol. 73, pp. 184–196, Feb. 2018. Accessed: Feb. 20, 2023, doi: 10.1016/j.ast.2017.12.007.

[36] B. Gao, S. Gao, Y. Zhong, G. Hu, and C. Gu, "Interacting multiple model estimation-based adaptive robust unscented Kalman filter," *Int. J. Control, Autom. Syst.*, vol. 15, no. 5, pp. 2013–2025, Oct. 2017. Accessed: Feb. 20, 2023, doi: 10.1007/s12555-016-0589-2.

[37] X. Gu, R. Mamon, M. Davison, and H. Yu, "An automated financial indices-processing scheme for classifying market liquidity regimes," *Int. J. Control*, vol. 94, no. 3, pp. 735–756, Mar. 2021. Accessed: Mar. 11, 2020, doi: 10.1080/00207179.2019.1616225.

[38] M. Husák, J. Komárková, E. Bou-Harb, and P. Čeleda, "Survey of attack projection, prediction, and forecasting in cyber security," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 640–660, 1st Quart., 2019. Accessed: Feb. 27, 2023, doi: 10.1109/COMST.2018.2871866.

[39] *Cost of a Data Breach Report 2021*, IBM Secur., Armonk, NY, USA, Jul. 2021. Accessed: Sep. 13, 2021. [Online]. Available: https://www.ibm.com/security/data-breach

[40] J. Jacobs. (Dec. 11, 2014). Analyzing Ponemon cost of data breach. Data Driven Security. Accessed: Sep. 13, 2021. [Online]. Available: http://datadrivensecurity.info/blog/posts/2014/Dec/ponemon/

[41] P. Jevtić and N. Lanchier, "Dynamic structural percolation model of loss distribution for cyber risk of small and medium-sized enterprises for tree-based LAN topology," *Insurance, Math. Econ.*, vol. 91, pp. 209–223, Mar. 2020. Accessed: Jan. 8, 2022, doi: 10.1016/j.insmatheco.2020.02.005.

[42] K. Jung, "Extreme data breach losses: An alternative approach to estimating probable maximum loss for data breach risk," *North Amer. Actuarial J.*, vol. 25, no. 4, pp. 580–603, Oct. 2021. Accessed: Jan. 8, 2022, doi: 10.1080/10920277.2021.1919145.

[43] C. Kleiber and S. Kotz, "Pareto distributions," in *Statistical Size Distributions in Economics and Actuarial Sciences*, 1st ed. Hoboken, NJ, USA: Wiley, 2003, pp. 59–106.

[44] V. Krishnamurthy and J. B. Moore, "On-line estimation of hidden Markov model parameters based on the Kullback–Leibler information measure," *IEEE Trans. Signal Process.*, vol. 41, no. 8, pp. 2557–2573, Feb. 1993. Accessed: Feb. 28, 2023, doi: 10.1109/78.229888.

[45] S. Li and Y. Bai, "Deep learning and improved HMM training algorithm and its analysis in facial expression recognition of sports athletes," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–12, Jan. 2022. Accessed: Feb. 28, 2023, doi: 10.1155/2022/1027735.

[46] C. Liu, H.-C. Li, K. Fu, F. Zhang, M. Datcu, and W. J. Emery, "Bayesian estimation of generalized gamma mixture model based on variational EM algorithm," *Pattern Recognit.*, vol. 87, pp. 269–284, Mar. 2019, doi: 10.1016/j.patcog.2018.10.025.

[47] R. S. Mamon, "Three ways to solve for bond prices in the Vasicek model," *J. Appl. Math. Decis. Sci.*, vol. 8, no. 1, pp. 1–14, Jan. 2004. Accessed: Nov. 23, 2020. [Online]. Available: https://downloads.hindawi.com/archive/2004/131526.pdf

[48] R. S. Mamon, C. Erlwein, and R. B. Gopaluni, "Adaptive signal processing of asset price dynamics with predictability analysis," *Inf. Sci.*, vol. 178, no. 1, pp. 203–219, Jan. 2008. Accessed: Mar. 11, 2020, doi: 10.1016/j.ins.2007.05.021.

[49] T. Mansouri, M. Sadeghimoghadam, and I. G. Sahebi, "A new algorithm for hidden Markov models learning problem," Feb. 2021, *arXiv:2102.07112*. Accessed Feb. 28, 2023

[50] G. J. McLachlan, S. X. Lee, and S. I. Rathnayake, "Finite mixture models," *Annu. Rev. Statist. Appl.*, vol. 6, no. 1, pp. 355–378, Mar. 2019. Accessed: Jan. 13, 2022, doi: 10.1146/annurev-statistics-031017-100325.

[51] A. Mukhopadhyay, S. Chatterjee, D. Saha, A. Mahanti, and S. K. Sadhukhan, "E-risk management with insurance: A framework using copula aided Bayesian belief networks," in *Proc. 39th Annu. Hawaii Int. Conf. Syst. Sci. (HICSS)*, Kauia, HI, USA, Jan. 2006, p. 126. Accessed: Jan. 11, 2022, doi: 10.1109/HICSS.2006.138.

[52] National Protection and Programs Directorate, U.S. Department of Homeland Security, and Cybersecurity and Infrastructure Security Agency, Arlington, VA, USA. (Nov. 2012). *Cybersecurity Insurance Workshop Readout Report*. Accessed: Nov. 29, 2020. [Online]. Available: https://www.cisa.gov/sites/default/files/publications/November%202012%20Cybersecurity%20Insurance%20Workshop.pdf

[53] B.-J. Park and D. Lord, "Application of finite mixture models for vehicle crash data analysis," *Accident Anal. Prevention*, vol. 41, no. 4, pp. 683–691, Jul. 2009. Accessed: Jan. 13, 2022, doi: 10.1016/j.aap.2009.03.007.

[54] Privacy Rights Clearinghouse. *Data Breaches*. Accessed: Jan. 11, 2022. [Online]. Available: https://privacyrights.org/data-breaches

[55] S. Romanosky, L. Ablon, A. Kuehn, and T. Jones, "Content analysis of cyber insurance policies: How do carriers write policies and price cyber risk?" *J. Cybersecur.*, vol. 5, no. 1, Feb. 2019, Art. no. tyz002. Accessed: Jan. 13, 2022, doi: 10.2139/ssrn.2929137.

[56] T. Rydén, "On recursive estimation for hidden Markov models," *Stochastic Process. Appl.*, vol. 66, no. 1, pp. 79–96, Feb. 1997. Accessed: Feb. 28, 2023., doi: 10.1016/S0304-4149(96)00114-7.

[57] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, Mar. 1978.

[58] H. L. Seal, "Survival probabilities based on Pareto claim distributions," *ASTIN Bull.*, vol. 11, no. 1, pp. 61–71, Jun. 1980. Accessed: Aug. 13, 2021, doi: 10.1017/S0515036100006620.

[59] M. A. Stephens, "EDF statistics for goodness of fit and some comparisons," *J. Amer. Stat. Assoc.*, vol. 69, no. 347, pp. 730–737, Sep. 1974. Accessed: Jan. 13, 2022, doi: 10.1080/01621459.1974.10480196.

[60] H. Sun, M. Xu, and P. Zhao, "Modeling malicious hacking data breach risks," *North Amer. Actuarial J.*, vol. 25, no. 4, pp. 484–502, Oct. 2021. Accessed: Dec. 13, 2021, doi: 10.1080/10920277.2020.1752255.

[61] W. Rabb, "State of the cyber insurance market—Top trends, insurers and challenges: A.M. Best," *Insurance J.*, Jun. 2019. Accessed: Nov. 13, 2020. [Online]. Available: https://www.insurancejournal.com/news/national/2019/06/18/529747.htm

[62] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989. Accessed: Feb. 16, 2023, doi: 10.1109/5.18626.

[63] R. J. Tibshirani and B. Efron, "Confidence intervals based on bootstrap 'tables,'" *An Introduction to the Bootstrap*, 1st ed. Boca Raton, FL, USA: CRC Press, 1993, pp. 153–167.

[64] Y. Wang, "On fast computation of the non-parametric maximum likelihood estimate of a mixing distribution," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 69, no. 2, pp. 185–198, Apr. 2007. Accessed: Jan. 11, 2022, doi: 10.1111/j.1467-9868.2007.00583.x.

[65] Y. Wang. (Mar. 5, 2017). *nspmix: Nonparametric and Semiparametric Mixture Estimation, R Package Version 1.5-0*. [Online]. Available: https://cran.r-project.org/web/packages/nspmix/index.html

[66] Wikipedia. *WannaCry Ransomware Attack*. Accessed: Nov. 25, 2021. [Online]. Available: https://en.wikipedia.org/wiki/WannaCry_ransomware_attack

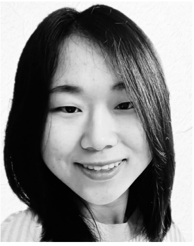
[67] S. Wheatley, T. Maillart, and D. Sornette, "The extreme risk of personal data breaches and the erosion of privacy," *Eur. Phys. J. B*, vol. 89, no. 1, pp. 1–12, Jan. 2016. Accessed: Jan. 11, 2020, doi: 10.1140/epjb/e2015-60754-4.

[68] M. Xu and L. Hua, "Cybersecurity insurance: Modeling and pricing," *North Amer. Actuarial J.*, vol. 23, no. 2, pp. 220–249, Apr. 2019. Accessed: Jan. 12, 2021, doi: 10.1080/10920277.2019.1566076.

[69] M. Xu, K. M. Schweitzer, R. M. Bateman, and S. Xu, "Modeling and predicting cyber hacking breaches," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 11, pp. 2856–2871, Nov. 2018, doi: 10.1109/TIFS.2018.2834227.

[70] M. Zhang Wu, J. Luo, X. Fang, M. Xu, and P. Zhao, "Modeling multivariate cyber risks: Deep learning dating extreme value theory," *J. Appl. Statist.*, vol. 50, no. 3, pp. 610–630, Feb. 2023. Accessed: Jan. 12, 2022, doi: 10.1080/02664763.2021.1936468.

[71] Y. Zhao and R. Mamon, "Annuity contract valuation under dependent risks," *Jpn. J. Ind. Appl. Math.*, vol. 37, no. 1, pp. 1–23, Jan. 2020, doi: 10.1007/s13160-019-00366-2.



YUYING LI received the B.S. degree in statistics from Sichuan University, Chengdu, China, in 2017, and the M.S. degree in actuarial science from The University of Western Ontario (UWO), London, Canada, in 2018, where she is currently pursuing the Ph.D. degree with the Department of Statistical and Actuarial Sciences (DSAS). Her research interests include statistical methods and stochastic approaches in cyber risk modeling.

ROGEMAR MAMON (Member, IEEE) is currently a Professor with DSAS, The University of Western Ontario (UWO). He is an elected fellow of the U.K.'s Institute of Mathematics and its Applications and a Chartered Scientist of the U.K.'s Science Council. His research interests include hidden Markov models and their estimation, including filtering, smoothing and prediction, and the applications of stochastic processes to financial and actuarial modeling and related areas.