Applied Stochastic Models in Business and Industry

WILEY

RESEARCH ARTICLE OPEN ACCESS

Pricing Cyber Insurance: A Geospatial Statistical Approach

L. V. Ballestra[1] | V. D'Amato[2] | P. Fersini[3] | S. Forte[4] | F. Greco[1]

[1]Department of Statistical Sciences, University of Bologna, Bologna, Italy | [2]Difarma Department, University of Salerno, Fisciano, Italy | [3]Department of Business and Management, Luiss Guido Carli University, Rome, Italy | [4]Faculty of Law, Giustino Fortunato University, Benevento, Italy

Correspondence: F. Greco (fedele.greco@unibo.it)

Received: 19 December 2023 | **Revised:** 26 July 2024 | **Accepted:** 6 August 2024

Keywords: Bayesian hierarchical models | cyber insurance | cyber risk | Gaussian Markov random fields | spatial correlation

ABSTRACT

Cyberspace is a dynamic ecosystem consisting of interconnected data, devices, and individuals, with multiple network layers comprising identifiable nodes. Location-based information can significantly improve cyber resilience decision-making and facilitate the development of innovative cyber risk pricing tools. This article is based on a methodology that uses company geospatial data to accurately estimate the number of expected losses arising from cyberattacks. Our approach aims to build and compare statistical spatial models that allow pricing cyber policies more effectively than traditional non-spatial methods by incorporating all available data. By accounting for spatial dependence, we can assess the risk of data breaches and contribute to the design of more efficient cyber risk policies for the insurance market.

1 | Introduction

Recent years have seen a significant increase in the frequency and impact of cyber incidents. According to a report by the European Systemic Risk Board [1], cyber risk now poses a systemic threat to the financial system, with potentially severe negative consequences for the real economy. The report cites industry estimates ranging from USD 45 billion to USD 654 billion for the global economy in 2018, highlighting the difficulty of accurately estimating the total cost of cyber incidents.

The National Institute of Standards and Technology (NIST) defines cyber risk as the "risk of financial loss, operational disruption, or damage resulting from the failure of digital technologies employed for informational and/or operational functions introduced to a manufacturing system via electronic means from unauthorized access, use, disclosure, disruption, modification, or destruction of the manufacturing system" [2]. Cyber risk can be classified as an operational risk, although it differs from more traditional sources of operational risk in several material ways.

The speed and scale of propagation, the potential for a major cyber incident to have a more widespread impact than many other shocks, the fact that it is not constrained by geographic boundaries, and the degree of disruption experienced by organizations all contribute to the specificity of cyber risk compared to operational risk. Companies and institutions can no longer ignore cyber threats. To protect business operations from both external and internal threats, cyber defense must be integrated into traditional security activities by aligning cybersecurity with strategic business activities. Organizations must quickly prioritize cyber threats to improve cyber resilience and quantify the impact of cyberattacks on business systems.

The European Systemic Risk Board [1] points out that cyber risk has the potential to trigger serious and systemic financial repercussions, highlighting that the materialization of cyber risk can trigger a systemic financial crisis. Thus, it is imperative to move away from the common approach of treating cyber risk as a purely information technology problem. Rather than relying on qualitative metrics and operational terms that treat cyber risk as solely an information technology problem, it is essential to quantify financial measures to inform investment decisions. Therefore, cyber risk must be viewed as a source of uncertainty that has a financial impact on the organization's business. This approach allows for a better understanding of the true significance of the risk as a critical part of enterprise risk management [3]. Carannante et al. [4] explore dependence among different cyber risk classes adopting vine copulas to capture dependence; Ruan [5] explains the need to establish data schemes such as International Digital Asset Classification (IDAC) and International Classification of Cyber Incidents (ICCI); Mukhopadhyay et al. [6] discuss cyber risk insurance products to minimize the impact of financial loss of security breach, while Aldasoro et al. [7] provide an interesting discussion concerning the drivers of cyber risk. For a comprehensive review on modeling and pricing of cyber insurance, the interested reader is referred to the work by Awiszus et al. [8].

This article proposes the spatial mapping of cyberattacks, leveraging the increasing volume and availability of location-based data to build statistical models that can improve the description and understanding of the complex cyberspace that includes layers of data and networks with strong interdependent structures.

The growing interest in geospatial data in the real economy stems from its ability to provide information about the location of objects, events, or phenomena, whether static or dynamic, throughout the world. This information can greatly enhance insight into the relationship between variables, revealing patterns and trends across all activities. The use of spatial modeling allows us to effectively quantify the impact and likelihood of risky scenarios in cyberspace, which can be used to design insurance policies that protect against cyber risk. In this regard, it is worth mentioning that Veerasamy, Moolla, and Dawood [9] have identified ten possible applications of geospatial data in cyber-security, namely, tracking, data analysis, visualization, situational awareness, cyber intelligence, collaboration, improved response to cyber threats, decision-making, cyber threat prioritization and protect cyber infrastructure.

To model the complex nature of cyber risk and its geographic patterns, we propose and compare five statistical models under a Bayesian inference paradigm. The simplest model assumes homogeneous risk, while the most complex model allows for both unstructured and spatially structured heterogeneity through the inclusion of random effects. By accounting for the spatial distribution of cyber risk, these models provide insights into the vulnerability of different areas and the potential impact of cyberattacks. This enables the development of more targeted risk management strategies, so as to better price cyber-risk.

We conduct an empirical analysis focusing on 49 states in the US, which shows that models able to exploit spatial correlation provide better fit performance and are thus more suitable for risk management. Specifically, based on widely accepted model comparison criteria and probability integral transform (PIT) histograms, we show that models incorporating spatially structured random effects provide the best estimation of cyberattack risk.

In addition, based on the five proposed models, we examine the pricing of an insurance policy against cyberattacks using three different premium principles. The results show that the premiums obtained using the spatially uniform cyberattack frequency model are significantly different from those obtained with the other four models. Our investigation indicates that it is inequitable to allocate premiums among the 49 states based on the assumption of homogeneous cyberattack frequency. These findings underscore the importance of using spatial modeling techniques in insurance pricing, as they allow more accurate estimates of cyberattack risk and more informed pricing decisions.

The remainder of this article is structured as follows. Section 2 provides an overview of the actuarial approach to cyber risk. Section 3 briefly describes alternative spatial models for assessing cyber risk, focusing on assumptions about the data generation process and the Bayesian approach used to construct and estimate the models. Section 4 presents the empirical results and key findings of the spatial models used, as well as the actuarial methodology we propose for estimating expected losses. Finally, Section 5 concludes.

2 | Cyber Risk in the Actuarial Domain

Historically, cyber risk analyses have focused on identifying technological vulnerabilities rather than quantifying financial losses. However, recent studies, such as the OECD 2017 report and the Ponemon Institute's 2019 study, have begun to address the issue of financial impact, shedding light on the growing importance of understanding the financial consequences of cyber incidents. For example, the Ponemon report notes that "the total cost for each company in the panel increased from 11.7 million US dollars in 2017 to a new high of 13.0 in 2018, with a rise of 12%."

Importantly, cyber events can result in a range of liabilities to third parties, including customers, suppliers, employees, and shareholders. In addition to direct financial losses, cyber incidents can also result in other costs, such as fines and penalties imposed by regulatory bodies (e.g., GDPR for EU states), incident response costs, and compensation for data breaches. Thus, it is becoming increasingly clear that the financial impact of cyber incidents can go well beyond direct losses.

As the cyber risk landscape continues to evolve and expand, the insurance industry has also recognized the growing importance of cyber risk and has begun to play an increasingly active role in the risk management process. Insurers have developed cyber insurance policies to help organizations manage and mitigate their cyber risks, thus becoming key players in the effort to address cyber risk.

The rapidly evolving nature of the cyber risk landscape presents significant challenges for actuaries working in this area. Unlike other fields with long histories and decades of historical data, the lack of long-term data on cyber risk makes it difficult for actuaries to accurately assess the risk and develop appropriate risk management strategies. As noted in Böhme, Laube, and Riek [10], the existing datasets "quickly become obsolete since the threats, vulnerabilities and mitigation methods develop rapidly." As a result, actuaries must rely on a range of alternative data sources and modeling techniques to effectively manage cyber risk.

The actuarial literature on cyber risk management is characterized by a wide range of methods and contexts, resulting in significant variations in the reported findings. For example, Biener, Eling, and Wirfs [11] compute the average cost per cyber incident at 40 million over 994 incidents occurring between 1971 and 2009. In contrast, NetDiligence [12, 13] report a much lower average cost of 0.7 million over 1201 claims filed between 2013 and 2017. This variation in reported costs highlights the challenge of accurately assessing cyber risk in the absence of long-term historical data.

Notably, the context of each cyber incident can vary significantly, adding to the complexity of accurately pricing cyber risk. According to Society of Actuaries [14], there are also differences between companies in different industries, further complicating the assessment of cyber risk. These challenges illustrate why pricing cyber risk remains difficult and why the insurance market for cyber risk is still in its infancy. Continued research and development in this area is necessary to improve current understanding of cyber risk and develop effective risk management strategies.

The cyber insurance market is developing mainly in the areas of "companies processing large amounts of personal data (telecommunication and media companies, health care, education, etc.), critical infrastructure companies (energy, communications), companies whose business is based on online transactions (retail, payment systems, financial institutions), a combination of the above (transport companies, health care)" [15]. In general, the cyber insurance market has some drawbacks, including a lack of standardization, limitations on the amount of coverage, and several exclusions in policy contracts. These factors can make it difficult for companies to accurately assess their cyber risk and select policies that provide adequate coverage. Addressing these challenges will require continued efforts to improve policy standardization and develop more comprehensive and flexible coverage options that can adapt to the evolving cyber risk landscape.

3 | The Statistical Methodology

We present a modeling strategy for estimating the cyberattack risk for $S = 49$ US states, including the continental states (excluding Alaska) and the District of Columbia. This study region is spatially connected and lends itself to straightforward spatial analysis. We explore several different models, paying particular attention to the hypotheses regarding the data generating process. The simplest model assumes homogeneous risk, while the most complex model allows for both unstructured and spatially structured heterogeneity through the inclusion of random effects. The range of models presented provides a flexible approach to capturing the complex nature of cyber risk and its spatial patterns.

To build and estimate the models, we use a Bayesian approach, which is a popular choice when dealing with complex hierarchical mixed models, especially those involving spatial data. This approach is particularly suited to our application because of its ability to propagate uncertainty about model parameters in the posterior distribution of cyberattack risk. By expressing this uncertainty in the posterior distribution, we can easily sample and combine it with simulations from the "claim size" distribution for insurance policy pricing.

To begin, we specify the same likelihood function for all considered models. Let N_i and F_i denote the number of cyberattacks and the number of firms in State i respectively, $i = 1, \dots, S$. Since N_i is count data, the Poisson model is a natural choice and provides a solid foundation for building more complex models that can capture spatial heterogeneity:

$$N_i | F_i, R_i \sim \text{Poisson}(F_i R_i), \qquad i = 1, \dots, S$$

where R_i denotes the risk of a cyberattack in State i. In this model, the expected cyberattack count is given by $F_i R_i$, where F_i is an offset that accounts for the size of each state (measured by the number of firms) and R_i is a model parameter. This approach allows us to estimate the relative risk of cyberattacks across different states while controlling for differences in state size. However, by incorporating additional factors, such as spatial correlation and other predictors, we can construct more sophisticated models that better capture the complexity of cyber risk. Specifically, to develop the model hierarchy, we use a log-linear predictor for R_i, which is specified as a generalized linear model. Below, we present five alternative models, denoted as $M1$–$M5$, that differ in their assumptions about the spatial structure and heterogeneity of cyber risk across states.

3.1 | M1: Intercept-Only Model

As a first naive model, we consider the intercept-only model, which implies the assumption that the risk is homogeneous across states, that is, $R_i = R$, $i = 1, \dots, S$. Thus, the linear predictor is:

$$\log(R_i) = \alpha \tag{1}$$

Specifying a diffuse Gaussian prior for the intercept term, $\alpha \sim \mathcal{N}(0, 1000)$, will give a posterior mean of R very close to the maximum likelihood estimate $\sum_{i=1}^{S} N_i / \sum_{i=1}^{S} F_i$.

While the homogeneous Poisson model assumes the same level of risk in all areas, empirical applications have shown that this model is often unrealistic. Insurers must therefore price cyber risk differently from state by state to account for differences in the risk of cyberattacks. However, the homogeneous Poisson model still serves as a natural starting point for building more complex models that can capture spatial heterogeneity. These models are proposed and compared in the following.

3.2 | M2: Fixed-Effects Model

To account for heterogeneity in cyberattack risk, the simplest approach is a fixed effects model with state-specific intercepts:

$$\log(R_i) = \alpha + \nu_i, \qquad \sum_{i=1}^{S} \nu_i = 0 \tag{2}$$

where ν_i is the deviation from the total intercept α. Note that the sum-to-zero constraint in Equation (2) is necessary to ensure model identifiability. The fixed effects structure of this model is reflected in the prior specification. Specifically, we use *independent* diffuse Gaussian priors for each model parameter:

$$\alpha \sim \mathcal{N}(0, 1000), \qquad \nu_i \sim \mathcal{N}(0, 1000), \quad i = 1 \dots, S$$

The priors are chosen to be non-informative, allowing the data to drive the inference and a wide range of model parameter values to be explored. Note that since the predictor is linear in the log scale, setting the prior variance to 1000 is sufficient to assign non-negligible prior probabilities to risk values observed in real-world applications.

Although more flexible than the intercept-only model, the fixed effects model in Equation (2) assumes prior independence among the risks. The use of diffuse priors results in posterior means that are very close to the state-specific maximum likelihood estimate of risk N_i/F_i. As a result, each state-specific risk estimate relies only on data from the state itself, neglecting potentially useful information provided by data available from other states.

The approach of using state-specific estimates may not fully capture the spatial patterns and correlations in cyber risk, which is a well-known limitation when modeling rare events such as cyberattacks. This limitation is known in the literature as the small-area problem, where the term "small" refers to the rarity of the phenomenon under study and the weak empirical evidence provided by individual state-specific data. This results in high sampling variability and high uncertainty associated with the estimates, which can lead to unreliable inferences.

To overcome the limitations of state-specific estimates, we follow a borrowing strength procedure that allows us to leverage information from neighboring areas and improve the statistical efficiency of the estimates. By borrowing strength across regions, we can obtain estimates that are a compromise between area-specific data and data collected from the entire spatial domain, capturing the spatial heterogeneity and correlation of cyber risk across different areas.

The process of borrowing strength is often performed locally, with neighboring regions playing a crucial role in determining the estimate for a given area. This is consistent with the hypothesis, often reasonable when analyzing socioeconomic phenomena, that things that are close in space are more similar than those that are far away. This principle is also relevant to disease mapping, which involves modeling the number of deaths observed in a human population exposed to risk. The disease mapping literature is built on these principles, and we can apply similar models to insurance data collected over space to estimate the spatial distribution of cyber risk. Below, we present several models that have been extensively studied in the disease mapping literature and that can be valuable tools for estimating cyber risk across regions.

3.3 | M3: Exchangeable Random Effects Model

The first extension involves exchangeable random effects, which assumes that the risks are heterogeneous and can be considered as random draws from a Gaussian population in the log scale. The structure of the linear predictor is the same as in model (2), but with the important difference that area-specific deviations ν_i from the overall intercept α are independent only conditionally on the heterogeneity parameter σ_ν^2, which unconditionally introduces dependence between the areas.

The conditional distribution of the random effects is

$$\nu_i|\sigma_\nu^2 \sim \mathcal{N}(0, \sigma_\nu^2), \quad i = 1 \dots , S$$

Following a fairly standard choice in Bayesian analysis, a Gamma prior is specified for the variance parameter: $\sigma_\nu^2 \sim \text{Gamma}(a, b)$.

The exchangeable random effects model provides estimates of the area-specific risk R_i, obtained as a weighted average of the observed data of area i and the overall risk. This approach introduces a shrinkage effect toward the global mean of the direct estimates, which is more pronounced for areas with a smaller number of firms F_i, and consequently with weaker empirical evidence. In contrast, direct estimates obtained from states with a high number of firms F_i and stronger empirical evidence are preserved. This shrinkage effect can improve the precision of the estimates and reduce the sampling variability and uncertainty associated with the estimation process, particularly for areas with limited data.

One of the drawbacks of model $M3$ is that it does not account for local spatial correlation, inducing shrinkage toward global risks, but neglecting local behavior. To address this limitation, we introduce models $M4$ and $M5$, which allow for more flexible spatial structures.

3.4 | M4: Spatial Random Effects Model

Spatial dependence between areas is introduced by specifying a Gaussian Markov random field (GMRF, see [16] for a full description) for the area-level random effects. GMRFs are typically defined by the inverse of their covariance matrix, known as the precision matrix, which describes the conditional dependence relationships between areas. The sparseness of this matrix provides notable computational advantages. Again, the linear predictor has the same structure as model (2), but the joint distribution of the random vector $\boldsymbol{\nu} = (\nu_1, \dots , \nu_S)^\top$ is constructed using the adjacency matrix $\mathbf{W}$. This is a symmetric S-dimensional matrix with entries set as follows:

$$\begin{cases} w_{ij} = 1 & \text{if} \quad i \sim j \\ w_{ij} = 0 & \text{otherwise} \end{cases}$$

The notation $i \sim j$ denotes that area i is a neighbor of area j. Following a standard choice, in this article, we consider areas as neighbors if they share a common boundary. The row sums of the adjacency matrix $d_i = \sum_{j=1}^S w_{ij}$ correspond to the number of neighbors of each area and are collected in the diagonal matrix $\mathbf{D} = \text{diag}(d_1, \dots , d_S)$, so that the precision matrix is obtained as

$$\mathbf{K}_\nu = \mathbf{D} - \mathbf{W}$$

and is positive semi-definite because the row sums are all equal to zero. Therefore, the joint distribution of the spatial random effects is improper and a sum-to-zero constraint is required for model identifiability. In particular, the joint distribution is

$$\boldsymbol{\nu} \sim \mathcal{N}_S(\mathbf{0}, \sigma_\nu^2 \mathbf{K}_\nu^-)$$

where σ_ν^2 is a scaling parameter and $\mathbf{K}_\nu^-$ is the generalized inverse of $\mathbf{K}_\nu$. The model hierarchy is completed by specifying the prior $\sigma_\nu^2 \sim \text{Gamma}(a, b)$.

3.5 | M5: Spatial and Exchangeable Random Effects Model

The last model is based on the Besag York and Mollié (BYM) specification [17], which is a popular approach designed to account for both spatially structured and unstructured heterogeneity in spatial data. The BYM model includes two random effects to capture these sources of variation, and the area-specific term of Equation (2) is modeled as follows:

$$\nu_i = \psi_i + \phi_i \tag{3}$$

where ψ and ϕ denote the exchangeable and spatial components, respectively, with priors

$$\psi_i|\sigma_\psi^2 \sim \mathcal{N}(0, \sigma_\psi^2) \quad i = 1, \dots, n, \qquad \boldsymbol{\phi} \sim \mathcal{N}_n(\mathbf{0}, \sigma_\phi^2 \mathbf{K}_\phi^-)$$

Non-identifiability of model (3) requires sum-to-zero constraints on both random effect vectors. Again, Gamma priors are assumed for the scaling parameters, that is, $\sigma_\phi^2 \sim \text{Gamma}(a_\phi, b_\phi)$ and $\sigma_\psi^2 \sim \text{Gamma}(a_\psi, b_\psi)$. There are numerous contributions in the literature on prior specification of the parameters, which aim to manage the a priori weight of the random effects. Some of these contributions propose interesting re-parameterizations of the model (see, for example, Riebler et al. [18]). In this article, we present a standard analysis of the model using common choices that are appropriate for the specific application we are considering.

3.6 | Model Estimation and Comparison

Bayesian inference involves summarizing the posterior distribution. If we denote the parameter vector as θ, then the posterior density is proportional to the product of the likelihood and the prior density, namely

$$\pi(\theta|\mathbf{N},\mathbf{F}) \propto \pi(\mathbf{N}|\theta, \mathbf{F})\pi(\theta)$$

where $\mathbf{N} = (N_1, \dots, N_i, \dots, N_S)$ and $\mathbf{F} = (F_1, \dots, F_i, \dots, F_S)$ are vectors containing the number of cyberattacks and firms, respectively. The parameter vector θ for each model is reported in Table 1. The posterior distribution for all models cannot be obtained in closed-form and must be computed through numerical approximation. For this purpose, two widely used strategies are Monte Carlo Markov chain (MCMC) sampling and integrated nested Laplace approximations (INLA; Rue, Martino, and Chopin [19]). INLA is particularly efficient for estimating latent GMRF models because it provides a highly accurate approximation of the posterior distribution and is computationally much faster than MCMC methods. INLA has been made easily accessible the `INLA` package [20, 21], a valuable tool for practitioners to use in applied Bayesian inference. Given these advantages, we use `INLA` to estimate our models, and the R code for our analysis is available upon request from the authors.

As a by-product of model estimation, INLA provides several measures of model performance and also allows us to draw random samples from the posterior distribution. This capability is critical to our study because we will use posterior samples for risks R_i, $i = 1, \dots, S$ and frequencies N_i, $i = 1, \dots, S$ along with simulations from the claim size distribution to compute insurance premiums.

TABLE 1 | Model parameters for models $M2$–$M5$.

Model	θ
$M1$: Intercept only	$\mathbf{R}, \alpha$
$M2$: Fixed effects	$\mathbf{R}, \alpha, \boldsymbol{\nu}$
$M3$: Exchangeable random effects	$\mathbf{R}, \alpha, \boldsymbol{\psi}, \sigma_\psi^2$
$M4$: Spatial random effects	$\mathbf{R}, \alpha, \boldsymbol{\phi}, \sigma_\phi^2$
$M5$: BYM model	$\mathbf{R}, \alpha, \boldsymbol{\psi}, \boldsymbol{\phi}, \sigma_\psi^2, \sigma_\phi^2$

TABLE 2 | Model comparison.

Model	DIC	WAIC	CPO
$M1$: Intercept only	1251.4	1286.2	643.5
$M2$: Fixed effects	392.9	378.0	242.7
$M3$: Exchangeable random effects	391.7	382.0	233.5
$M4$: Spatial random effects	388.2	379.1	222.4
$M5$: BYM model	388.2	379.1	222.4

To assess how well the five models describe the empirical data, we use three different measures of fit: the deviance information criterion (DIC), the widely applicable information criterion (WAIC), and the conditional predictive ordinate (CPO; [22]). The first two are widely used information criteria that assess model fit while taking into account model complexity [23, 24]. The CPO is a cross-validation criterion that is computed as follows:

$$\text{CPO} = -\sum_{i=1}^{S} \ln(\text{CPO}_i)$$

where

$$\text{CPO}_i = \pi(N_i|N_{-i}) = \int \pi(N_i|N_{-i}, \theta)\pi(\theta|N_{-i})\text{d}\theta, \quad i = 1, \dots, S$$

and N_{-i} denotes all the observations but the ith. For all three measures, lower values indicate better fit.

Table 2 shows the results of the model comparison. These results are obtained using data of breach cyber risk provided by the Privacy Rights Clearinghouse as described in the next section. We observe that the intercept-only model performs significantly worse than the other models, suggesting that cyberattack risk varies spatially across the United States. Both the DIC and CPO criteria favor spatial models, with no preference between the model with only spatial random effects ($M4$) and the model that includes both structured and unstructured random effects ($M5$). However, according to the WAIC criterion, the fixed effects model provides a slightly better fit.

In Figure 1, we present the PIT histograms for models $M2$ to $M5$ (excluding the intercept-only model, which provides very poor fit). As we can see, the performance of $M2$ is worse than that of all other models. Based on the goodness of fit measures in Table 2 and the PIT histograms in Figure 1, we can conclude that models

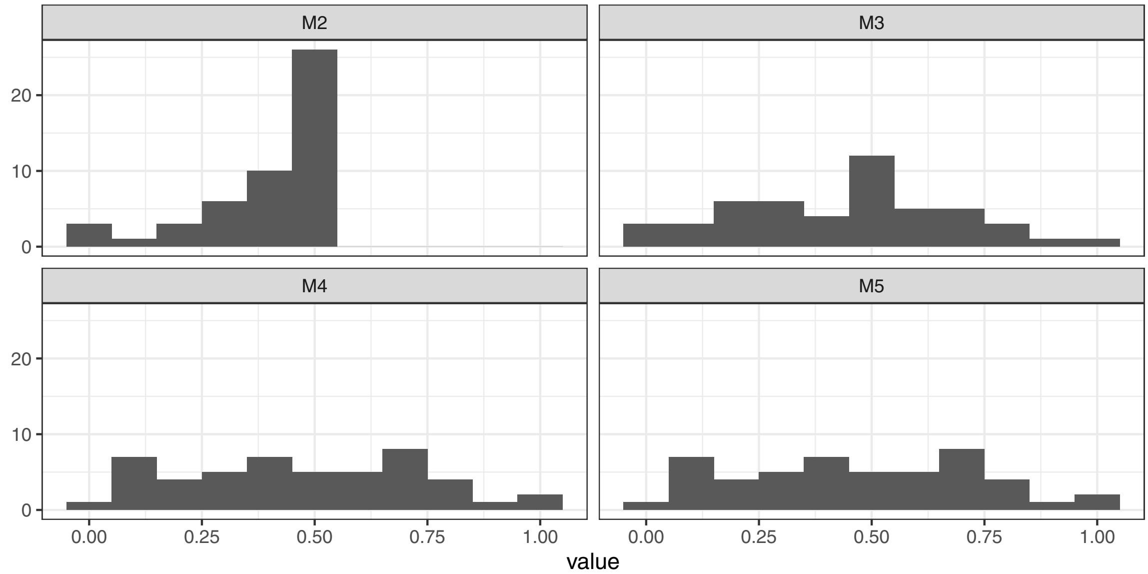

FIGURE 1 | Probability integral transform histogram for models $M2$–$M5$.

$M4$ and $M5$, which include spatially structured random effects, provide a better estimate of cyberattack risk. In the next section, we will further explore the differences between the five models when evaluating cyber insurance premiums.

4 | Insurance Application

In this section, we price a data breach insurance policy for cyber risk. To estimate the probability distribution of claim frequency, we use the five spatial models ($M1$, $M2$, …, $M5$) described in Section 3.

4.1 | Data

To estimate the five models, we need data on the number of data breaches. For this purpose, we utilize the data provided by the Privacy Rights Clearinghouse (https://privacyrights.org/data-breaches). This dataset reports the number of data breach attacks experienced by US firms from 2005 to 2019, as well as the type and geographic location (latitude and longitude) of each attack. The database covers the entire time interval from 2005 to 2019, so the spatial models M1–M5 evaluate the posterior distribution of the number of cyberattacks over a 15-year period. However, in the following, we will compute the premium for a data breach insurance policy that provides coverage for one year. Therefore, since the premium will be calculated using Monte Carlo simulation (see below), we will divide the number of cyberattacks sampled from the posterior distribution provided by models $M1$–$M5$ by 15.

Furthermore, we need the number of firms in each state. We gather these data from the Statistics of U.S. Businesses (SUSB) Annual Data Tables for the year 2017, provided by the United States Census Bureau (https://www.census.gov/data/tables/2017/econ/susb/2017-susb-annual.html).

For the ith state (and federal district), $i = 1, 2, \dots, 49$, and each of the five models $M1, M2, \dots, M5$ we will compute the probability distribution of the total claims paid in one year by an insurer offering protection against the cyber risk of data breaches to all F_i firms in that state. To this aim, we will also need the probability distribution of the cost $Y_{j,i}$ incurred by the insurer for the jth cyber risk attack in the ith state. Unfortunately, data for the on cyber risk losses are not available at the state level (at least to the best of our knowledge). Therefore, we model $Y_{j,i}$ as i.i.d. log-normally distributed random variables, and set the mean of $Y_{j,i}$ equal to the average cost attributed to the US market in IBM [25], which is 9.05 million dollars, and assume a coefficient of variation (the ratio of the standard deviation (SD) to the mean) of 10.95, as reported in Biener, Eling, and Wirfs [11]. With the coefficient of variation and the mean, we can compute the variance and fully determine the (LogNormal) distribution of $Y_{j,i}$ using the method of moments. Specifically, the above calculation yields $Y \sim \text{LogNormal}(13.621, 2.190)$, with $E(Y) = 9.05$ million dollars and $\sigma(Y) = 99.11$ million dollars.

Summarizing, the data that we retrieved from the aforementioned databases and data sources to perform our analysis are: the number of data breach attacks experienced from 2005 to 2019 in the 49 states, with the geographic location of each attack, the number of firms per state (F_i, $i = 1, 2, \dots, 49$), and the mean and coefficient of variation of the probability distribution of the data breach incidents, modeled as i.i.d. log-normal random variables.

4.2 | Results

The total claims paid in a year in the ith state is computed as follows:

$$Z_i = \sum_{j=1}^{N_i} Y_{j,i} \tag{4}$$

where N_i is the number of cyber risk attacks in the ith state in one year and $Y_{j,i}$ is the cost incurred by the insurer for the jth cyber risk attack in the ith state. Based on (4), we can evaluate the probability distribution of Z_i by Monte Carlo simulation. Specifically, we first simulate N_i by drawing it from the posterior distribution computed with models $M1$–$M5$ (we divide the Monte Carlo value by 15 since $M1$–$M5$ are estimated using data covering a 15-year time interval). Then, for all of these cyberattacks, we simulate $Y_{j,i}$ as i.i.d. log-normally distributed random variables with $E(Y) = 9.05$ million dollars and $\sigma(Y) = 99.11$ million dollars (the probability distribution of the severity that we estimated in the previous subsection).

For the convenience of readers, the entire Monte Carlo simulation procedure is outlined below:

1. Select one of the spatial models: $M1$, $M2$, $M3$, $M4$, or $M5$.
2. Specify the number N of Monte Carlo simulations.
3. For $i = 1, 2, \dots, 49$, simulate the number N_i of cyberattacks in the ith state by drawing from the posterior Poisson distribution.
4. Simulate the losses due to the N_i cyberattacks by drawing them from the LogNormal distribution.
5. Compute the total loss for the ith state as the sum of the N_i losses previously obtained.
6. Recursively iterate the process N times and obtain the simulated probability distribution of the total loss for each state.

To compute the premium, we use three different premium principles, which we describe below (see Klugman, Panjer, and Willmot [26] and Pitacco and Olivieri [27]). Specifically, for the ith state, we compute the expense-loaded premium as follows:

$$EP_i = \frac{E[Z_i] + \delta_{(i,k)}}{1-\beta} \tag{5}$$

where $\delta_{(i,k)}$ is the safety loading for the ith state under the kth premium principle, and β is the percentage of expense loading (in the numerical experiments presented in this article, we use the values common to insurance practitioners, namely $\beta = 20\%$ and $\gamma = 15\%$).

The premium principles used to evaluate the safety loading are described below:

1. Percentile (P75) principle. We compute the safety loading as follows:
$$\delta_{(i,1)} = VaR_\alpha(Z_i) - E[Z_i] \tag{6}$$
where $VaR_\alpha(Z_i)$ is the α quantile of Z_i. In the numerical experiments we choose $\alpha = 75\%$.
2. Cost of Capital (CoC) principle. Let ρ denote the CoC, and assuming that the cyber risk will expire after one year, let $i_{rf}(0,1)$ denote the 1-year risk-free interest rate.
According to the Solvency II Directive 2009/138/EC of the European Parliament [28] to evaluate the Solvency Capital Requirement (SCR), we compute the safety loading as follows:
$$\delta_{(i,2)} = \frac{\rho \cdot SCR_i}{1 + i_{rf}(0,1)} = \frac{\rho(VaR_{99.5\%}(Z_i) - E[Z_i])}{1 + i_{rf}(0,1)} \tag{7}$$
In the numerical experiments, we choose $\rho = 6\%$ (according to the Solvency II standard formula for quantifying the risk margin for insurance liabilities) and $i_{rf}(0,1) = 0$.

TABLE 3 | Average premiums and coefficients of variations.

	P75		CoC		SD	
Model	**Mean**	**CV**	**Mean**	**CV**	**Mean**	**CV**
$M1$	101.60	1.09	111.29	0.99	102.92	1.02
$M2$	102.19	1.36	111.15	1.23	103.20	1.28
$M3$	102.13	1.35	111.13	1.23	102.74	1.28
$M4$	102.04	1.35	111.15	1.23	103.20	1.28
$M5$	102.05	1.35	111.11	1.23	102.70	1.28

3. Standard deviation (SD) principle. We compute the safety loading as follows:
$$\delta_{(i,3)} = \gamma \cdot \sigma[Z_i] \tag{8}$$
where $\sigma[Z_i]$ is the SD of the probability distribution of Z_i (computed according to the Monte Carlo simulation procedure described above). In the numerical experiments we use $\gamma = 15\%$, a value commonly used by practitioners.

Figure 2 reports the spatial distribution of the state-level premiums divided by the number of firms F_i, $i = 1, \dots, S$. These maps are determined by the spatial distribution of the cyber-attack risk, while observed differences are determined by the different principles. It is worth noticing that the CoC approach always yields the highest premiums, indicating that the right tail of the probability distribution of losses due to cyber attacks is quite heavy. Indeed, the distance between the 99.5% percentile and the mean $E[Z_i]$, when multiplied by the small ρ value of 6% and divided by $1 + i_{rf}(0,1)$ (see formula (7)), exceeds both the distance between the 75th percentile and the mean (as per formula (6)) and the SD multiplied by the value of γ, which is 15% (as per formula (8)).

We report the average premiums (across all 49 states) and coefficients of variation below.

As shown in Table 3, the premium depends strongly on the principle used to compute it, but is less sensitive to the spatial model used. In particular, for each premium principle, the average premium across the 49 states remains relatively constant across the models. Consistent with the assumptions of the models, the premium computed using $M1$ is the most homogeneous across the 49 states, as evidenced by a coefficient of variation (CV) of 1.09 for P75, 0.99 for CoC, and 1.01 for SD, which is significantly lower than that of the other models. In addition, models $M2$–$M5$ produce similar coefficients of variation (as shown in Table 3).

Below we show the premiums obtained in each of the 49 states using models $M1$–$M5$ and the P75 premium principle (Table 4), the CoC premium principle (Table 5), and the SD premium principle (Table 6).

For each of the three premium principles, the results obtained with $M1$ differ significantly from those obtained with the other four models. For example, the risk scenario implied by the assumption that the frequency of cyberattacks is homogeneous across states is very different from the risk scenario obtained by assuming heterogeneous frequency. The model comparison presented in the previous section (where $M1$ fits the empirical

TABLE 4 | Pricing using the P75 principle.

STATE	$M1$	$M2$	$M3$	$M4$	$M5$
Alabama	63,827,829	41,019,006	41,984,799	41,920,041	41,788,630
Arizona	85,409,213	72,889,364	73,168,759	73,900,144	73,671,879
Arkansas	42,906,796	27,039,694	27,977,988	27,619,532	27,750,157
California	592,905,530	833,035,984	832,457,346	831,773,962	829,952,353
Colorado	112,745,358	106,777,959	106,105,652	104,698,700	104,850,218
Connecticut	55,548,125	85,305,920	83,253,447	84,449,348	84,661,750
Delaware	14,725,689	13,888,696	13,737,128	13,990,128	13,962,395
Distr. of Columbia	12,617,056	111,430,309	104,068,378	105,660,259	105,723,366
Florida	358,505,493	250,829,864	251,876,633	251,237,412	251,264,161
Georgia	152,316,803	155,751,966	155,238,197	153,410,604	153,360,335
Idaho	31,639,306	12,236,776	14,348,733	15,683,587	15,698,736
Illinois	205,621,217	195,211,628	194,663,955	193,798,514	194,065,921
Indiana	94,190,402	120,872,750	119,676,982	118,265,147	118,675,022
Iowa	53,802,728	41,944,893	42,213,664	41,068,321	41,043,989
Kansas	48,416,766	96,237,483	93,764,064	90,491,755	90,286,518
Kentucky	58,381,626	56,744,630	56,212,418	56,586,062	56,551,452
Louisiana	68,902,058	34,908,407	36,728,304	35,781,619	35,958,511
Maine	26,595,792	18,300,640	18,881,636	19,265,710	19,195,796
Maryland	89,046,630	149,110,846	146,672,703	147,417,642	147,225,126
Massachusetts	113,659,701	146,434,708	144,898,636	146,915,055	147,420,005
Michigan	143,366,249	80,965,483	82,349,992	84,387,187	84,349,203
Minnesota	98,064,055	81,881,675	81,882,734	78,493,679	78,402,628
Mississippi	38,622,437	15,782,377	17,776,637	18,032,361	18,035,421
Missouri	96,765,768	79,309,395	79,364,776	80,095,494	79,875,776
Montana	25,588,373	16,518,106	17,330,958	14,465,541	14,449,865
Nebraska	35,517,430	22,594,429	23,490,675	23,898,338	23,849,735
Nevada	40,233,509	38,453,828	38,116,590	37,441,041	37,451,104
New Hampshire	23,648,698	30,705,753	29,377,379	29,979,045	30,042,066
New Jersey	154,140,372	103,174,099	104,075,719	105,074,258	105,393,172
New Mexico	27,617,257	26,251,748	25,998,690	25,124,201	25,159,985
New York	361,282,494	515,354,793	514,542,573	512,388,486	513,666,061
North Carolina	149,284,941	125,610,291	125,335,412	125,880,881	125,500,406
North Dakota	15,542,686	2,762,328	5,431,239	4,906,943	4,956,888
Ohio	156,209,076	162,657,065	162,181,117	161,027,389	161,103,962
Oklahoma	60,719,661	38,457,031	39,362,419	41,366,810	41,543,993
Oregon	76,355,156	71,478,469	71,187,720	71,145,859	71,108,472
Pennsylvania	188,861,246	153,219,860	153,262,015	155,360,866	154,890,502
Rhode Island	17,689,620	23,589,686	22,286,349	23,888,025	23,902,909
South Carolina	69,688,643	44,576,986	45,430,907	45,738,727	45,722,914
South Dakota	17,447,788	4,471,178	6,820,263	6,998,145	6,964,088
Tennessee	86,057,380	88,015,060	87,176,578	84,637,402	84,685,599
Texas	364,564,252	313,509,000	314,436,904	311,792,224	312,238,840
Utah	52,392,488	39,424,335	39,814,929	39,072,881	39,010,686
Vermont	13,200,243	22,787,340	20,681,310	21,753,226	21,758,731
Virginia	131,418,277	139,591,574	139,063,455	141,877,243	141,762,736
Washington	124,362,534	117,256,091	117,136,813	116,611,266	116,769,547
West Virginia	22,447,244	13,021,422	14,189,333	16,128,123	16,228,850
Wisconsin	92,084,077	62,174,293	62,693,578	62,364,983	62,751,243
Wyoming	13,697,947	3,624,363	5,745,520	5,944,138	5,948,057
TOTAL	4,978,634,017	5,007,189,580	5,004,472,005	4,999,808,303	5,000,629,755

Note: Premiums are expressed in dollars.

TABLE 5 | Pricing using the CoC principle.

STATE	$M1$	$M2$	$M3$	$M4$	$M5$
Alabama	75,053,813	50,015,480	51,557,059	51,391,463	50,999,309
Arizona	97,425,670	83,560,438	83,771,962	84,899,871	84,342,495
Arkansas	53,194,035	35,331,854	36,389,169	35,945,298	36,100,556
California	595,583,313	828,522,524	827,035,551	828,113,318	826,031,709
Colorado	124,486,616	117,999,596	117,528,306	116,351,241	115,906,089
Connecticut	66,522,758	95,528,750	94,525,015	94,780,814	94,956,093
Delaware	21,741,838	21,176,028	21,009,667	20,905,363	21,095,447
District of Columbia	19,840,568	122,606,136	114,498,007	116,520,962	117,648,478
Florida	366,168,956	259,883,484	261,617,101	261,747,288	261,166,354
Georgia	164,177,546	167,574,127	165,530,171	164,747,314	163,734,711
Idaho	41,096,415	18,923,244	21,721,813	23,528,088	23,233,031
Illinois	217,194,371	205,894,728	205,005,116	203,897,989	203,871,481
Indiana	104,961,686	131,700,823	131,733,178	129,104,423	129,964,195
Iowa	63,998,634	51,771,915	52,318,577	50,155,358	50,737,193
Kansas	58,494,016	107,625,792	104,555,118	101,438,046	101,150,951
Kentucky	68,982,265	67,902,484	66,163,740	67,346,053	67,012,642
Louisiana	79,298,458	44,901,004	46,395,155	45,158,369	44,630,282
Maine	35,555,982	25,947,025	26,720,593	27,129,845	26,954,602
Maryland	100,452,498	159,363,689	157,256,189	158,448,923	158,738,042
Massachusetts	124,698,276	157,100,501	155,342,144	157,685,195	158,095,408
Michigan	154,781,019	92,159,635	93,206,764	95,272,151	95,831,980
Minnesota	109,347,192	92,622,164	93,020,362	89,680,895	88,905,909
Mississippi	48,331,355	23,305,791	25,418,033	26,098,089	25,602,484
Missouri	107,556,856	90,107,472	90,124,505	90,433,281	90,203,397
Montana	34,583,510	23,592,698	24,949,053	21,575,899	21,798,386
Nebraska	45,878,501	30,858,985	31,942,107	32,398,501	32,522,572
Nevada	50,009,026	47,818,784	47,842,183	47,020,950	47,080,708
New Hampshire	32,529,042	39,684,417	38,411,938	39,061,351	39,597,261
New Jersey	163,600,412	114,542,138	114,796,367	116,393,914	115,763,705
New Mexico	36,359,279	35,052,656	34,501,142	33,640,541	33,603,442
New York	367,892,733	518,618,416	517,069,184	516,788,577	516,235,638
North Carolina	160,810,894	136,304,359	136,476,437	136,662,969	136,938,898
North Dakota	22,849,770	6,906,331	10,527,761	9,950,587	9,853,460
Ohio	166,207,502	172,902,204	172,214,773	172,465,905	170,668,062
Oklahoma	70,977,834	48,034,462	48,677,321	51,025,668	51,313,004
Oregon	87,227,261	82,382,922	81,803,958	81,982,657	81,083,382
Pennsylvania	199,220,370	164,178,066	164,272,395	165,225,700	165,577,202
Rhode Island	25,532,902	31,961,559	30,439,516	32,109,724	32,345,369
South Carolina	81,262,491	53,732,089	55,123,644	55,614,358	55,271,206
South Dakota	25,403,686	9,225,304	12,287,927	12,500,248	12,601,974
Tennessee	96,796,179	99,040,796	98,666,561	95,603,119	95,620,645
Texas	372,413,180	320,601,287	321,099,539	320,988,107	321,830,205
Utah	63,271,943	49,029,452	49,142,072	48,326,617	48,723,463
Vermont	20,255,409	30,929,856	28,830,517	29,593,188	30,241,040
Virginia	142,385,927	150,931,644	150,277,422	152,552,673	153,344,234
Washington	134,641,944	128,090,042	128,175,792	126,384,163	127,061,322
West Virginia	31,123,918	20,220,100	21,410,531	23,455,393	23,604,845
Wisconsin	102,561,549	72,266,920	72,890,190	72,782,918	73,517,542
Wyoming	20,688,078	7,814,402	10,877,360	11,429,486	11,140,170
TOTAL	5,453,427,475	5,446,244,573	5,445,148,987	5,446,312,851	5,444,250,574

Note: Premiums are expressed in dollars.

TABLE 6 | Pricing using the SD principle.

STATE	$M1$	$M2$	$M3$	$M4$	$M5$
Alabama	67,144,816	44,724,553	45,636,241	44,851,773	45,339,661
Arizona	90,513,983	74,750,466	75,973,074	77,991,464	75,028,760
Arkansas	46,491,205	30,452,781	31,422,999	31,208,235	31,552,330
California	569,673,321	798,767,233	797,366,532	800,527,908	797,954,615
Colorado	113,056,513	107,207,467	107,777,114	105,566,035	105,562,360
Connecticut	59,669,744	91,726,343	84,805,489	85,574,353	85,597,494
Delaware	19,561,625	17,962,726	17,923,808	18,920,669	18,082,782
Distr. of Columbia	17,230,877	113,456,850	103,950,854	104,865,835	106,919,842
Florida	344,947,822	243,532,815	243,984,481	248,075,187	244,625,440
Georgia	151,415,515	154,533,699	152,879,092	152,494,583	151,013,932
Idaho	35,828,781	16,739,604	18,167,607	21,212,220	19,498,671
Illinois	201,256,741	191,714,018	197,010,985	192,059,956	189,997,317
Indiana	96,275,687	120,929,241	119,591,951	119,074,212	118,878,593
Iowa	56,897,442	45,401,380	45,643,492	45,033,216	44,383,898
Kansas	51,349,104	98,709,789	94,741,752	91,761,701	91,263,084
Kentucky	61,013,114	61,625,857	59,079,612	61,149,745	60,400,141
Louisiana	71,831,379	39,616,078	40,909,634	39,377,507	38,737,879
Maine	31,038,465	22,032,830	23,022,136	23,079,598	23,226,541
Maryland	91,324,195	146,472,443	147,650,899	145,226,494	146,363,504
Massachusetts	114,740,829	145,480,198	143,969,849	145,118,589	146,923,402
Michigan	142,088,615	83,551,385	83,822,482	86,851,509	87,913,657
Minnesota	98,914,577	83,704,868	83,417,592	80,449,210	81,631,929
Mississippi	42,221,217	22,006,386	22,157,986	22,879,468	21,586,083
Missouri	97,669,446	80,680,304	80,631,483	83,320,974	81,377,819
Montana	30,013,340	19,979,431	21,398,509	18,495,524	18,400,374
Nebraska	39,313,803	32,954,509	28,508,103	27,626,932	28,718,050
Nevada	43,881,643	42,996,882	41,827,789	40,568,711	42,607,551
New Hampshire	28,280,038	34,296,045	33,845,237	36,647,263	35,529,289
New Jersey	150,554,073	103,454,006	103,895,075	107,450,148	105,325,367
New Mexico	31,863,445	30,247,594	29,997,871	29,083,849	29,079,454
New York	349,431,129	497,187,896	494,716,819	494,437,712	492,229,601
North Carolina	146,885,025	125,166,305	124,429,266	125,742,265	124,103,827
North Dakota	19,567,884	6,698,932	10,027,585	9,403,711	7,956,189
Ohio	154,199,240	159,458,284	159,874,353	159,523,633	158,290,379
Oklahoma	63,769,578	41,416,504	42,340,151	45,267,813	44,932,159
Oregon	78,333,898	74,486,537	73,056,924	79,046,809	73,195,255
Pennsylvania	185,795,095	151,652,302	150,834,366	153,449,987	153,043,132
Rhode Island	25,129,475	29,158,130	25,884,412	27,981,876	27,729,988
South Carolina	89,933,829	47,022,718	49,355,560	48,860,577	49,002,542
South Dakota	20,968,259	7,841,791	10,922,064	10,765,353	11,616,222
Tennessee	88,068,183	91,186,045	88,073,305	87,775,740	87,248,459
Texas	353,883,874	302,978,686	304,686,306	308,843,440	306,022,175
Utah	59,798,862	49,584,346	43,470,636	43,242,682	42,975,086
Vermont	18,460,784	26,929,556	24,980,140	25,843,967	25,588,583
Virginia	130,513,886	139,446,399	140,090,660	141,033,948	141,324,557
Washington	123,010,220	117,741,739	118,721,608	115,035,127	115,619,286
West Virginia	27,243,105	17,736,312	18,026,208	20,070,022	19,946,603
Wisconsin	93,397,160	64,908,314	64,408,216	64,537,262	67,643,960
Wyoming	18,385,639	6,298,442	9,121,538	9,438,755	10,306,108
TOTAL	5,042,836,476	5,056,607,018	5,034,029,846	5,056,843,549	5,032,293,929

Note: Premiums are expressed in dollars.

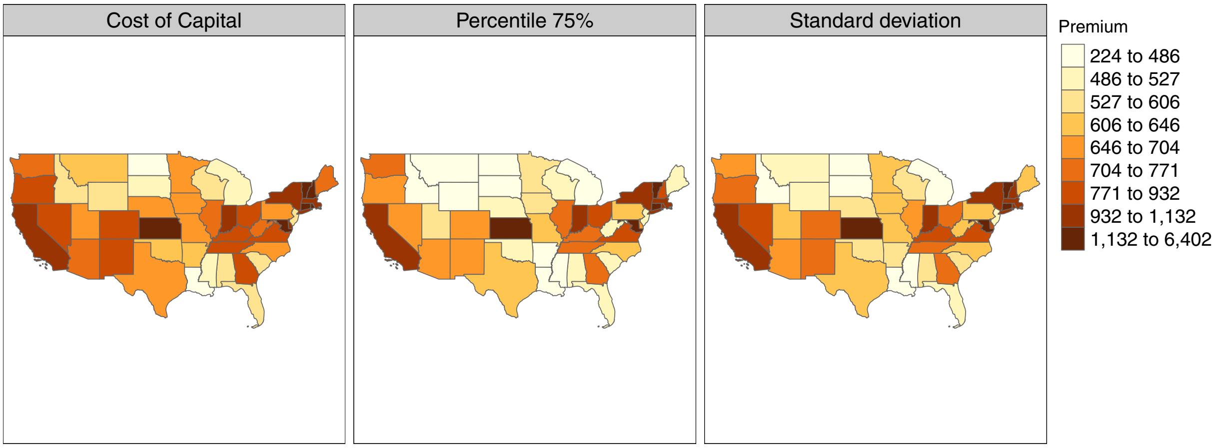

FIGURE 2 | Premiums divided by the number of firms in the continental US states obtained with the cost of capital principle (left), percentile 75% principle (middle) and standard deviation principle (right).

risk distribution much worse than the other models) suggests that it is unequitable to allocate premiums among the 49 states based on the assumption of homogeneous cyberattack frequency. When using the P75 and CoC principles, models $M4$ and $M5$ give very similar results (the maximum difference between the premiums is less than 1% for P75 and less than 2.6% for CoC). In addition, for most states, the premium obtained using $M2$ and $M3$ is similar to that obtained with $M4$ and $M5$. However, there are some states for which the results obtained using $M2$ and $M3$ differ from those obtained using $M4$ and $M5$ (the maximum difference between the $M2$ and $M5$ premiums is approximately 44%, and the maximum difference between the $M3$ and $M5$ premiums is approximately 19%). The consistency between the results obtained with models $M4$ and $M5$ is expected because the P75 and CoC principles rely on two quantiles of the posterior risk distribution (the 75th and 99.5th percentiles) that are relatively close to each other and are not affected by the extreme tail behavior.

When using models $M4$ and $M5$ and the SD criterion, the results for Idaho, North Dakota, South Dakota, and Wyoming differ significantly from the results obtained for the remaining 45 states. It is worth noting that Idaho, North Dakota, South Dakota, and Wyoming have a relatively small number of firms and, together with Montana, form a spatial cluster of low-risk states. Furthermore, these four states experienced the lowest number of cyberattacks during the study period (15 for Idaho, 4 for North Dakota, 6 for South Dakota, and 5 for Wyoming).

In this case, the difference between models $M4$ and $M5$, and in particular the fact that the spatial-only model $M4$ is less sensitive to the variability across states than model $M5$ (which includes both an unstructured and a structured random effect), leads to significant differences in the premiums (in relative terms).

Finally, when considering the impact of such differences on the overall premium distribution, worth noting is that the combined contribution of the four states to the total premium amount is approximately 1%.

5 | Conclusions

In this article, we use a statistical spatial modeling framework to estimate the risk of cyberattacks across geographic areas corresponding to the continental US. We implement the models following the Bayesian paradigm, which allows obtaining the posterior distribution of the number of data breaches or cyberattacks and to naturally simulate from the posterior predictive distribution the number of attacks, which is useful for pricing policies.

We show that models able to exploit spatial correlation provide better fit performance (model comparison based on widely adopted selection criteria, such as DIC, CPO, and PIT), making them more suitable for sensitive policy pricing.

In addition, we use the posterior predictive distribution of the number of data breaches to calculate the premium for a one-year cyberattack insurance policy. To evaluate safety loadings, we use three different premium principles: the 75% percentile, the CoC, and the SD principle. For each of these principles, the results obtained using the spatially uniform cyberattack frequency model differ significantly from those obtained with the other four models.

As pointed-out by an anonymous reviewer, the spatial precision matrix $\mathbf{K}_\nu$ could be specified by adopting different criteria and possibly by embedding prior knowledge and expert opinions concerning the similarity of US states with respect to the phenomenon being analyzed. An interesting proposal in this spirit can be found in Majumdar et al. [29]. For the sake of brevity, we do not show results obtained by using different spatial structures and a covariance matrix obtained by combining the spatial structure of the map with weights determined by the correlation between states with respect to cyber attacks. In fact, policy pricing envisioned in the article depends on the posterior distribution of the linear predictor, which shows negligible sensitivity to the adopted structure in the application being studied.

In particular, our investigation shows that it is inequitable to allocate premiums among the 49 states based on the assumption of homogeneous cyberattack frequency. Therefore, it is important to consider the spatial correlation and heterogeneity of cyberattack frequency when pricing policies.

Acknowledgment

Open access publishing facilitated by Universita degli Studi di Bologna, as part of the Wiley - CRUI-CARE agreement.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. European Systemic Risk Board, "Systemic Cyber Risk" (technical report, European Systemic Risk Board, 2020).

2. K. Stouffer, T. Zimmerman, C. Tang, J. Lubell, J. Cichonski, and J. McCarthy, "Cybersecurity Framework Manufacturing Profile" (technical report, National Institute of Standards and Technology, 2019).

3. T. W. Moore, S. B. C. D. Tandy, and F. Chang, "Identifying How Firms Manage Cybersecurity Investment" (working paper, Darwin Deason Institute for Cyber Security, Southern Methodist University, 2015).

4. M. Carannante, V. D'Amato, P. Fersini, S. Forte, and G. Melisi, "Vine Copula Modeling Dependence Among Cyber Risks: A Dangerous Regulatory Paradox," *Applied Stochastic Models in Business and Industry* 39, no. 4 (2023): 549–566.

5. K. Ruan, "Introducing Cybernomics: A Unifying Economic Framework for Measuring Cyber Risk," *Computers & Security* 65 (2017): 77–89.

6. A. Mukhopadhyay, S. Chatterjee, D. Saha, A. Mahanti, and S. K. Sadhukhan, "Cyber-Risk Decision Models: To Insure It or Not?" *Decision Support Systems* 56 (2013): 11–26.

7. I. Aldasoro, L. Gambacorta, P. Giudici, and T. Leach, "The Drivers of Cyber Risk," *Journal of Financial Stability* 60 (2022): 100989.

8. K. Awiszus, T. Knispel, I. Penner, G. Svindland, A. Voß, and S. Weber, "Modeling and Pricing Cyber Insurance," *European Actuarial Journal* 13 (2023): 1–53.

9. N. Veerasamy, Y. Moolla, and Z. Dawood, "Application of Geospatial Data in Cyber Security," in *Proceedings of the 21st European Conference on Cyber Warfare and Security* (Reading, UK: Academic Conferences International Limited, 2022), 305–313, https://doi.org/10.34190/eccws.21.1.447.

10. R. Böhme, S. Laube, and M. Riek, "A Fundamental Approach to Cyber Risk Analysis," *Variance* 11 (2018): 161–185.

11. C. Biener, M. Eling, and J. Wirfs, "Insurability of Cyber Risk: An Empirical Analysis," *Geneva Papers on Risk and Insurance - Issues and Practice* 40, no. 1 (2015): 131–158.

12. NetDiligence, "Cyber Claims Study" (technical report, NetDiligence, 2016).

13. NetDiligence, "Cyber Claims Study" (technical report, NetDiligence, 2018).

14. Society of Actuaries, "Quantification of Cyber Risk for Actuaries" (technical report, Society of Actuaries, 2020).

15. M. F. Carfora, F. Martinelli, F. Mercaldo, and A. Orlando, "Cyber Risk Management: An Actuarial Point of View," *Journal of Operational Risk* 14, no. 4 (2019): 77–103.

16. H. Rue and L. Held, *Gaussian Markov Random Fields: Theory and Applications*. Monographs on Statistics and Applied Probability, vol. 104 (London, UK: Chapman & Hall, 2005), https://doi.org/10.1201/9780203492024.

17. J. Besag, J. York, and A. Mollié, "Bayesian Image Restoration, With Two Applications in Spatial Statistics," *Annals of the Institute of Statistical Mathematics* 43, no. 1 (1991): 1–20.

18. A. Riebler, S. H. Sørbye, D. Simpson, et al., "An Intuitive Bayesian Spatial Model for Disease Mapping That Accounts for Scaling," *Statistical Methods in Medical Research* 25, no. 4 (2016): 1145–1165, https://doi.org/10.1177/0962280216660421.

19. H. Rue, S. Martino, and N. Chopin, "Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations," *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 71, no. 2 (2009): 319–392.

20. F. Lindgren and H. Rue, "Bayesian Spatial Modelling With R-INLA," *Journal of Statistical Software* 63, no. 19 (2015): 1–25, https://doi.org/10.18637/jss.v063.i19.

21. T. G. Martins, D. Simpson, F. Lindgren, and H. Rue, "Bayesian Computing With INLA: New Features," *Computational Statistics & Data Analysis* 67 (2013): 68–83, https://doi.org/10.1016/j.csda.2013.04.014.

22. A. E. Gelfand, "Model Determination Using Sampling-Based Methods," in *Markov Chain Monte Carlo in Practice*, eds. W. R. Gilks, S. Richardson, and D. Spiegelhalter (London, UK: Chapman and Hall/CRC, 1996), 145–161.

23. D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde, "The Deviance Information Criterion: 12 Years On," *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 76, no. 3 (2014): 485–493.

24. S. Watanabe, "A Widely Applicable Bayesian Information Criterion," *Journal of Machine Learning Research* 14, no. 1 (2013): 867–897.

25. IBM, "Cost of Data Breach" (technical report, IBM, 2021).

26. S. Klugman, H. Panjer, and G. Willmot, *Loss Models: From Data to Decisions* (London, UK: John Wiley & Sons, 1998).

27. E. Pitacco and A. Olivieri, *Introduction to Insurance Mathematics* (London, UK: Springer, 2011).

28. European Union Parliament, "Solvency II Directive 2009/138/EC of the European Parliament," 2009.

29. A. Majumdar, C. E. Lennert-Cody, M. N. Maunder, and A. A. da Silva, "Spatio-Temporal Modeling for Estimation of Bigeye Tuna Catch in the Presence of Pandemic-Related Data Loss Using Parametric Adjacency Structures," *Fisheries Research* 268 (2023): 106813, https://doi.org/10.1016/j.fishres.2023.106813.