Cloud failure and cyber insurance: calibration of stress scenarios and diversification

Olivier Lopez, Daniel Nkameni

Cloud failure and cyber insurance: calibration of stress scenarios and diversification

Olivier Lopez[1], Daniel Nkameni[1,2]

October 11, 2024

Abstract

The expansion of the cyber insurance market is constantly under the threat of an accumulation event that would affect simultaneously a large number of policyholders. Very few experiences exist on such catastrophes, apart from worldwide cyber attacks like Wannacry and NotPetya in 2017 (in a context where cyber insurance coverage was lower). Nevertheless, the very nature of cyber risk makes the occurrence of such events plausible in the future. There is therefore a need for stress-testing in order to be sure that a portfolio can resist to such a crisis. In this perspective, the EIOPA recently published methodological guidelines specific to cyber. One of the most concerning scenarios is the potential vulnerability of a cloud outage catastrophe, that is the failure of a cloud provider whose solution is shared by a significant part of the portfolio. In the present work, we propose a way to model and calibrate such kind of cloud outage scenario. We also provide way to measure the level of diversification of a cyber insurance portfolio, and how this diversification may protect against such events. A by-product of our methodology is to provide guidelines to underwriters to help reduce the vulnerability of a portfolio to these cloud outage scenarios.

Key words: Cyber insurance ; stress scenarios ; cloud outage ; portfolio optimization.

Short title: Cloud failure and cyber insurance.

[1] CREST Laboratory, CNRS, Groupe des Écoles Nationales d'Économie et Statistique, Ecole Polytechnique, Institut Polytechnique de Paris, 5 avenue Henry Le Chatelier 91120 PALAISEAU, France

[2] Detralytics, Paris, France.

1 Introduction

The adoption of cloud solutions spectacularly increased over the last decades, as pointed by several reports[1] (see also Naldi and Mastroeni [2016]). Many corporate nowadays strongly rely on cloud capacities to perform their everyday activities, not only in the tech sector. Cloud usages are multiple: cloud computing allowing to process complex code on distant servers, while storage solutions allowing to access and share data efficiently within an organization, see for example Hsu [2022]. Security is also an expected outcome of this evolution, since the solution provider is supposed to be better armed to guarantee the permanence of the data, due to its expertise in the field. But no solution is perfectly secured (see Bisong et al. [2011]), and despite the apparent immateriality of cloud solutions, one should not forget that they rely on servers that may stop working, be physically destroyed, or even hacked. Among some recent incidents, one may mention the burning of some servers of OVH in 2021[2]. Although purely accidental, this incident shows the potential weaknesses of the sector. In this precise case, the existence of backups did not even protect some customers, since some of these backups were stored in the same building that burnt[3].

From a cyber insurance perspective, the scenario of a massive cloud failure is threatening the viability of the sector. Such an event could instantly generate an important amount of claims among a large number of policyholders relying on the same public cloud solution. With no more independence between the policyholders, the mutualization principle, which is at the core of the insurance business, is broken. This is one of the reasons explaining why the cloud outage scenario is not only a question of national crisis management[4], but also mentioned as one of the scenarios that require attention in performing cyber stress testing in a report from the European Insurance and Occupational Pensions Authority (EIOPA), see EIOPA [2023]. The projection of the total impact of such an episode on the economy made by Eling et al. [2023] also demonstrates the huge losses that may be generated by such an incident.

The present paper aims at providing a generic framework to model the impact of a cloud outage (or more generally of a critical infrastructure failure), and to measure

[1] See for example Flexera's "State of the cloud" report, `https://info.flexera.com/CM-REPORT-State-of-the-Cloud-2024-Thanks?revisit`

[2] see `https://w.media/what-was-the-possible-cause-of-ovhcloud-data-centre-fire-in-2021/`

[3] see `https://www.transatlantic-lawyer.com/ovh-must-pay-more-than-400000-e-after-a-fire-destroyed-it`

[4] see `https://www.dhs.gov/publication/dhs-resilience-framework`, section 4.

the exposure of a portfolio to such a catastrophe. The focus is not put on pricing (see Mastroeni and Naldi [2017] for an example of pricing methodology applied to the specific insurance of cloud outages and Mastroeni et al. [2022] for the pricing of related CatBonds), but on defining a measure of the exposition to such a systemic[5] risk. The main focus of this work is indeed to look at what the EIOPA defines as a "cyber underwriting risk", where the difficulty is to develop a collective vision, at the scale of a group of exposed entities, and not only at an individual level. At an individual level, the question of evaluating the impact of the disruption of a cloud service has been developed by several authors, see for example Abualkishik et al. [2020]. This vision essentially assumes a precise knowledge of the information systems of the victim. Such kind of level of knowledge is much harder to achieve at a collective level, restraining feasible approaches to rely on macroscopic data to calibrate the models.

Moreover, a collective vision naturally raises the question of diversification. Clearly, the ability to absorb a massive event will be weakened if all members of the portfolio are dependent from the same service experiencing a failure. In climate insurance, underwriting rules are established to ensure a sufficiently high level of diversification: in flood insurance, an insurance company will pay attention not to ensure the whole neighborhood of a given river, see Aerts et al. [2008], Denaro et al. [2020]. In the case of insurance against a cloud failure, one of the expected outcome of the present paper is to produce guidelines to underwriters to achieve a composition of the portfolio that is compatible with diversification requirements, by ensuring that the proportion of policyholders relying on the different available cloud solutions is sufficiently balanced. This requires to define a proper measure of diversification, which is adapted to the present context.

The present paper aims at modeling the consequences of a cloud failure, and formalizing this notion of diversification in this context of exposition to a massive event. The idea is to distinguish two regimes:

- first a situation when isolated claims happen. This is the classical insurance framework where mutualization operates due to the independence (or the weak form of dependence) between the policyholders. This case is the situation from which the insurance company is expecting to generate some positive results;

- a crisis regime, where a large part of the portfolio simultaneously collapses because of

[5]Let us note that the word "systemic" is not used here in the sense of financial systemic risk, but in a broader sense. Here, a systemic event may not destroy the whole insurance system, but could endanger significantly the survival of a company.

the interruption of a common shared solution. This stress-test framework measures the ability of the portfolio to resist to such catastrophic event.

Two risk measures are combined to define the resilience of the portfolio in both regimes, leading to a simple quadratic criterion to optimize to reach the appropriate level of diversification. The idea is to provide a criterion that can be used to define underwriting guidelines that will contribute to better securing the portfolio. It is also a way to measure the danger of concentration of the digital solutions into a too small number of tools.

The rest of the paper is organized as follows. In section 2, we introduce the modeling of the cyber insurance portfolio, distinguishing between the two regimes (standard one and catastophic one). The risk measures used to quantify the exposure of the portfolio to cloud failure are exposed in section 3, leading to the definition of a criterion to optimize to improve diversification. A practical example of application of the method is given in section 4.

2 Cloud outage model

In this section, we introduce the model used to describe the loss caused by cloud incidents at a portfolio level. We first explain, in section 2.1, the general framework that is used to describe the total loss at a portfolio level. From an analysis of the consequences of a cloud failure for the victim, we explain how to model the severity of the episode, first via the modeling of the timeline of the attack (section 2.1.1), then the different losses experienced throughout the different stages of the attack (section 2.1.2). Sections 2.2 and 2.3 are respectively devoted to the analysis of the behavior in a "standard regime", and in a stressed regime corresponding to a systemic failure of a cloud provider.

2.1 Describing the loss: portfolio and individual level

Let us consider that we have $j = 1, \cdots, d$ cloud providers, and a portfolio of $i = 1, \cdots, n$ policyholders. We introduce a binary random variable $\delta_{i,j}$ indicating if policyholder i suffered from a defect from provider j. The damage experienced at a portfolio level is then

$$\mathfrak{T}_s = \sum_{i=1}^{n} \sum_{j=1}^{d} \delta_{i,j} w_{i,j} \tau_i L_i^{(j)},$$

where $w_{i,j} \in [0, 1]$ denotes the exposure of policyholder i to cloud provider j (that is the proportion of its cloud dependent business activities that are relying on j, we assume

in the following that $\sum_{j=1}^{d} w_{i,j} = 1$), τ_i denotes the turnover of policyholder i, and $L_i^{(j)}$ denotes a unitary cost of the attack on j that generated loss for i. The variables $L_i^{(j)}$ are supposed to have the same distribution as a variable $L^{(j)}$. Behind this representation of the loss, there is the idea that the size of the company (here understood as its turnover) is a scale factor for the loss.

Regarding the indicator of an incident $\delta_{i,j}$, we will be distinguishing two cases in the following:

- a so-called "standard regime" where $\delta_{i,j}$ is stochastic (see section 2.2);
- a stressed regime, that is corresponding to a "systemic" event where all users of j are simultaneously affected corresponding to a crisis where $\delta_{i,j} = 1$ for all i.

To model the loss experienced in case of a cloud failure (whether it is an isolated or systemic event), we first discuss the timeline of consequences of such an event for the victim. Then we consider how these different steps affect the loss.

2.1.1 Timeline of a business interruption

We consider the case where the user i experiences a cloud outage of provider j, that is a total interruption of its services for a duration $T_i^{(j)}$. In the case of non-isolated attacks (stressed regime where $\delta_{i,j} = 1$ for all i), $T_i^{(j)} = T^{(j)}$ for all i , otherwise the durations for policyholders i and i' with $i \neq i'$ may be considered as independent. We characterize the timeline of the attack by three variables:

1. the initial duration of the unavailability of the cloud service $T_i^{(j)}$. This duration is common to every entity that we consider in the portfolio in the case of a non-isolated attack;
2. the time U_i after which the entity is able to react via the introduction of a backup plan. This backup plan does not fully replace the functions of the cloud servers, but will allow to restart a certain level of activity, hence reduces the loss. Let us also note that this backup plan may not exist (in which case, $U_i = \infty$), or may come too late if U_i is larger than the total duration of the crisis;
3. the additional time V_i after which the activity restarts at full regime after restoration of the service. Indeed, restoration of the cloud service does not mean immediate recovery for the cloud user, who will usually need additional time before going back to normal. Hence, for a given entity, the total duration of the crisis is $T_i^{(j)} + V_i$.

This decomposition of the timeline of the outage is aligned with the approach considered in Lloyd's [2018]. In full generality, the distribution of the variables U_i and V_i may be different from a victim to another, and could also depend on the cloud provider affected by the outage. Covariates may be introduced to materialize this heterogeneity. However, to simplify the notations and discussion, we will here assume that the policyholders are "identical", in the sense that $(U_i)_{1\leq i\leq n}$ and $(V_i)_{1\leq i\leq n}$ are identically distributed random vectors. We also assume that the policyholders are independent, and that $T_i^{(j)}$ is independent from U_i and V_i.

2.1.2 Loss experienced by the policyholder

Lloyd's [2018] introduces a rough way to go from a time of service interruption to the value of the loss experienced by an entity victim of a cloud failure. The basic idea is to assume that 1 day of business interruption leads to a loss of 1/365 times the annual turnover. The same report also suggests to restrict the turnover to the online activities of the company. This restriction may not be legitimate: although a cyber attack disrupting the information systems has direct consequences on online activities, it may also disrupt even physical dimensions of the business. Indeed a factory may stop working because of a loss of access to some digital applications. But we can keep the idea that it may seem reasonable to consider that the loss is proportional to the turnover τ_i, and to the duration of inactivity. Also, as in section 2.2, we keep the idea that the magnitude of the loss for policyholder i is also proportional to the exposure of this policyholder to the attacked cloud provider j (that is $w_{i,j}$, the proportion of cloud activities of i depending on j).

Following this idea, we propose to model the loss experienced by entity i as $w_{i,j}\tau_i L_i^{(j)}$, where

$$L_i^{(j)} = \mathfrak{a}_{i,j}(T_i^{(j)} - (T_i^{(j)} - U_i)_+(1-\mathfrak{b}_{i,j})) + \frac{\mathfrak{a}_{i,j}\left\{1-(1-\mathfrak{b}_{i,j})\epsilon_i^{(j)}\right\}V_i}{2}, \tag{2.1}$$

where we introduce two random variables $(\mathfrak{a}_{i,j}, \mathfrak{b}_{i,j})$ that are specific to this entity and to the cloud provider, and $\epsilon_i^{(j)} = \mathbf{1}_{U_i\leq T_i^{(j)}}$ is an indicator function equal to 1 if the backup plan manage to activate before the end of service interruption. This form of the loss is the superposition of the effects of the three periods described above:

- first, the entity suffers a loss which is $\mathfrak{a}_{i,j}$ per unit of time, for a duration $T_i^{(j)} - (T_i^{(j)} - U_i)_+$;
- then, during $(T_i^{(j)} - U_i)_+$, the loss is reduced to $\mathfrak{a}_{i,j}\mathfrak{b}_{i,j}$;

- finally, service is back, and we assume a continuous and homogeneous diminution of the loss, decreasing from $\mathfrak{a}_{i,j}$ (resp. $\mathfrak{a}_{i,j}\mathfrak{b}_{i,j}$) to 0 if the backup plan failed (resp. managed) to activate, in a period of time of length V_i, leading to the last term in (2.1).

The variable $\mathfrak{a}_{i,j}$ materializes the cost of a day of service interruption for entity i (in absence of an active backup plan) when the cloud provider j is stroke, and we assume that this cost stays constant through time. Making this cost depend on j stresses the fact that the cloud provider do not necessarily supply the same services: some may have a higher degree of criticality that reflects in the fact that $\mathfrak{a}_{i,j}$ is then higher. The variable $\mathfrak{b}_{i,j} \in [0,1]$ measures the efficiency of the back-up plan developed by i. We assume in the following that $(\mathfrak{a}_{i,j}, \mathfrak{b}_{i,j})_{1\leq i\leq n}$ are independent, and that, for all j, $(\mathfrak{a}_{i,j}, \mathfrak{b}_{i,j})_{1\leq i\leq n}$ have the same distribution.

The distribution of $\mathfrak{a}_{i,j}$ is a way to represent the heterogeneity of impacts of business interruption caused by the cloud, while a deterministic approach would give a rougher approximation by not allowing fluctuations. In the same spirit, the distribution of $\mathfrak{b}_{i,j}$ is a way to take into account the variable quality of crisis management.

We assume that $(U_i, V_i, \mathfrak{a}_{i,j}, \mathfrak{b}_{i,j})_{1\leq i\leq n}$ is an i.i.d. random vector independent from $T_i^{(j)}$, and that its four components are also mutually independent. It is of course possible to relax this last assumption of independence between U_i, V_i, $\mathfrak{a}_{i,j}$ and $\mathfrak{b}_{i,j}$ (for example, one can assume that if the backup plan started soon enough, this has implications on the value of V_i that should be shorter), but this of course increase the complexity of the preliminary statistical analysis required to calibrate the model.

According to this model, we see that the differentiation between the standard regime and the stressed one is solely characterized by the fact that, during a crisis, all the policyholders share the same duration of attack $T^{(j)}$ (and are all affected by the event). One could of course generalize the model by considering a specific loss distribution for systemic event compared to standard ones.

2.2 Profitability of the portfolio in the standard regime

We first consider the standard case where we only have "isolated" claims. Using the term "isolated", we do not necessarily mean that there is no correlation between policyholders and/or incidents, but simply that we are working in the classical framework where mutualization is valid. In this framework, optimizing the portfolio composition is a matter

of optimizing the profitability. To simplify the model, we will consider that every policyholder pays a premium that is proportional to its expected loss (under the standard regime) with the same loading factor. More precisely, this means that if a policyholder i generates a loss X_i, the premium is $(1+\theta)E[X_i]$, with the same value of θ. In this framework, the higher average value corresponds to the higher expected return (but the variance can of course give a more mixed picture). The approach that we consider can be extended to the case where the loading factors are not the same for all individuals, also introducing more evolved pricing schemes taking correlation into account as in Mastroeni et al. [2023]. In the present case, we shall keep the idea that imposing a constraint on the expected loss of the portfolio is equivalent to imposing a constraint on the average profitability.

In the following, we assume that the probability of failure for provider j is $\mathbb{P}(\delta_{i,j} = 1) = p_j$, that is only depends on the service j, and not on the policyholder's use. In this framework, the policyholder i can experience at most one incident from provider j in the time period that is considered. However, the model can be generalized by considering that $w_{i,j}\tau_i L_i^{(j)}$ is the aggregated amount of all losses generated by j for i.

Assuming that $\delta_{i,j}$ and $L_i^{(j)}$ are independent, the average loss of the portfolio becomes

$$\sum_{j=1}^{d}\sum_{i=1}^{n} p_j m_j w_{i,j}\tau_i = \sum_{j=1}^{d} \pi_j \bar{w}_j = \boldsymbol{\pi}'\bar{\mathbf{w}} = \mathfrak{L}_s(\bar{\mathbf{w}}), \tag{2.2}$$

where $m_j = \mathbb{E}(L^{(j)})$, $\pi_j = p_j m_j$, $\boldsymbol{\pi} = (\pi_j)'_{j=1,\cdots,d}$, and $\bar{w}_j = \sum_{i=1}^{n} w_{i,j}\tau_i$, and $\bar{\mathbf{w}} = (\bar{w}_j)'_{1\le j\le d}$ (with $\mathbf{w}'$ denoting the transposed vector of $\mathbf{w}$). Clearly we have $\sum_{j=1}^{d}\bar{w}_j = \sum_{i=1}^{n}\tau_i$ (which can be considered as the aggregated turnover of the portfolio), and $\bar{w}_j$ can be understood as the exposure of the insurance company (as an aggregator of individual risks) to cloud provider j.

Of course, if the target were to minimize the average loss, the most favorable case would be a portfolio of policyholders with 100% exposure to the cloud provider with the highest average profitability. But in such a situation, the insurance company becomes too exposed to a systemic default of this provider, and, even in a standard regime (meaning: outside of a systemic event), the variance of the result may be too high. Following the classical quadratic vision of modern portfolio theory as defined by Markowitz (see Markowitz [1991]), the idea is then to control the variance.

At a portfolio level, the variance of the loss is

$$\sum_{j,k=1}^{d}\sum_{i_1,i_2=1}^{n} w_{i_1,j}\tau_{i_1} Cov(\delta_{i_1,j}L_{i_1}^{(j)}, \delta_{i_2,k}L_{i_2}^{(k)}) w_{i_2,k}\tau_{i_2},$$

where $Cov(X, Y)$ denotes the covariance between two random variables X and Y.

In the following, we make the assumption that

$$Cov(\delta_{i_1,j} L_{i_1}^{(j)}, \delta_{i_2,k} L_{i_2}^{(k)}) = \sigma_j \sigma_k \rho_{j,k}, \tag{2.3}$$

where

$$\sigma_j^2 = Var(\delta_{i,j} L_i^{(j)}) = p_j(E[L^{(j)2}] - p_j E[L^{(j)}]^2).$$

Typically, (2.3) means that the correlation between the costs of two incidents is driven by the cloud provider and that, on the other hand, the policyholders are uncorrelated otherwise than through the providers.

This assumption is relatively realistic if we consider that the size of the portfolio is large enough, and if we recall that we are in a "standard regime" where we exclude the existence of systemic incidents that would strike simultaneously a significant part of the portfolio. This last situation will be considered in a second step. For now, assumption (2.3) allows to simplify the writing of the variance of $\mathfrak{T}_s$, that is

$$\mathfrak{V}_s(\mathbf{w}) = \bar{\mathbf{w}}' \begin{pmatrix} \cdots & \cdots & \cdots \\ \cdots & \sigma_j \sigma_k \rho_{j,k} & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix} \bar{\mathbf{w}} = \bar{\mathbf{w}}' \Sigma \bar{\mathbf{w}}. \tag{2.4}$$

If we were not considering the case of a massive cloud failure, optimizing the portfolio would reduce to minimizing $\mathfrak{V}_s(\mathbf{w})$ subject to a constraint of average loss, that is on $\mathfrak{L}_s(\bar{\mathbf{w}})$. This framework corresponds to the classical portfolio theory framework of Markowitz, where the assets would be the different cloud providers, and we want determine the optimal proportion of the portfolio turnover to be (ideally) exposed to each of these providers.

This rough definition of an optimal portfolio in view of diversification of the risk is clearly not adapted to protecting this portfolio against systemic events, which is our aim. The modeling that we propose in the present section only covers standard situations, but is not sufficient to take into account the occurrence of a cyber crisis, in which a cloud provider may entirely fail during a significant amount of time. Therefore, we develop a specific modeling of these particular situations, which will add constraints to define the optimal portfolio.

2.3 Systemic risk measure in the stressed case

At a portfolio level, in case of an attack of the $j-$th cloud provider, the insurance company will suffer the aggregated loss $\mathbf{L}^{(j)} = \sum_{i=1}^{n} w_{i,j} \tau_i L_i^{(j)}$, where $L_i^{(j)}$ has been defined in section

2.1.2, and depends on the same duration of attack $T^{(j)}$, that applies to every policyholder i. In this section, our aim is to measure, in the eventuality of such an event, how resilient the portfolio can be.

The distribution of $\mathbf{L}^{(j)}$ may be difficult to get analytically. A possibility is to use Monte Carlo simulations. But this would lead to difficulties in understanding the impact of diversification of the portfolio: indeed the distribution of the loss would depend on a too large number of parameters (namely the proportions $w_{i,j}$ for each policyholder). On the other hand, our objective to analyze how diversified the portfolio is, requires a more macroscopic vision. Indeed, if we want to act to improve diversification of the portfolio, it will be difficult to act at an individual level (that is changing $w_{i,j}$ for all i, or recruiting new policyholders with a specific value of $w_{i,j}$).

Therefore, we would like, ideally, a measure of the risk that only depends on $\bar{w}_j$, say $\mu(\bar{w}_j)$, since $\bar{w}_j$ represents the part of the turnover exposed to failure of the cloud provider j in the portfolio. From an underwriting perspective, we claim that is then easier to act on this allocation, by trying to optimize this issue while acquiring new business.

To build such kind of measure, a solution is to approximate the distribution using asymptotic theory, assuming that n is sufficiently large. This approximation exhibits the impact of $\bar{w}_j$, which can be understood as an aggregated level of exposure, on which it is much easier to act.

Let

$$s_w^2 = \sum_{i=1}^{n} \left(\frac{w_{i,j}\tau_i}{\bar{w}_j} \right)^2.$$

Introducing

$$\begin{aligned} m_t &= E\left[L_i^{(j)} | T^{(j)} = t \right], \\ \sigma_t^2 &= Var(L_i^{(j)} | T^{(j)} = t), \end{aligned}$$

one can approximate (for n sufficiently large) the distribution of $\mathbf{L}^{(j)}/\bar{w}_j$ by the one of a random variable $\mathcal{L}^{(j)}$ such that

$$\mathcal{S}^{(j)}(l) = \mathbb{P}\left(\mathcal{L}^{(j)} \geq l \right) = \int_0^\infty \bar{\Phi}\left(\frac{l - m_t}{s_w \sigma_t} \right) d\mathbb{P}_j(t) = \mathfrak{f}(s_w, l), \tag{2.5}$$

where $\mathbb{P}_j$ denotes the distribution of $T^{(j)}$. The proof of this approximation can be found in section 6.

The interesting feature of this approximation is that we essentially reduce the distribution of $\mathbf{L}^{(j)}$ to $\bar{w}_j$ and s_w. To measure how much a portfolio is exposed to the risk of massive

losses generated by a cloud provider attack, it is natural to use some characteristics of the distribution of $\mathbf{L}^{(j)}$.

Based on this approximation, we then discuss three possible choices in the following:

1. Expectation of $\mathbf{L}^{(j)}$: this quantity is simply $\bar{w}_j E[m_T] = \bar{w}_j \int_0^\infty m_t d\mathbb{P}_j(t)$, which is already of the form $\mu(\bar{w}_j)$. Using this measure to identify the weakness of the portfolio in the context of a provider failure presents the disadvantage to focus on the average loss scenario in case of such a catastrophe. The next two approaches allow to consider more pessimistic issues.

2. Quantile of $\mathbf{L}^{(j)}$ (or Value-at-Risk): let q_{α_j} denote the α_j-upper quantile of $\mathbf{L}^{(j)}$, defined as the solution of $\mathbb{P}(\mathbf{L}^{(j)} \geq q_{\alpha_j}) = \alpha_j$, where $\alpha_j \in (0,1)$ (and essentially, α_j is close to zero). Focusing on q_{α_j} allows to consider more pessimistic scenarios, compared to the central version offered by the expectation. If we use the approximation (2.5), we get

$$q_{\alpha_j} \approx \bar{w}_j \mathcal{S}^{(j)-1}(\alpha_j).$$

   This expression is not of the proper form, since, from (2.5), we have $\bar{w}_j \mathcal{S}^{(j)-1}(\alpha_j) = \bar{w}_j \mathfrak{f}^{-1}(s_w, \alpha_j)$. The dependence in s_w does not fit with our need to have a representation of the risk measure only in terms of $\bar{w}_j$. Indeed, s_w can be seen as a "micro-level" measure, since it measures the variation of weights among members of the portfolio. Nevertheless, it is easy to check that $s_w \to \mathfrak{f}^{-1}(s_w, \alpha_j)$ is a non-decreasing function. Hence, if we consider that $s_w \leq s$, then we can use

$$\mu(\bar{w}_j) = \bar{w}_j \mathfrak{f}^{-1}(s, \alpha_j),$$

   as an upper bound of the previous quantity that does not depend on s_w. Regarding the constraint $s_w \leq s$, it can be added to the set of constraints (see section 3.1), or relaxed as we will do in the application part, arguing that, if s large enough, the condition may be easily fulfilled.

3. CTE of $\mathbf{L}^{(j)}$: among risk measures, the Conditional Tail Expectation (CTE) is sometimes prefered to the quantile since it provides information on the magnitude of the excess when the loss is larger than the quantile. The CTE is defined as

$$CTE(\alpha_j) = E\left[\mathbf{L}^{(j)} | \mathbf{L}^{(j)} \geq q_{\alpha_j}\right].$$

If we use the approximation (2.5), this CTE can be considered as proportional to $\bar{w}_j$, since

$$CTE(\alpha_j) \approx q_{\alpha_j} + \frac{\bar{w}_j}{\alpha_j} \int_{\mathcal{S}^{(j)-1}(\alpha_j)}^{\infty} \mathcal{S}^{(j)-1}(t)dt = \bar{w}_j \mathfrak{g}(s_w, \alpha_j).$$

Again, the function $s_w \to \mathfrak{g}(s_w, \alpha_j)$ is non-decreasing, pleading for introducing the measure $\mu(\bar{w}_j) = \bar{w}_j \mathfrak{g}(s, \alpha_j)$.

In view of these three examples, we will consider a risk measure of the type

$$(\mu(\bar{w}_1), \cdots, \mu(\bar{w}_d)) = \mathbf{m}'\bar{\mathbf{w}}, \tag{2.6}$$

to describe the exposure of the portfolio to stress scenarios corresponding to the different providers.

Remark 2.1 *We here rely on a Gaussian approximation of the losses. This approximation is legitimate if the number of policyholders is high enough, and if the distribution of the loss is not heavy tail (otherwise, high quantiles may be better approximated by Generalized Pareto distribution, see Mikosch and Nagaev [1998]). We here implicitly assume that the individual losses are moderate, and that the solvency of the portfolio is at risk more because of the number of victims than because of the value of a single claim. Relying on a Generalized Pareto distribution to approximate* $\mathbf{L}^{(j)}$ *is an alternative, but would not provide risk measures that are linear with respect to* $\bar{\mathbf{w}}$ *as in (2.6). Moreover, let us note that the probabilities* $(\alpha_j)_{j=1,\cdots,d}$ *may not necessarily need to be as close to zero as the* 0.005 *value used in the Solvency II regulation for the Value-at-Risk: we are here looking at stress scenarios, whose probabilities of occurrence are not quantified. The values* α_j *are essentially used to get a more or less conservative risk measure.*

Remark 2.2 *In this stressed approach, we consider, for each cloud provider, the consequences of a failure of his solution, striking all his users. An even more pessimistic approach would consist in the joint failure of several actors. The dependence between these actors is considered in the standard regime. In the stressed case, one can easily add scenarios like the joint failure of two cloud providers* j *and* k*, and a similar approach applies to the study of the resulting loss* $\mathbf{L}^{(j)} + \mathbf{L}^{(k)}$.

3 Diversification and optimal insurance portfolio

In section 2, we defined a way to measure the profitability and the risk associated to a portfolio, based on the distribution $\bar{\mathbf{w}}$ of the exposed capital between the different cloud

providers. The purpose of the present section is to determine an optimization problem that will allow to determine what is the best configuration in terms of $\bar{\mathbf{w}}$ if one wants to achieve enough diversification in both standard and stressed regimes. Section 3.1 is devoted to the definition of an optimization criterion that merges the constraints of section 2.2 (standard regime) and section 2.3 (stressed regime). This criterion depends on a parameter quantifying the relative importance given to the two situations, and section 3.2 aims at helping its practical calibration. In section 3.3, we compare the simple approach we develop to other diversification measures used in portfolio theory.

3.1 Optimization problem

In section 2.2 and 2.3, we defined two ways to measure the risk at a portfolio level. The first one corresponds to a "standard regime", where variance, as a quadratic function of $\bar{\mathbf{w}}$, is a way to measure uncertainty. On the other hand, in case of a systemic cloud failure, the risk measures we considered in section 2.3 can be approximated by $\mathbf{m}'\bar{\mathbf{w}}$, for some vector $\mathbf{m}$. We propose to mix these two ways to measure risk in a single criterion, say

$$\mathfrak{C}_\lambda(\bar{\mathbf{w}}) = \frac{1}{2}\bar{\mathbf{w}}'\Sigma\bar{\mathbf{w}} + \lambda\mathbf{m}'\bar{\mathbf{w}},$$

where $\lambda > 0$ is a parameter to be fixed by the user to quantify the comparative importance of the risk measure in the standard case (materialized by $\bar{\mathbf{w}}'\Sigma\bar{\mathbf{w}}$), and in the catastrophic case. We will propose, in section 3.2, a way to calibrate this parameter λ.

For a given total exposure W and an average loss expected to be ρ, finding the optimal portfolio is then a matter of quadratic programming under constraints, that is

$$\begin{aligned} Min_{\bar{\mathbf{w}}} \ & \frac{1}{2}\bar{\mathbf{w}}'\Sigma\bar{\mathbf{w}} + \lambda\mathbf{m}'\bar{\mathbf{w}}, \\ s.t. \ & \bar{\pi}'\mathbf{w} = \rho, \\ & \bar{\mathbf{w}} \geq 0, \\ & \sum_{j=1}^{d} \bar{w}_j = W. \end{aligned}$$

Since Σ is definite positive, the problem is a convex optimization problem, and has a unique solution if the polyhedra defined the constraints is nonempty. Many standard methods may be used to solve this classical problem, see Floudas and Visweswaran [1995].

3.2 Practical choice of the criterion

The criterion proposed in section 3.1 is designed to be simple: we want both the variance and the risk measure in case of catastrophe to be small, so we make a linear combination of the two objectives. However, the practical meaning of the parameter λ is relatively obscure. This parameter is used to describe the importance of one criterion with respect to the other, but these two criteria are not of the same nature: the quadratic part is a variance, while the second part, related to the systemic risk measure, corresponds to a quantile (in case of VaR) or to a conditional expectation (in case of CTE).

The aim of this section is to provide guidelines that will allow to consider a realistic value of λ which corresponds to the needs of the user.

To do so, let us first consider the standard regime. If the size of the portfolio is large enough and if the policyholders are independent, we can approximate the distribution of the random total loss of the portfolio by a Gaussian random variable, say $\mathfrak{T}_s \sim \mathcal{N}(\mathfrak{L}_s(\bar{\mathbf{w}}), \mathfrak{V}_s(\bar{\mathbf{w}}))$. This Gaussian approximation is reasonable if the size of the portfolio is large enough, and considering that the distributions of the individual losses are not heavy tailed. If these distributions are in fact heavy tailed, other kind of distributions may be used, see for example Mikosch and Nagaev [1998] for detailed result on collective risk models approximations. Staying, to simplify, in the Gaussian approximation case, if we consider an insurer planning to spend a maximal amount of capital $\mathfrak{c}$ to face the risk, the probability of this reserve to be insufficient can be approximated by

$$\bar{\Phi}\left(\frac{\mathfrak{c}-\mathfrak{L}_s(\bar{\mathbf{w}})}{\mathfrak{V}_s(\bar{\mathbf{w}})^{1/2}}\right), \tag{3.1}$$

where $\bar{\Phi}$ is the survival function of a $\mathcal{N}(0,1)$ random variable. Under the constraint that the expected loss is ρ (which is a constraint of the optimization problem), the probability of exceeding the reserve (3.1) will be less than a given value α provided that

$$\frac{1}{2}\mathfrak{V}_s(\bar{\mathbf{w}}) = \frac{1}{2}\bar{\mathbf{w}}'\Sigma\bar{\mathbf{w}} \leq \frac{(\mathfrak{c}-\rho)^2}{2\bar{\Phi}^{-1}(\alpha)}. \tag{3.2}$$

On the other hand, if we consider the risk measures of section 2.3, we see that they all correspond to an amount of capital (average loss in case of catastrophe, capital value corresponding to a high quantile in case of VaR, average value of the loss beyond this quantile in case of CTE). Let us consider the case where the user does not want these values to exceed $\mathfrak{c}$, where $\mathfrak{c}$ is the same limit value as the one use in the first situation where we were only considering the standard regime. This means that we expect, at the

optimal, to have $\mathbf{m}'\bar{\mathbf{w}} \leq d\mathfrak{c}$ (since we have d providers, and $\mathbf{m}'\bar{\mathbf{w}}$ corresponds to the sum of capital corresponding to failure scenarios for each of the providers.

An optimal portfolio for the program defined in section 3.1 is associated to a certain level of variance, and a certain value of the systemic risk measure. If we expect, for the optimal portfolio, to have a variance close to the constraint (3.2), and a value of the systemic risk also close to the limit $d\mathfrak{c}$, and if we give equal importance to both objectives, then, in $\mathfrak{C}_\lambda$, the quadratic part and the linear part should have the same importance at the optimum, so we expect to have

$$\lambda d\mathfrak{c} = \frac{(\mathfrak{c} - \rho)^2}{2\bar{\Phi}^{-1}(\alpha)}, \tag{3.3}$$

which is our way to set a reasonable value for λ. Of course, this rule may be substituted to any other rule that makes sense, and one may chose not to treat equally the risk measure for the standard case and for the systemic one.

3.3 Comparison with other diversification measures

We made the choice to study the standard regime using a quadratic approximation, then to study separately a stressed event, leading to a modification of the classical criterion used in modern portfolio theory. The introduction of this additional factor is motivated by the need to protect the portfolio in case of systemic episodes. The same motivation lead to other strategies in portfolio optimization, like the one based on optimizing the CTE (or Conditional Value at Risk) like in Angelelli et al. [2008] or Ray and Bhattacharyya [2017] (see also Cesarone et al. [2011]). In these approaches, the authors do not distinguish between two regimes, but directly model the tail of the distribution of the loss variables to compute the CTE. To use such type of approaches, one could for example approximate the tail of the loss of the portfolio, using extreme value theory results, see for example Mainik et al. [2015]. This method would present the advantage to model the loss variable as a whole.

However, it would have many operational drawbacks that are directly linked to our context of insurance and of lack of data on systemic cyber events: due to a poor statistical experience on extreme claims, historical data may not be sufficient to correctly estimate the tail of the distribution. Moreover, this distinction between two regimes corresponds relatively well to a distinction between catastrophes (that escape the framework of statistical modeling and whose probability is impossible to precisely evaluate, since referring

to potentially never encountered situations that are yet to come) and a situation where classical rules of insurance apply.

Let us also mention again that the complete optimization of the portfolio would require to determine the optimal value of $w_{i,j}$ for all i and j, while we only focus on $\bar{\mathbf{w}}$. Considering this would considerably increase the dimension of the problem, giving more importance to evolved optimization techniques (see for example Gunjan and Bhattacharyya [2023] for a review), while the optimization problem we solve is usually of small dimension (due to the small number of available providers) and only requires standard techniques. But let us also stress that this choice to simplify the optimization problem is motivated by the fact that we are studying an insurance portfolio and not a financial portfolio as in the classical situation of portfolio theory. This distinction is important, because our degrees of freedom are fewer, in insurance, compared to the problem of selecting financial asset. Here, the asset is a policyholder, and is completely free not to join the portfolio, even if we want him to. Considering optimization with respect to an aggregated indicator $\bar{\mathbf{w}}$ is more efficient for risk management to track the evolution of this distribution of cloud providers within the portfolio, and to establish simple subscription guidelines to improve it.

4 Empirical illustration

In this section, we illustrate on a practical example the modeling of the impact of cloud failures on a portfolio, and the behavior of the solution of the optimization problem we mentioned. We first describe in section 4.1 the distribution we consider to model cloud outage. In section 4.2, we describe the distribution of the cloud providers among the portfolio, the probability of failure and covariance structure that we use in our numerical analysis. Section 4.3 provides a discussion on the optimal portfolio and its evolution with respect to the input parameters of the model.

4.1 Distribution of the loss

In this illustration, we consider different type of distributions to describe the loss experienced by a victim following a cloud outage. Let us recall that, according to our model (2.1), we have 5 variables that we need to specify:

1. the time of service interruption $T_i^{(j)}$: we consider a Weibull distribution, that is $\mathbb{P}(T_i^{(j)} \geq t) = \exp(-(\gamma t)^a)$. For an average interruption of $\mathfrak{d}$ days, we consider

three values of a (0.5, 1 and 1.5) and set the value of γ to reach an expected time of interruption of 2 days (in the main case) or 5 days (for a comparison). This allows to compare situations where the average interruption time is the same, but where the hazard rate of the distribution is different (decreasing hazard rate for $a = 0.5$, constant for $a = 1$ which corresponds to the particular case of an exponential distribution, increasing for $a = 1.5$).

2. the time before the reaction (i.e. start of the back-up plan) of the victim U_i. Again we consider the three types of distributions above, with parameters so that the expectation is the same. We consider that there is a probability that the backup plan does not work. We consider $\mathbb{P}(U_i = \infty) = 0.2$. which corresponds to the value given by Lloyd's [2018] for companies with turnover above 1 billion USD (a more pessimistic value is assumed by Lloyd's [2018] for a turnover below 1 billion USD with a value of 0.5, we keep the optimistic vision in the present scenario). We set $E[U_i|U_i < \infty] = 2.1$ which again corresponds to the magnitude mentioned in Lloyd's [2018], considering that U_i has a Weibull distribution.

3. the distribution of $\mathfrak{a}_{i,j}$ which reflects the heterogeneity between policyholders. We will consider two cases: a deterministic case where $\mathfrak{a}_{i,j}$ is equal to a constant depending on j (see Table 1), and normal random variables with same means as in the deterministic case (and different values for the variance to modify the variability between individuals).

4. the distribution of $\mathfrak{b}_i$ is inspired again from Lloyd's [2018] and is taken as uniformly distributed between 0.3 and 0.5.

Let us note that, to simplify the discussion, we consider the same distribution for each cloud provider in each setting. In practice, this may not be the case: some cloud provider may operate on critical functions of the information systems while less important tasks may be left to others. This is where a careful analysis of the way these solutions are used among the policyholders could convey a precious additional information. Some market analysis, like for example Flexera [2023], could be used since they provide statistics on the type of workloads that are performed by the different cloud solutions, although these statistics may be biased compared to the specific population of the considered portfolio.

4.2 Probability of failure

We consider a situation where 5 cloud providers, named A to E[6], are available on the market. This small number is inspired by the report Flexera [2023] which shows that 5 companies (namely, by alphabetical order, AWS, Azure, Google Cloud, IBM Cloud, Oracle Cloud, but the letters A to E do not refer to any of them) shared more than 95% of the market in 2022.[7] As explained in section 4.1, we consider that the task performed by the different providers are not of the same level of criticality. Table 1 describes the value of the daily losses $\mathfrak{a}_{i,j}$ in the deterministic case (average losses in the random case).

Cloud provider j	Daily loss $\mathfrak{a}_{i,j}$
A	2.7
B	3.1
C	2.4
D	2.6
E	2.3

Table 1: Daily loss for the different cloud providers in the deterministic case. Recall that $\mathfrak{a}_{i,j}$ is the same for all i.

To describe the situation in the so-called "standard regime", we need to specify the probability of failure. Again, since we assumed a similar loss distribution for the providers in case of failure, the reliability of the different solutions only differ through this aspect. The probability of failure are gathered in Table 1.

Cloud provider	A	B	C	D	E
p_j	0.36	0.22	0.70	0.49	0.81

Table 2: Values taken for the probability of failure of the cloud providers.

Regarding the covariance matrix Σ that contains the information on the dependence between the cloud providers, its value depends on the distribution of the loss that varies from one setup to another. If $\Delta_{j,k} = E[\delta_{i,j}\delta_{i,k}]$, and $l_j = E[L_i^{(j)}]$, the coefficient $\mathfrak{s}_{j,k}$ (line

[6]With a slight inconsistency of notations, instead of j spanning from 1 to d, we use letters to denote the cloud providers, in order to distinguish more easily what refers to the index j (the cloud providers) and the index i (the individuals).

[7]Adoption almong the enterprises "running significant workloads" on these systems.

j, column k) of Σ is

$$\mathfrak{s}_{j,k} = l_j l_k(\Delta_{j,k} - p_j p_k).$$

The coefficients $\Delta_{j,k}$ are given in Table 3.

j \ k	1	2	3	4	5
1		0.36	0.47	0.43	0.38
2			0.40	0.34	0.32
3				0.49	0.52
4					0.39
5					

Table 3: Matrix Δ where the coefficients of line j and column k is $\Delta_{j,k}$. Since the matrix is symetric, we only show the upper part for readability. We also do not show the diagonal, since $\Delta_{j,j} = p_j$ can be deduced from Table 2.

4.3 Optimal portfolio

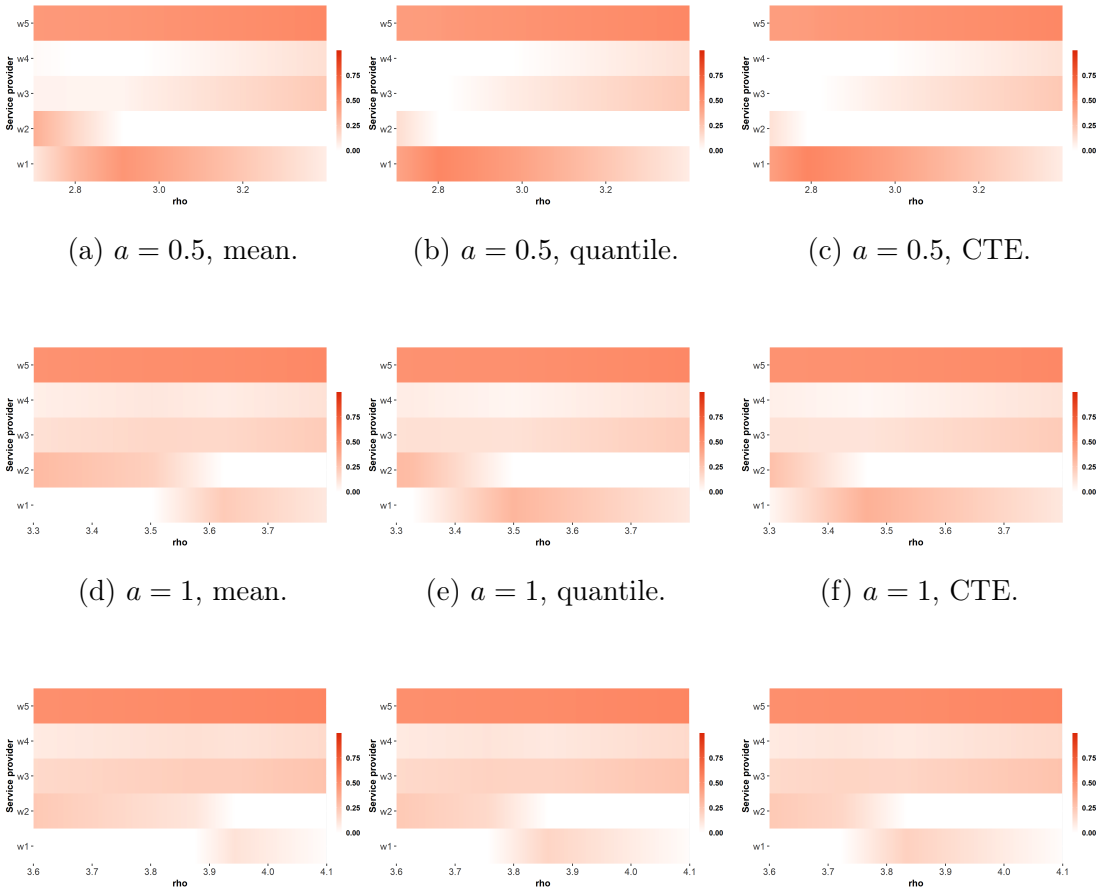
In this section, we take $\alpha = 0.95$, we consider the value of λ as in (3.3) with a value of the capital $\mathfrak{c} = 25$ and we make ρ to vary. We essentially consider the case where $\mathfrak{a}_{i,j}$ is deterministic, and then show in a second time how the impact of randomness of this factor changes the results. For the quantile and conditional tail expectation risk measures, we consider $\alpha_j = \alpha$ for all $j = 1, \cdots, 5$, and we make α vary. We explore in detail the case where the average unavailability of the service is 2 days (Figures 1 to 4). In Figure 5 and 6, we briefly show extension of these observations when this parameters varies by consider the case where it is equal to 5 days.

To obtain a few indications about the optimal portfolio and how it depends from the different types of parameters, we first consider in Figure 1 the impact of ρ, that is of the expected profitability in the standard regime. The choice of ρ also impacts the penalizing constant λ from (3.3).

Starting with a general remark, we see that $w_5 = w_E$ has the strongest value. This is expected from the setting that we considered: the probability of experiencing a failure is high for cloud provider 5, but this is also mechanically related to a higher expected profitability by construction.

With no surprise, the choice of the risk measure has impact on the composition of the optimal portfolio. However, we see that the pattern of dependency of this composition with respect to ρ is the same for the three risk measures for a given distribution of the time of unavailability. For example, in the case of an exponential distribution ($a = 1$), we see that $w_2 = w_B$ vanishes at some point, with, depending on the risk measure, a slight variation of the value of ρ after which one observes this effect.

On the other hand, the distribution of the time of interruption of service has a considerable impact on the structure of this optimal portfolio: in the case $a = 0.5$, we see that $w_3 = w_C$ is relatively low (approximately of the same magnitude as $w_4 = w_D$), but becomes much higher for $a = 1.5$ and $a = 1$.



(a) $a = 0.5$, mean. (b) $a = 0.5$, quantile. (c) $a = 0.5$, CTE.

(d) $a = 1$, mean. (e) $a = 1$, quantile. (f) $a = 1$, CTE.

(g) $a = 1.5$, mean. (h) $a = 1.5$, quantile. (i) $a = 1.5$, CTE.

Figure 1: Evolution of the weights of the optimal portfolio depending on the value of ρ. The average duration of the interruption is set to 2 days. The left column displays the results taking the mean as risk measure, the middle column shows the results for the quantile, and the right column shows the case of Conditional Tail Expectation.

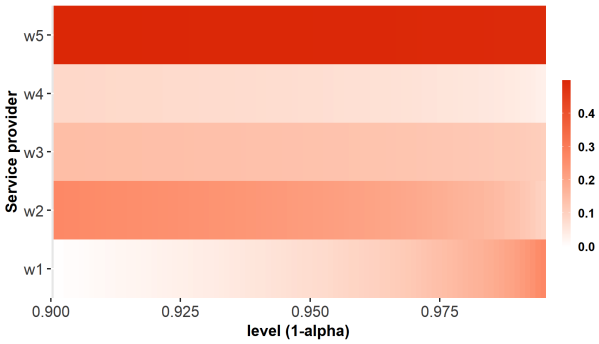
This impact of the distribution for the time of interruption is also observed when we consider the dependence of the results with respect to α in Figure 2 (this time only for quantile and CTE). For example, the value of w_2 is almost 0 for the case $a = 0.5$, while much higher in case of $a = 1$ and $a = 1.5$.

If the variables $\mathfrak{a}_{i,j}$ are random, the pattern is similar as one can see in the example of Figure 3, but we see some sensitivity with respect to the level of the noise. The most obvious case is related to the values of w_4 and w_5, where we see that, depending on ρ, there is a threshold after which provider E is prefered to provider D: this threshold tends to become higher when the variance of the noise (describing the heterogeneity of the population) increases.

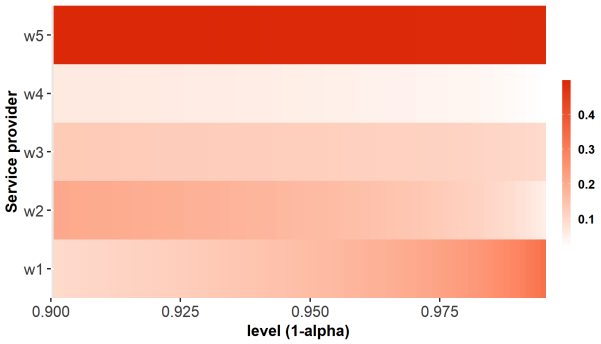
We next consider the impact of the penalizing constant λ on the portfolio, focusing on the specific value of one of the weights (to avoid redundancy, we only show results for w_4 and for the CTE). In this case, we do not necessarily consider the case where λ is defined by (3.3), but make the penalizing constant vary around $\lambda = 1$. Note that if (3.3) is used, the value $\lambda = 1$ corresponds to $\alpha = 0.95$, $\mathfrak{c} = 25$ and $\rho = 3.37, 3.37$ or 3.77 for $a = 0.5, 1$ or 1.5 respectively. Here, the three distributions provide the same shape but with a lag on ρ.

If we change the expectation of the variable T, that is the average duration of the episode of unavailability of the cloud service, the composition of the portfolio evolve. In Figure 5 and 6, we show this evolution for the case of an exponential duration $a = 1$. In Figure 5, we see that, in this configuration, the main difference is that $w_1 = w_A$ becomes very weak compared to w_2, while there was, in a case with a lower average time of interruption, a switch from w_2 to w_1 when ρ increases. We see a similar phenomenon in Figure 6. Additionally, we also see that the proportion w_3 tends to increase compared to w_4.

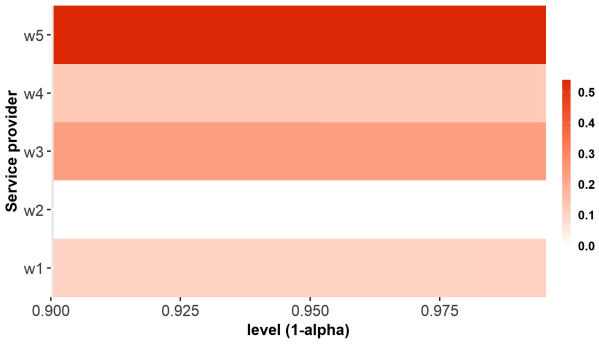
Of course, the setting that we propose here is really simplified: first of all, we assume that the average profitability is proportional to the probability to have a claim, which creates an incentive for taking higher risk. Although the losses, in case of an attack may not be the same from a cloud provider to another, since these providers may be used to perform different type of tasks: some cloud provider are specialized in storage, other in computation, and some are more trusted than others to manipulate sensitive information or vital tasks. The model can of course be easily modified to take these aspects into account.



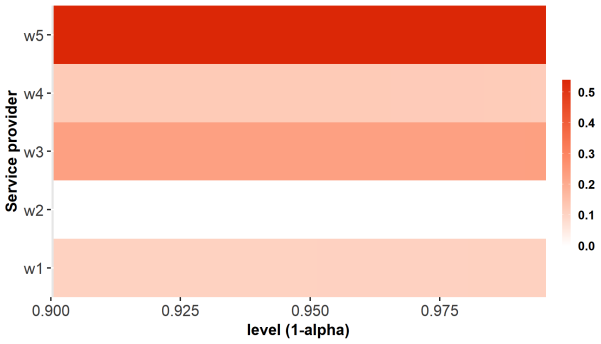
(a) $a = 1$, quantile.



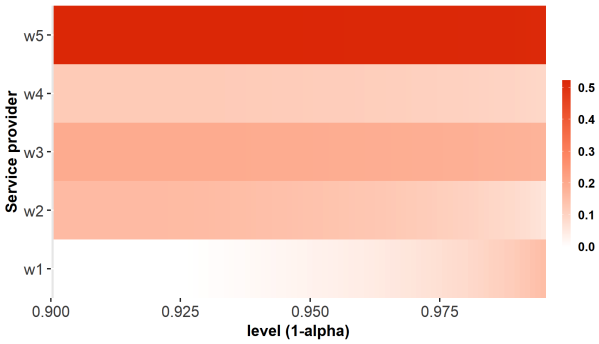
(b) $a = 1$, CTE.



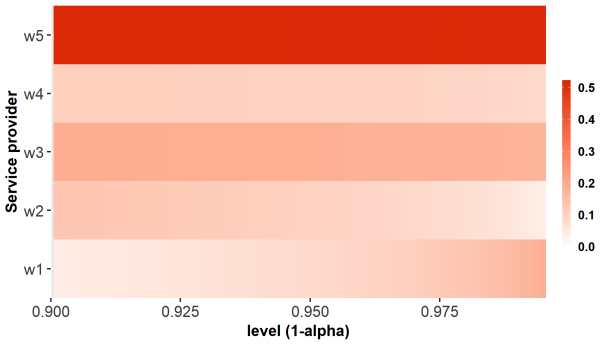
(c) $a = 0.5$, quantile.



(d) $a = 0.5$, CTE.

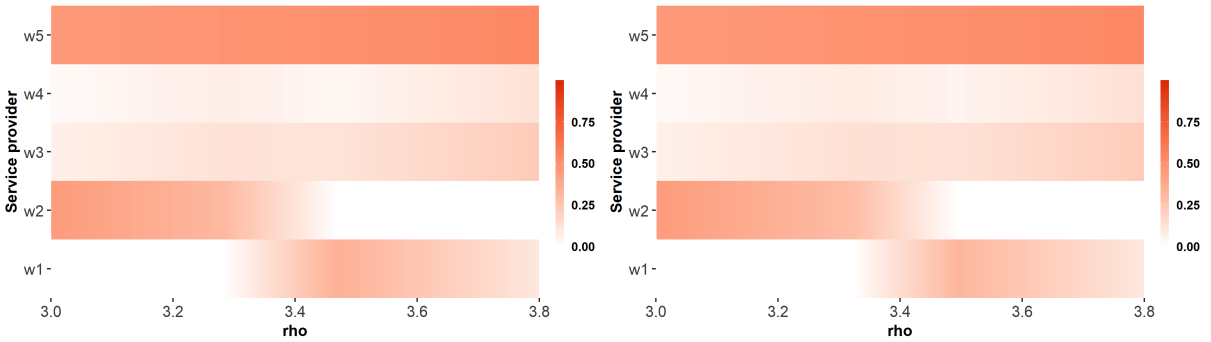


(e) $a = 1.5$, quantile.



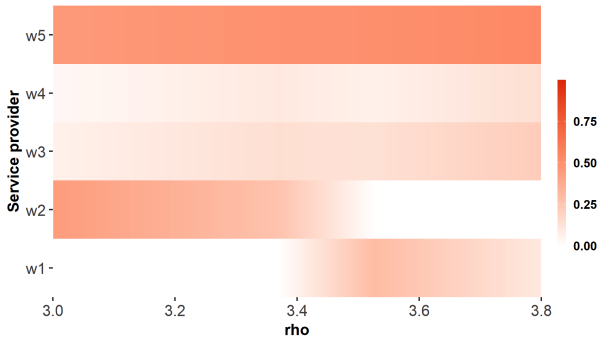
(f) $a = 1.5$, CTE.

Figure 2: Evolution of the weights of the optimal portfolio depending on the value of α_j (which is taken as equal for $j = 1, \cdots, 5$. The average duration of the interruption is set to 2 days. The left column displays the results taking the quantile as risk measure, the right column shows the case of Conditional Tail Expectation.



(a) Standard deviation 0.1.

(b) Standard deviation 0.75.



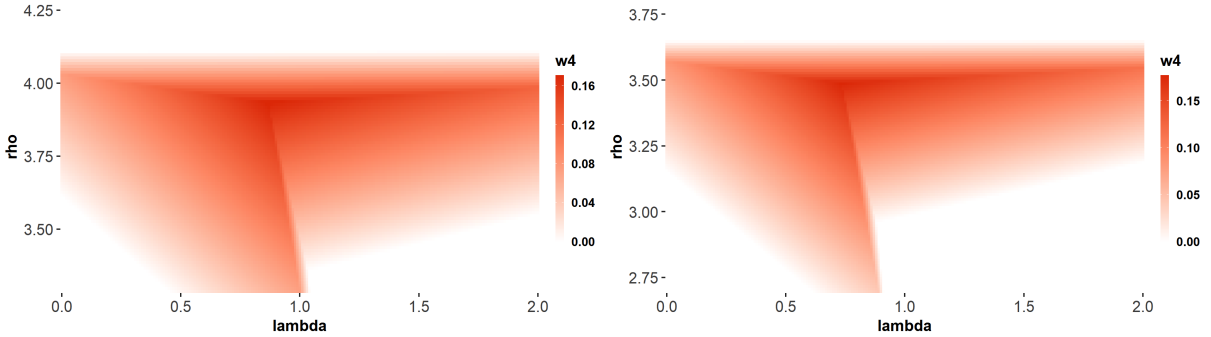
(c) Standard deviation 1.4.

Figure 3: Evolution of the weights of the optimal portfolio depending on the value of ρ in the case where $a = 1$ (exponential distribution), the risk measure is CTE, and with a random value of $\mathfrak{a}_{i,j}$. The distribution of $\mathfrak{a}_{i,j}$ is Gaussian with mean defined as in Table 1, and different values of the standard deviation are considered.

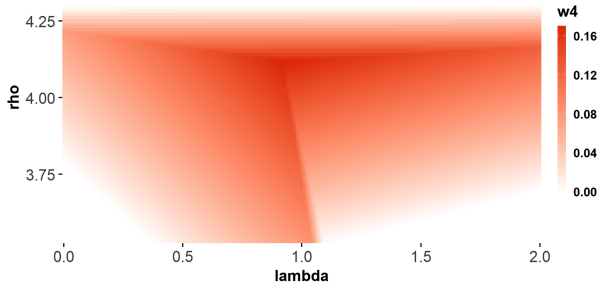
5 Conclusion

Cloud interruption is one of the most concerning scenarios when looking at the collective risk management of cyber risk. It can especially destabilize an insurance portfolio, and is identified as one of the situation that should be modeled by EIOPA in its guidelines on cyber risk stress testing. Measuring the consequences of such an episode first requires to understand what are the consequences of such a disruption. We propose a framework, inspired by Lloyd's [2018], that studies the different step of business interruption, and shows the different quantity that need to be modeled to assess this impact.

We would like to stress that the consequences may be very different for two policy-



(a) $a = 1$. (b) $a = 0.5$.

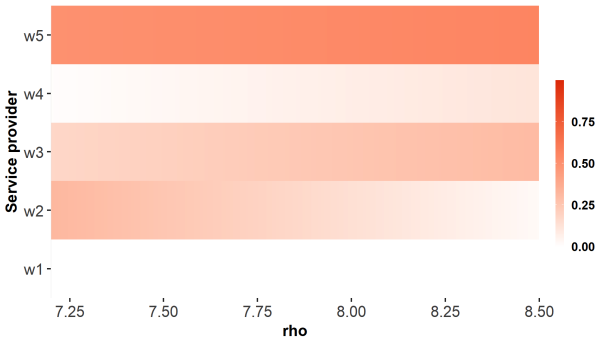


(c) $a = 1.5$.

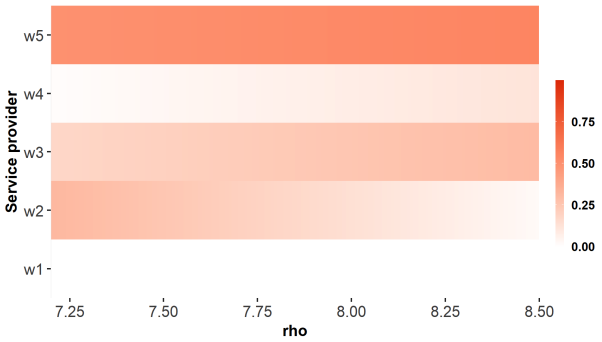
Figure 4: Evolution of the weight w_4 using the CTE as risk measure.

holders in two different industrial sectors: depending on the nature of the victim, time may not have the same value, since some activities may differ operations with low damage, while others are not able to. Therefore, there is a specific analysis that should be developed to study the link between various characteristics of the policyholder and the cost of one day business interruption. This is not the objective of this paper, but the model is easy to adapt to this situation.

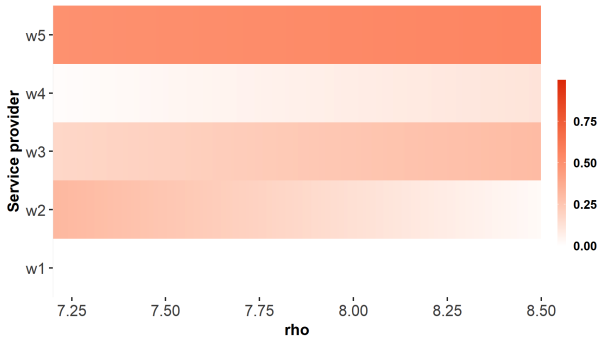
Our focus was to propose a way to measure the exposure of the portfolio, but also to put it in perspective with the expected profitability. The metric we develop to optimize the portfolio is a combination of a measure of profitability in a "standard regime", and a stressed one. These two regimes have deliberately been distinguished in modeling the behavior of the portfolio. The measure that is developed can be used to compare two portfolios, but is also designed to provide guidelines to underwriters: through the simple optimization procedure that we propose, it is possible to identify an ideal composition,



(a) $a = 1$, mean.



(b) $a = 1$, quantile.

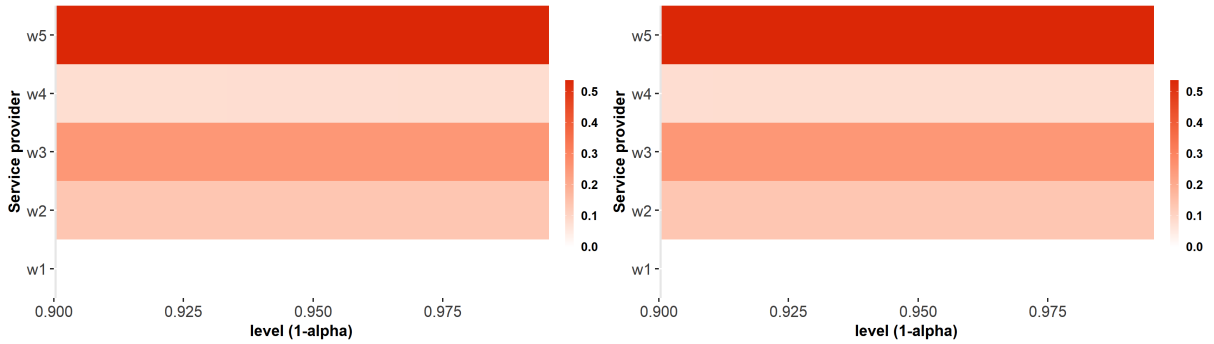


(c) $a = 1.5$, conditional tail expectation.

Figure 5: Evolution of the weights of the optimal portfolio depending on the value of ρ. The average duration of the interruption is set to 5 days. Only the results regarding the exponential distribution ($a = 1$) are reported for the different risk measures.

which can be considered as a target for underwriters. Of course, moving towards this ideal composition is not automatic, since, in insurance, the "assets" - i.e. the policyholder - have to be convinced to join the portfolio.

The illustration we provided show the sensitivity to the model assumptions (essentially the distribution of the time of interruption). To calibrate it in such a catastrophic scenario, historical data may be of help but needs to be supplemented with expertise (from IT specialists, but also from threat intelligence since the nature and motivation of the attackers may have impact on the duration of the phenomenon).



(a) $a = 1$, quantile.

(b) $a = 1$, CTE.

Figure 6: Evolution of the weights of the optimal portfolio depending on the value of α_j (which is taken as equal for $j = 1, \cdots, 5$). The average duration of the interruption is set to 5 days. Only the results regarding the exponential distribution ($a = 1$) are reported for the different risk measures.

6 Appendix: Proof of (2.5)

In case of a systemic event impacting the cloud provider j, $T_i^{(j)}$ is the same for all i, that $T_i^{(j)} = T^{(j)}$. On the event $\{T^{(j)} = t\}$, the portfolio loss $\mathbf{L}^{(j)}$ can be rewritten as

$$\frac{\mathbf{L}^{(j)}}{\bar{w}_j} = \sum_{i=1}^{n} \frac{w_{i,j}\tau_i}{\bar{w}_j} \varphi_t(\mathbf{Z}_i),$$

where $\mathbf{Z}_i = (\alpha_i, \beta_i, U_i, V_i)$. $(\mathbf{Z}_i)_{1 \leq i \leq n}$ are i.i.d. and independent from $T^{(j)}$ by assumption. On the other hand, $E[\varphi_t(Z_i)] = m_t$, and $Var(\mathbf{L}^{(j)}/\bar{w}_j) = s_w^2 \sigma_t^2$.

From an asymptotic point of view, if

$$\sum_{i=1}^{n} \left(\frac{w_{i,j}\tau_i}{\bar{w}_j} \right)^4 < \infty, \tag{6.1}$$

the weighted Central Limit Theorem applies (see Weber [2006]), and

$$\mathbb{P}\left(\frac{\mathbf{L}^{(j)}}{\bar{w}_j} \geq l | T^{(j)} = t \right) \sim_{n \to \infty} \bar{\Phi}\left(\frac{l - m_t}{s_w \sigma_t} \right).$$

The approximation (2.5) then comes from the fact that

$$\mathbb{P}\left(\frac{\mathbf{L}^{(j)}}{\bar{w}_j} \geq l \right) = \int_0^\infty \mathbb{P}\left(\frac{\mathbf{L}^{(j)}}{\bar{w}_j} \geq l | T^{(j)} = t \right) d\mathbb{P}_j(t).$$

Let us note that condition (6.1) may be weakened, see Theorem 1 in Weber [2006].

References

A. Z. Abualkishik, A. A. Alwan, and Y. Gulzar. Disaster recovery in cloud computing systems: An overview. *International Journal of Advanced Computer Science and Applications*, 11(9), 2020.

J. C. Aerts, W. Botzen, A. van der Veen, J. Krywkow, and S. Werners. Dealing with uncertainty in flood management through diversification. *Ecology and Society*, 13(1), 2008.

E. Angelelli, R. Mansini, and M. G. Speranza. A comparison of mad and cvar models with real features. *Journal of Banking & Finance*, 32(7):1188–1197, 2008.

A. Bisong, M. Rahman, et al. An overview of the security concerns in enterprise cloud computing. *arXiv preprint arXiv:1101.5613*, 2011.

F. Cesarone, A. Scozzari, and F. Tardella. Portfolio selection problems in practice: a comparison between linear and quadratic optimization models. *arXiv preprint arXiv:1105.3594*, 2011.

S. Denaro, A. Castelletti, M. Giuliani, and G. Characklis. Insurance portfolio diversification through bundling for competing agents exposed to uncorrelated drought and flood risks. *Water Resources Research*, 56(5):e2019WR026443, 2020.

EIOPA. Methodological principles of insurance stress testing - cyber component, 2023. URL `https://www.eiopa.europa.eu/publications/methodological-principles-insurance-stress-testing-cyber-component_en`.

M. Eling, M. Elvedi, and G. Falco. The economic impact of extreme cyber risk scenarios. *North American Actuarial Journal*, 27(3):429–443, 2023. doi: 10.1080/10920277.2022.2034507. URL `https://doi.org/10.1080/10920277.2022.2034507`.

Flexera. State of the cloud report, 2023. URL `https://info.flexera.com/CM-REPORT-State-of-the-Cloud`.

C. A. Floudas and V. Visweswaran. Quadratic optimization. *Handbook of global optimization*, pages 217–269, 1995.

A. Gunjan and S. Bhattacharyya. A brief review of portfolio optimization techniques. *Artificial Intelligence Review*, 56(5):3847–3886, 2023.

P.-F. Hsu. A deeper look at cloud adoption trajectory and dilemma. *Information Systems Frontiers*, 24(1):177–194, 2022.

Lloyd's. Cloud down impacts on the us economy. air worldwide., 2018. URL https://www.lloyds.com/~/media/files/news-and-insight/risk-insight/2018/cloud-down/aircyberlloydspublic2018final.pdf.

G. Mainik, G. Mitov, and L. Rüschendorf. Portfolio optimization for heavy-tailed assets: Extreme risk index vs. markowitz. *Journal of Empirical Finance*, 32:115–134, 2015.

H. M. Markowitz. Foundations of portfolio theory. *The journal of finance*, 46(2):469–477, 1991.

L. Mastroeni and M. Naldi. Insurance pricing and refund sustainability for cloud outages. In *Economics of Grids, Clouds, Systems, and Services: 14th International Conference, GECON 2017, Biarritz, France, September 19-21, 2017, Proceedings 14*, pages 3–17. Springer, 2017.

L. Mastroeni, A. Mazzoccoli, and M. Naldi. Pricing cat bonds for cloud service failures. *Journal of Risk and Financial Management*, 15(10):463, 2022.

L. Mastroeni, A. Mazzoccoli, and M. Naldi. Cyber insurance premium setting for multi-site companies under risk correlation. *Risks*, 11(10):167, 2023.

T. Mikosch and A. V. Nagaev. Large deviations of heavy-tailed sums with applications in insurance. *Extremes*, 1:81–110, 1998.

M. Naldi and L. Mastroeni. Economic decision criteria for the migration to cloud storage. *European Journal of Information Systems*, 25:16–28, 2016.

J. Ray and S. Bhattacharyya. Particle swarm optimization technique for optimizing conditional value-at-risk based portfolio. *International Journal of Computer Sciences and Engineering*, 5(2), 2017.

M. Weber. A weighted central limit theorem. *Statistics and Probability Letters*, 76(14): 1482–1487, 2006. ISSN 0167-7152. doi: https://doi.org/10.1016/j.spl.2006.03.007. URL https://www.sciencedirect.com/science/article/pii/S0167715206000824.