esma European Securities and
Markets Authority

ESMA Working Paper
No. 2, 2022

Financial stability risks from cloud outsourcing

Carolina Asensio, Antoine Bouveret, Alexander Harris

May 2022 | ESMA/2022/WP-2022-2

ESMA50-165-2138

ESMA Working Paper, No. 2, 2022

Authors: Carolina Asensio, Antoine Bouveret, Alexander Harris

Authorisation: This Working Paper has been approved for publication by the Selection Committee and reviewed by the Scientific Committee of ESMA.

Legal reference of this Report: Regulation (EU) No 1095/2010 of the European Parliament and of the Council of 24 November 2010 establishing a European Supervisory Authority (European Securities and Markets Authority), amending Decision No 716/2009/EC and repealing Commission Decision 2009/77/EC, as amended by Regulation (EU) 2019/2175 of the European Parliament and of the Council of 18 December 2019, Article 32 ‘Assessment of market developments, including stress tests’, ‘1. The Authority shall monitor and assess market developments in the area of its competence and, where necessary, inform the European Supervisory Authority (European Banking Authority), and the European Supervisory Authority (European Insurance and Occupational Pensions Authority), the ESRB, and the European Parliament, the Council and the Commission about the relevant micro-prudential trends, potential risks and vulnerabilities. The Authority shall include in its assessments an analysis of the markets in which financial market participants operate and an assessment of the impact of potential market developments on such financial market participants.” The charts and analyses in this report are, fully or in parts, based on data not proprietary to ESMA, including from commercial data providers and public authorities. ESMA uses these data in good faith and does not take responsibility for their accuracy or completeness. ESMA is committed to constantly improving its data sources and reserves the right to alter data sources at any time.

European Securities and Markets Authority (ESMA)
Risk Analysis and Economics
201-203, Rue de Bercy
FR–75012 Paris
risk.analysis@esma.europa.eu

Financial stability risks from cloud outsourcing[♦]

Carolina Asensio[♣], Antoine Bouveret[♣♣], Alexander Harris[♣♣♣]

Abstract

Financial institutions increasingly rely on Cloud Service Providers (CSPs). Cloud services can increase resilience of individual firms. However, given the high concentration of CSPs, a single CSP outage could generate simultaneous firm-level outages, posing systemic risk. Our model examines this possibility. We calibrate the model with operational risk data to simulate outages among CCP clearing members and show that CSPs need to be significantly more resilient than firms to improve the safety of the financial system. In financial settings where only longer (multi-period) outages impose systemic costs, CSPs can best address systemic risk by strongly reducing incident resolution time, rather than incident frequency. Finally, we show that the use of an idealized back-up CSP successfully mitigates systemic risk from CSPs. Back-up requirements may need to be imposed by policymakers however, as the systemic risk is an externality to individual firms.

JEL Classifications: D62, D82, D85, G23

Keywords: Operational risk, cloud service provider, systemic risk, concentration risk, central counterparty

[♦] The views expressed are those of the authors and do not necessarily reflect the views of the European Securities and Markets Authority. Any error or omissions are the responsibility of the authors. A summary of the findings of this paper was published in the ESMA Report on Trends, Risks and Vulnerabilities no. 2, 2021. The authors would like to thank Adrien Amzallag, Keith Bear, Richard Berner, Jon Danielsson, Claudia Guagliano, Leonardo Gambacorta, Vikram Haksar, Victor Herraez, Steffen Kern, Caroline Le Moign, Klaas Lenaerts, Maurizio Naldi, Christian Winkler and seminar participants at the BIS and IMF for discussions and feedback on earlier versions of the work. They are grateful to Yanis El Omari for help sourcing the data.

[♣] Trainee, Risk Analysis and Economics Department, European Securities and Markets Authority (ESMA), CS 80910, 201-203 rue de Bercy, 75589 Paris Cedex 12, France. E-mail: carolina.asensio@esma.europa.eu.

[♣♣] Senior Economist, Risk Analysis and Economics Department, European Securities and Markets Authority (ESMA), CS 80910, 201-203 rue de Bercy, 75589 Paris Cedex 12. E-mail: antoine.bouveret@esma.europa.eu.

[♣♣♣] Senior Risk Analysis Officer, Risk Analysis and Economics Department, European Securities and Markets Authority (ESMA), CS 80910, 201-203 rue de Bercy, 75589 Paris Cedex 12. E-mail: alexander.harris@esma.europa.eu.

1 Introduction

The use of cloud computing services by financial institutions has expanded over the last few years, as firms are increasingly outsourcing parts of their IT infrastructure (FSB, 2019). While migrating to the cloud provides a range of benefits to firms including scalability and flexibility, the high concentration of the CSP market can present risks to financial stability, especially from an operational risk perspective (Danielsson and Macrae, 2019).

CSPs can increase the resilience of financial institutions as CSPs invest heavily in security and spread their infrastructures across geographical areas. However, the high degree of concentration in the market implies that if a CSP were to suffer a major outage, a large number of clients would face operational challenges.

This paper analyses the impact of migrating to the cloud on the stability of the financial system. It shows the importance of trade-offs between higher individual resilience for firms using CSPs and higher risk of tail events, where multiple firms suffer an outage at the same time.

First, in a stylized theoretical framework, we characterize average outage time for firms in terms of incident frequency and average incident resolution time (Proposition 1). We show that while the use of CSPs typically improves the resilience of financial institutions at individual level, concentration risk can lead to systemic risk due to a higher probability of simultaneous outages (Proposition 2). For systemic risk to be lower, CSPs need to be substantially more resilient than individual firms (lower frequency and/or lower duration of outages) to compensate for concentration risk. Using our model, we also examine how that the use of a second CSP or an operationally separate zone in a single CSP as back-up (‘multi-cloud’) for a given core financial activity mitigates risks to financial stability.[1] Successful mitigation is possible if CSPs do not share common vulnerabilities. However, there exists an equilibrium in which firms outsource to the cloud but do not back up (Proposition 3), suggesting that policy intervention may be warranted.

Second, we use outage data from financial institutions and CSPs to estimate the likelihood of systemic events in a network of clearing members. Using a small sample, we find that CSPs have less frequent outages than clearing members (proxied by CCP outage data) but of longer duration. In that set-up, outsourcing to CSPs creates systemic risk due by increasing the likelihood of simultaneous outages.

Overall, our results provide a framework to analyse the benefits and risks related to third-party outsourcing, which can be used by policymakers and regulators in the context of ongoing policy work on CSPs (FSB, 2020). We also shed light on what kind of data on outages is needed to estimate risks and costs related to third-party outsourcing.

This paper complements recent work on risks related to CSPs. Lloyd’s uses a scenario analysis to quantify the losses related to an outage of CSPs at global level and more specially for the largest US companies (Lloyd’s, 2017, 2018). Using a Value-at-Risk approach, (Naldi, 2017) provides a measure of potential losses per CSP, by calibrating frequency and severity distributions on outage data. (Aldasoro et al., 2020) find that a higher dependence on CSPs, measured by investment in cloud services at country-level, is associated with lower cyber losses, yet the authors note that since they only have small losses in their database, this result might not apply to more extreme events. To our knowledge, our paper is first to model the marginal systemic risk that arises from cloud-based outsourcing of critical financial services, compared with a no-cloud baseline. Our simple framework calculates this risk analytically, taking

[1] We model an idealised case of back-up whereby cloud services are fully and instantly portable between providers or availability zones. As we discuss in section 4, these conditions may not hold in practice.

mean duration and frequency of outages as inputs. We also show how back-up of critical financial services via multi-cloud may be under-provided in equilibrium compared to the social optimum.

The remainder of the paper is as follows: Section 1 provides an overview of the use of CSPs by financial institutions and some of the risks that can arise. Section 2 outlines a stylized model to assess risks related to the use of CSPs. Section 3 applies the model to a CCP network and Section 4 concludes.

2 Motivation

2.1 Cloud service providers and use by financial institutions

Cloud computing is an innovation that allows for the use of an online network ('the cloud') of hosting processors to increase the scale and flexibility of computing capacity (FSB, 2019).

There are three main service models for cloud computing:

- *Infrastructure as a Service (IaaS):* Provides computing resources and IT infrastructure to clients, on which they can deploy and run arbitrary software.
- *Platform as a Service (PaaS):* Provides clients with an on-demand environment for developing, testing, delivering and managing software online.
- *Software as a Service (SaaS):* Clients use the provider's apps on a cloud infrastructure, which are accessible from various devices through either a client or program interface.

When it comes to how these cloud services are deployed, there are three deployment models:

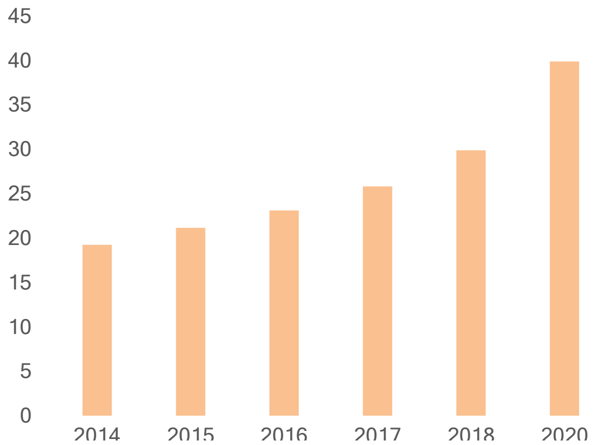
- *Public Cloud:* The cloud infrastructure is provided for open use by the public over the internet and exists on the premises of the cloud provider. Some common uses of public cloud are storage, testing and development environments, and web-based email.
- *Private Cloud*: The service is provisioned for exclusive use by a single organization, with all services kept on a private network. This deployment eases customization to meet IT requirements or to guarantee an enhanced control over the outsourced services.
- *Hybrid Cloud*: Combination of private and public cloud infrastructures that remain unique units but are linked by proprietary technology enabling data and app portability.

Given these possible deployment models, firms are starting to rely on a 'multi-cloud' approach. In this context, they outsource many services (through both public and/or private deployments) to more than one CSP to reduce dependence on a single provider.

While cloud computing is still a topic of research, it has become key to the digital economy. The use of cloud has significantly increased in the last few years (see Chart 1), a trend which has been further accelerated by the pandemic as firms have had to set up remote working facilities.

Migrating to the cloud, once seen as an option, is now viewed by many businesses as inevitable, and more firms in the financial sector are increasingly outsourcing or intending to outsource to CSPs in the near future to remain competitive. According to McKinsey's results in a poll of banking and securities organizations, while 81% of respondents claimed to currently hold less than 25% of their environments in the public cloud, 54% said they were targeting to raise this amount to over 50% in the next five years (McKinsey, 2021).

Chart 1
Firms increasingly purchasing cloud services



Note: Percentage of businesses purchasing cloud computing services by year in 22 EU countries. Countries included: AT, BE, CZ, DE, DK, EE, ES, FI, FR, GR, HU, IE, IT, LV, LT, LU, NL, PL, PT, SI, SK, SE. Firms across the economy with at least 10 employees were surveyed. Sources: OECD, ESMA.

There are many benefits associated with the use of cloud computing in the financial system. Cloud can help firms cut costs related to the development and maintenance of IT infrastructures that keep up with the pace of innovation, as financial services firms seldom have the scale and capacity to set up in-house such sophisticated and highly automated infrastructures. This can enable financial institutions to direct internal resources that were previously focused on administrating IT infrastructure towards innovating and delivering new and improved products and services. Cloud computing can also help firms expedite and scale processes, increase flexibility and operational efficiency, as well as enhance the ability to identify business opportunities and revenue streams. Another key benefit related to cloud outsourcing is risk mitigation through enhanced information security and disaster recovery plans, given that CSPs can provide efficient solutions to mitigate traditional technology risks, such as capacity, redundancy, and resiliency concerns. Moreover, cloud migration plays a huge role as an enabler of the use of other innovative technologies such as artificial intelligence, big data and distributed ledger technology, allowing for higher automation.

2.2 Potential risks related to the use of CSPs

Cloud outsourcing can offer several benefits, but it can also raise challenges at firm level in terms of governance, data protection and information security. Operational risks result from inadequate or failed internal processes, people, and systems, or from external events, and they may impact financial institutions in different ways. For instance, data losses could happen due to failures, deletion or disasters that occur at CSPs. For example, in February 2012, customers were unable to access Azure hosted services due to a ‘leap day bug’ related to 29 February. In August 2015, a series of lightning strikes to the local power grid caused failures in a Google data centre in Belgium (Lloyd’s, 2018).

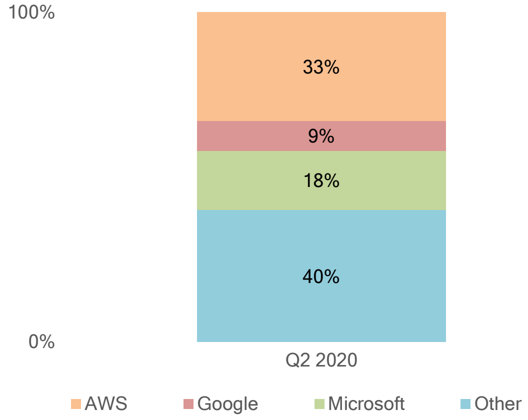
These risks can also arise when CSPs outsource some of their functions to other third parties, or ‘fourth parties’. Another important type of operational risk to consider is cyber risk; as massive amounts of data are stored in cloud ecosystems, these become very attractive targets for cyber criminals. It is important to note that all these risks can be mitigated by financial institutions in the due diligence process when selecting an adequate CSP, when drafting the corresponding service level agreements (SLA) and by adhering to regulations and operational resilience principles.

Another important risk embedded in all new technologies is ‘vendor lock-in’, where a financial institution relies strongly on the services of one CSP, for instance due to use of software technology only supported by one CSP. This could lead to severe difficulties when migrating to another provider. Depending on the level of dependency and on the CSP’s commitments, it may even lead to a catastrophic business failure should the cloud provider go bankrupt (or decide for business reasons to stop providing cloud services, etc).

In addition, cloud computing can bring risks at the level of the wider financial system. An increasing concern in the regulatory and supervisory landscape relating to outsourcing and third-party risk management is the possibility of systemic risk arising from the concentration in the provision of cloud services, which becomes higher as the number of financial institutions that outsource critical or important functions to CSPs increases (FSB, 2020). Given the limited number of CSPs that can meet the high standards of resiliency requirements that financial institutions demand, it is plausible that a sufficiently large number of them become dependent on a small number of CSPs. This implies that operational incidents may become more correlated among those financial institutions that outsource critical or important functions to a common CSP. Even though cloud computing may yield increased data security and operational resilience at firm level, it could also increase the risk of simultaneous incidents among several firms and lead to potential negative outcomes for financial stability (Danielsson and Macrae, 2019; FSB, 2019). Concentration risk in this context is thus a form of systemic risk.

Synergy Research Group estimated that the largest three CSPs made up for 60% of market share for the IaaS market. A major disruption, outage, or failure at one of these CSPs, even if unlikely, could create potential concentration risk in terms of a single point of failure, with likely severe consequences for financial stability.

Chart 2
Cloud outsourcing is a concentrated market



Note: Global market share of cloud infrastructure services in Q2 2020, by vendor. Source: Synergy Research Group

Although operational, vendor lock-in and financial stability are the most obvious risks, there are several other risks related to cloud outsourcing that are important to take into consideration when assessing a potential migration to the cloud. For an extensive review of other risks related to cloud computing refer to the cloud risk assessment performed by the European Network and Information Security Agency (ENISA, 2012).

3 Systemic risks related to CSPs: a model

3.1 Literature review

The increasing use of CSPs has been accompanied by an emerging literature on the risks and potential impact of CSP outages.

A series of studies estimate the costs related to outages of cloud providers. Using scenario analysis, Lloyd's estimates global losses ranging from USD 4bn to USD 53bn for an outage duration of between 0.5 and 3 days (Lloyd's, 2017), and losses for the largest US firms at around USD 10bn for an outage of the top three CSPs lasting between 3 and 6 days (Lloyd's, 2018).

Using a Value-at-Risk approach, Naldi (2017) provides a measure of potential losses for CSPs, based on outage data and estimated loss per minute. The author models outage frequency using a Poisson distribution and outage duration using a Generalized Pareto Distribution, frequently used to model fat tails in operational risk (Bouveret, 2019). Our model builds on this approach, distinguishing between outage frequency and duration. For tractability, and to prevent time-consistency (i.e. time-overlapping outages for a single area of a firm's operations), we do so in a two-state Markov chain framework. This allows us to analyse alternative technology-based approaches to mitigating systemic risk: preventing outages versus quickly resolving them.

A related strand of the literature examines the impact of using CSPs on the cost of cyber events for individual firms. Using a large dataset of cyber losses, Aldasoro et al (2020) find that a higher dependence on CSPs, measured by investment in cloud services at country-level, is associated with lower costs. However, the authors note that this result might not apply to more extreme events since they only have small losses in their database. Harmon, Vytelingum and Babaie-Harmon (2020) put forward an agent-based model with banks and CSPs in a settlement context. CSPs can face outages, the duration of which is assumed to follow an exponential distribution. When a CSP suffers an outage, banks using the CSP cannot proceed with settlement, creating credit risk. The authors estimate the impact on other banks in the network, using contagion measures based on market-based data for banks (Demirer et al, 2018).

3.2 Overview of the model and scenario 1 (no-cloud baseline)

We now introduce a model to investigate the conditions under which outsourcing to the cloud by financial sector firms may generate systemic operational risk. The model considers a set of financial sector firms in three main alternative scenarios:

1. A setting where no cloud outsourcing is available (the 'no-cloud scenario');
2. A setting where each financial sector firm outsources the time-critical IT service to one among several CSPs (the 'cloud scenario'); and
3. A setting where each financial sector firm outsources the time-critical IT service to a primary CSP and to a secondary provider (the 'multi-cloud scenario').

In this section, we do not explicitly model firms' decision as to whether to outsource to the cloud. Instead, we examine the risk implications of the first two scenarios. In section 4, we also consider risks in scenario 3. Yet the model can readily be understood in a strategic context. Firms will have an incentive to move operations to the cloud – other things equal and neglecting frictional costs – if cloud outsourcing prevents incidents or improves their resolution speed. In Annex 1, we model this strategic decision formally. There we show that firms may not find it individually optimal to use a back-up cloud provider, even if the system would be more efficient if all firms were to back-up. In short, there is an externality that may warrant policy intervention.

Throughout the model, the number of financial sector firms is denoted n and the number of CSPs is denoted n_{cloud}. Unless otherwise stated, firms are indexed by i, operational incidents by j and CSPs by k. Time is discrete: $t = 1,2,3, \dots$

At any given time t, firm i is in one of two states: $\omega_i(t) = 0$ (no outage) or $\omega_i(t) = 1$ (outage occurs). We define $m(t)$ to be the total number of firms suffering an outage at time t, i.e.:

$$m(t) = \sum_{i=1}^{n} \omega_i(t) \tag{1}$$

Assumption 1: Firms' states are independent. Transition probabilities between states are constant.

Assumption 1 – particularly the independence of firms' states – is important. In reality, firms' outages may be correlated – for instance, if there is a natural disaster affecting data systems in a geographical region, or if a malicious attack targets several firms at once. The assumption contrasts with scenario 2 (detailed below), where all firms that have the same CSP suffer perfectly correlated outages. Independence can to some extent be justified by interpreting the model as a means to study the difference in systemic risk between scenarios 1 and 2, abstracting away from those risk drivers that are common to both settings. For instance, to the extent the two scenarios face a common risk of a multi-firm malicious attack – which can be perpetrated directly against the firms or via the cloud – we can regard the effect as 'cancelling out' between the scenarios. However, the independence assumption clearly reduces baseline systemic risk in scenario 1, which therefore overstates the extent to which CSPs create *additional* systemic risk via concentration.

The second part of Assumption 1, that transition probabilities are constant, implies that outages follow a Markov chain. Regardless of the system's initial configuration, it has long-run steady state properties that we can study. For example, given the transition probabilities we can calculate the average amount of time a firm spends in outage, the average amount of time that two or more firms are in simultaneous outage, and the frequency that a firm suffers a multi-period outage of given duration.

Returning to the model specification, suppose that $\omega_i(t) = 0$ at arbitrary time period t, i.e. the firm is not suffering an operational outage. The probability that the firm remains in that state at the next time period $t + 1$ (i.e. that no outage occurs in $t + 1$) is a constant probability λ, known as the *incident rate*. The mean length of time the firm remains free of an outage is as follows:

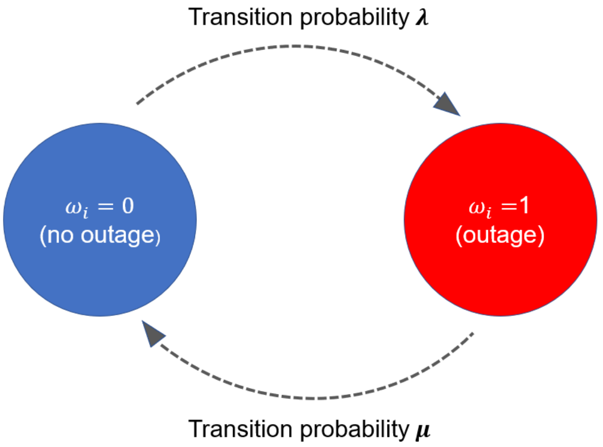
$$Mean\ time\ between\ outages\ = \frac{1}{\lambda} \tag{2}$$

Now suppose instead that $\omega_i(t) = 1$, i.e. an operational outage is underway at t. The probability that the outage is resolved in the next period $t + 1$ is a constant probability μ, known as the *repair probability*. We have that:

$$Mean\ time\ of\ an\ outage\ = \frac{1}{\mu} \tag{3}$$

Chart 3 illustrates the transition process for a single firm.

Chart 3
Markov chain governing outages over time for a single firm


Transition probability $\boldsymbol{\lambda}$
$\omega_i = 0$
(no outage)
$\omega_i = 1$
(outage)
Transition probability $\boldsymbol{\mu}$

Note: Markov chain diagram for a single firm in the no-cloud baseline scenario, in which possible states of the firm are represented by coloured circles.

The long-run average share of time that firm i spends in the outage state is defined as τ:

$$\tau \coloneqq \lim_{t\to\infty} Prob[\omega_i(t) = 1] \tag{4}$$

***Proposition 1**: Irrespective of the initial state, the long-run average share of time a firm spends in the outage stage is* $\boldsymbol{\tau = \frac{\lambda}{\lambda+\mu}}$

Proposition 1 says that the share of time spent in the outage stage is increasing in the incident rate λ (measuring the probability of transitioning to an outage state) and is decreasing in the recovery rate μ (the probability of transitioning to a no-outage state).[2] This result is intuitive: a higher incident rate leads to more frequent outages, while the greater the recovery rate, the quicker they are resolved. Henceforth we make the following assumption:

Assumption 2: At $t = 0$, each firm's state is independently drawn, with $Prob[\omega_i(t) = 1] = \tau$

Assumption 2 does not drive the main results of this paper but is a simple way to enable us to investigate the steady state properties of the system of firms. We assume that the system starts off in a state that is 'natural', in the sense that it is generated by the same probability distribution to which the system converges in the long run from any starting condition. A consequence of this assumption is that ex-ante, the probability distribution over firms' states is the same at any

[2] Proof: Proposition 1 is a standard result for a finite, discrete Markov chain. Let $\boldsymbol{V}$ denote the 2x1 steady state vector, whose first entry is the probability of no outage and whose second entry is the probability of outage. Let $\boldsymbol{P}$ denote the 2x2 transition matrix $\begin{pmatrix} 1-\lambda & \lambda \\ \mu & (1-\mu) \end{pmatrix}$, where element $\boldsymbol{P}_{ij}$ is the transition probability between states i and j.
It is straightforward to verify that $\boldsymbol{V} = (1-\tau, \tau)$, where $\tau = \frac{\lambda}{\lambda+\mu}$, by substitution into the steady state equation $\boldsymbol{VP} = \boldsymbol{V}$.

point in time. Assumption 2 is not critical; we could instead have specified arbitrary initial conditions and studied the long-run stochastic properties of the system, as the ex-ante probability of being in the outage state will converge to τ.[3]

Given that Assumption 2 specifies that firms' initial states are identically and independently distributed (i.i.d), the expected number of firms suffering an outage at a given point in time is the product of the individual probabilities of each firm suffering an outage. Let $E[.]$ denote ex-ante expectations under Assumption 2. It follows:

$$E[m(t)] = n\tau \tag{5}$$

Equation (5) states that on average there are $n\tau$ firms in an outage state at any point in time.

More generally, the number of firms suffering an outage follows a binomial distribution.[4] For example, the probability that at least two firms suffer an outage is given by $1 - B_n(1,\tau)$, where $B_n(x,p)$ is the cumulative binomial distribution with n trials, x is the number of trials with a given outcome and p is the probability of that outcome. Explicitly, $B_n(1,\tau) = \mathrm{Prob}[m(t) = 0] + \mathrm{Prob}[m(t) = 1] = (1-\tau)^n + \tau(1-\tau)^{n-1}$, where $m(t)$ is a random variable measuring the number of firms in outage in arbitrary period t. Thus, the probability of at least 2 firms being in simultaneous outage in any time period is:

$$\mathrm{Prob}[m(t) \geq 2] = 1 - (1-\tau)^n - \tau(1-\tau)^{n-1} \tag{6}$$

In general, we can consider the probability that there are at least $S > 0$ firms in outage in a given time period. The approach of equation (6) can be used to calculate this probability:

$$\mathrm{Prob}[m(t) \geq S] = 1 - \mathrm{Prob}[m(t) \leq S-1] = 1 - B_n(S,\tau) = 1 - \sum_{i=0}^{S-1} \binom{n}{i} \tau^i \, \left(1-\tau^i\right)^{n-i} \tag{7}$$

According to equation (7), the probability of a systemic event is decreasing in S, the threshold number of firms, and increasing in τ, the average time a firm spends in outage. Both these dependencies are strongly intuitive, as $\tau = \frac{\lambda}{\lambda+\mu}$ (Proposition 1) implies that the probability of a systemic event is increasing in the incident rate λ and the repair rate μ.

3.3 Scenario 2: Systemic cloud outsourcing

Suppose now that there are $n' \geq 1$ cloud providers, each of which serves an equal share of the n firms.[5] Each cloud provider j is in one of two states at any given time: $\omega_j(t) = 0$ (no operational outage) or $\omega_j(t) = 1$ (outage occurs).

Firms only suffer outages via the cloud, and a cloud provider outage affects all the firms that use this cloud provider. More formally, if firms i and j uses cloud provider k, then $\omega_i(t) = 1$ if and only if $\omega_j(t) = 1$. Each cloud provider is equally likely to suffer an outage, and hence each firm is equally likely to suffer an outage.[6]

[3] The convergence property is guaranteed by the specification of the system as an aperiodic, irreducible Markov chain.

[4] The binomial distribution arises from the fact that each firm's state is identically and independently distributed (i.i.d) in period 1 under Assumption 2. Ex-ante, firms' states are i.i.d. in subsequent periods as the Markov process starts in its steady state under Assumption 2.

[5] We assume without loss of generality that n is divisible by n_{cloud} without remainder.

[6] This consequence holds for any assignment of firms such that each has a single provider.

It follows that at any given time, the number of firms in outage $m(t)$ equals the number of cloud providers in outage multiplied by the number of firms per cloud provider. More formally,

$$m(t) = \frac{n}{n'} \sum_{j=1}^{n'} \omega_j(t) \tag{8}$$

All cloud providers have the same incident rate, denoted λ', and same repair rate, denoted μ'.

Analogously with the no-cloud case, let us define the average time in outage for a firm using a cloud provider, denoted τ':

$$\tau' \coloneqq \lim_{t\to\infty} Prob[\omega_{cloud}(t) = 1] \tag{9}$$

so that

$$\tau' = \frac{\lambda'}{\lambda' + \mu'} \tag{10}$$

In Annex 1, we assume that cloud providers have lower average outage time than do firms in the no-cloud scenario, *i.e.* $\tau' < \tau$. Under this assumption we set out a simple equilibrium framework in which all firms find it optimal to outsource to the cloud.

3.4 Comparing simultaneous outages between scenarios

Recall that each cloud provider is assumed to service $\frac{n}{n'}$ firms. It follows that at least S firms suffer an outage if at least $\frac{S\cdot n'}{n}$ providers are in simultaneous outage.

Denote $S' \coloneqq \frac{S\cdot n'}{n}\right\rceil < S$. The quantity S' is the number of cloud providers in outage (rounded up to the nearest integer) that results in S firms being in outage.

The risk that at least S firms suffer a simultaneous outage in any time period in scenario 2 is:

$$\mathrm{Prob}[m(t) \geq S \mid Cloud] = 1 - B_{n'}(S', \tau')$$

where $Cloud$ denotes the event that scenario 2 takes place, i.e. that firms outsource to the cloud.

The cumulative binomial distribution $B_n(.,.)$ is decreasing in its first argument (i.e. the threshold number of trials) and increasing in its second argument (the probability that a trial is 'successful' – in the present context, that an outage occurs.)

Define the *odds ratio*, R, as follows:

$$R \coloneqq \frac{\mathrm{Prob}[m(t) \geq S' |\, Cloud]}{\mathrm{Prob}[m(t) \geq S |\, No\ cloud]} = \frac{1 - B_{n'}(S', \tau')}{1 - B_n(S, \tau)} \tag{11}$$

The odds ratio describes how many times more likely a simultaneous outage of at least S firms is in scenario 2 than scenario 1. Intuitively, it describes how much more likely such an outage is made by concentration risk due to cloud outsourcing.

The numerator of the right-hand side of equation (11) can be rewritten as $1 - B_{n\cdot\frac{n'}{n}}\left(S \cdot \frac{n'}{n}, \tau'\right)$, where the number of trials and the threshold number of outages are scaled by a common factor

$\frac{n'}{n}$. If we set $\tau' = \tau$, it follows that $B_{n'}(S', \tau') < B_n(S, \tau)$ and hence that $R > 1$. In other words, if the probability of outage for any given firm is the same across the two scenarios, then the probability of simultaneous outage of S firms is greater in the cloud scenario. If we decrease τ' continuously, $B_{n'}(S', \tau')$ will increase continuously, with $B_{n'}(S', 0) = 1$. Propositon 2 follows.

***Proposition 2*: There is a unique outage probability $\tau' < \tau$ for cloud providers that yields an odds ratio of 1, i.e. the probability that at least S firms are out is equalized across the no-cloud and cloud scenarios. The odds ratio is strictly increasing in τ'.**

3.5 A stylized numerical example

For concreteness, let us consider 20 firms and assume that in scenario 2 we have 5 CSPs with 4 firms each as clients. As the purpose at this stage is simply to illustrate the properties of the model, we set parameters such that the risks are relatively large.[7]

Table 1: Parameter values in stylized example

Parameter	Interpretation	Value
n	Number of firms	20
n'	Number of CSPs	5
S	Minimum number of firms in simultaneous outage for systemic event	4
λ	Per-period probability of new outage in no-cloud baseline	5%
μ	Per-period probability that an outage is resolved in no-cloud baseline	45%
λ'	Per-period probability of new outage in cloud scenario	1.6%
μ'	Per-period probability that an outage is resolved in cloud scenario	78%

We obtain the average time spent in outage in the baseline scenario:

$$\tau = \frac{\lambda}{\lambda + \mu} = 10\% \tag{12}$$

The average time a firm spends in outage in the cloud scenario is:

$$\tau' = \frac{\lambda'}{\lambda' + \mu'} = 2.0\% \tag{13}$$

In this example, a firm in the baseline scenario would spend around five times as long in outage compared with if it outsourced to the cloud.

The probability of a systemic event in the no-cloud baseline is given by:

[7] We do not interpret time periods here but note that the incident frequency estimates would be unrealistically high if we were working on the basis of hourly periods as in the application of section 3. Conversely, those estimates would be more realistic if were we suppose a single time period to be an appropriately long interval (e.g. days or weeks, depending on the application), but the assumed recovery parameters μ would then be unrealistically low. The aim of the stylized model is to aid intuition; a more realistic calibration is provided in section 3.

$$\text{Prob}[m(t) \geq 4] = 1 - \text{Prob}[m(t) \leq 2] = 1 - \sum_{i=0}^{3} \binom{20}{i} 0.1^{i}\ (1-0.1)^{20-i}$$
$$= 13.3\% \qquad (14)$$

The probability of a systemic event under the cloud scenario is given by:

$$\text{Prob}[m(t) \geq 4] = 1 - \text{Prob}[m(t) \leq 3] = 1 - (1-0.02)^{5}$$
$$= 9.7\% \qquad (15)$$

In this stylized example, the improved resilience of the cloud provider – echoed in the inequality $\tau' < \tau$ – more than compensates the concentration risk brought by the cloud outsourcing model.

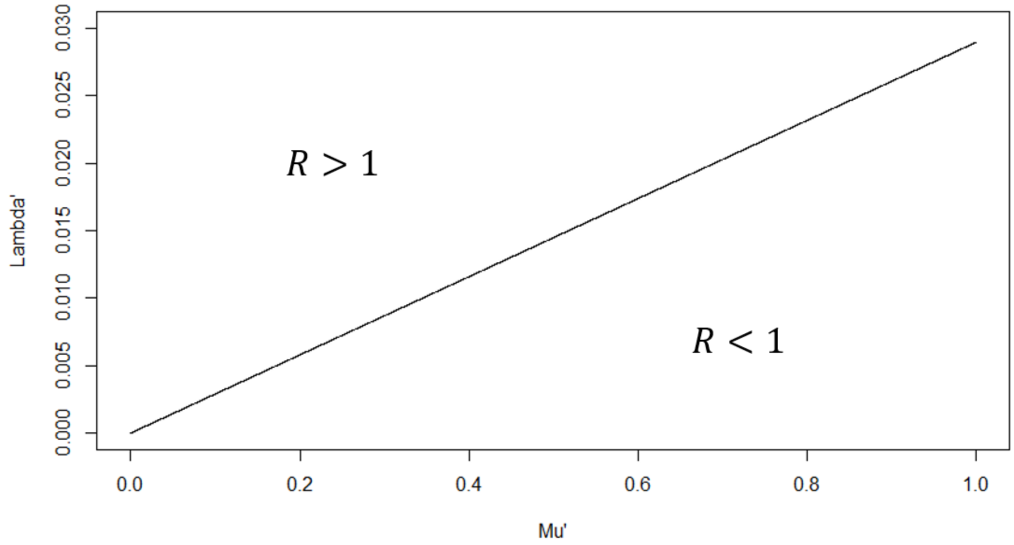
What is the approximate threshold value τ' such that $R = 1$, fixing $\tau = 10\%$?

Solving numerically yields that if we set $\tau' = 2.8\%$,

$$\text{Prob}[m(t) \geq 3] = 1 - \text{Prob}[m(t) \leq 2] = 1 - \left(1 - 0.035^{i}\right)^{4}$$
$$= 13.3\% \qquad (16)$$

- *Remark: Any values for the underlying parameters λ', μ' such that $\frac{\lambda'}{\lambda'+\mu'} = 2.8\%$ are consistent with this result. Chart 4 illustrates regions of parameter values for which systemic risk is higher in the cloud scenario than the no-cloud scenario, and vice versa. The solution relies on the specification that a systemic event occurs whenever there is a simultaneous outage. Note that the scale for μ' (horizontal axis) is around two orders of magnitude larger than that for λ'. This is to be expected, since doubling the repair rate will only approximately double the among of time in outage if the incident rate is small.*

Chart 4
Values of cloud incident rate (λ') and cloud repair rate (μ') that equalize systemic risk with baseline in stylized example



Note: The line plots values of cloud incident rate λ' and cloud repair rate μ' for which $\tau'(\lambda'', \mu') = 2.8\%$. In the region above the line, τ'> 2.8%. Assuming τ'= 10% in the no-cloud scenario, the odds ratio R > 1 in this region, i.e. the probability of simultaneous outage of at least 4 firms is greater when outsourcing takes place in the no-cloud baseline.

3.6 Minimum-time conditions

If instead we assume that a systemic event requires simultaneous outage among at least S of the same firms in two or more consecutive time periods, then the cloud recovery rate μ' plays a new role in determining systemic events. Specifically, we define:

$$m_2(t) \coloneqq |\{i: \omega_i(t) = 1 \cap \omega_i(t-1) = 1\}| \tag{17}$$

The quantity, $m_2(t)$ is the number of firms that have been in outage for at least 2 periods. Now:

$$\mathrm{Prob}[Systemic\ event\ at\ t] = \ \mathrm{Prob}[m_2(t) \geq S] \tag{18}$$

To evaluate this expression, first recall that in the no-cloud model, a firm remains in outage from one period to the next with probability $(1-\mu)$. Outage length therefore follows a geometric distribution. The cumulative density function for total outage length is as follows:[8]

$$F(t) = \mu \sum_{\nu=1}^{t} (1-\mu)^{\nu-1} = 1-(1-\mu)^{\nu} \tag{19}$$

Consequently, conditional on observing that a firm is *out* during a given period, the probability that the outage is in at least its second period is $1 - F(1) = (1-\mu)$, which is simply the probability of remaining in outage from one period to the next. Hence:

$$\mathrm{Prob}[m_2(t) \geq S] = 1 - \mathrm{Prob}[m_2(t) \geq S] = 1 - B_n(S, (1-\mu)\tau) \tag{20}$$

Let $m_k(t)$ be defined analogously with equation (17) to be the number of firms in outage for at least k consecutive periods:

$$m_k(t) \coloneqq |\{i: \omega_i(t) = 1 \cap \omega_i(t-1) = 1 \cap \ldots \cap\ \omega_i(t-k)\}| \tag{21}$$

Conditional on observing that a firm is *out* during a given period, the probability that the outage is in at least its kth period is $1 - F(k-1) = (1-\mu)^{k-1}$. If we assume that a systemic event requires simultaneous outage in k or more consecutive time periods, applying the formula for $F(t)$ yields the following expression for the no-cloud model:

$$\mathrm{Prob}[m_k(t) \geq S] = 1 - B_n\big(S, (1-\mu)^{k-1}\tau\big) \tag{22}$$

We can extend the simple illustrative example of section 3.5 to include minimum-time conditions for a systemic event. Suppose first that a systemic event requires the same four (or more) firms to be in outage for two consecutive periods. In this case, using the same parameters as before, we have that $(1-\mu)\tau = 55\% \times 10\% = 5.5\%$, while $(1-\mu')\tau' = 22\% \times 2\% \cong 0.44\%$.

The probability of a systemic event in the no-cloud baseline is now:

$$\begin{aligned}\mathrm{Prob}[m_2(t) \geq 4] = 1 - \mathrm{Prob}[m_2(t) \leq 3] = 1 - \sum_{i=0}^{3} \binom{20}{i} 0.055^{i}\ \big(1 - 0.055^{i}\big)^{20-i} \\ = \ 2.2\%\end{aligned} \tag{23}$$

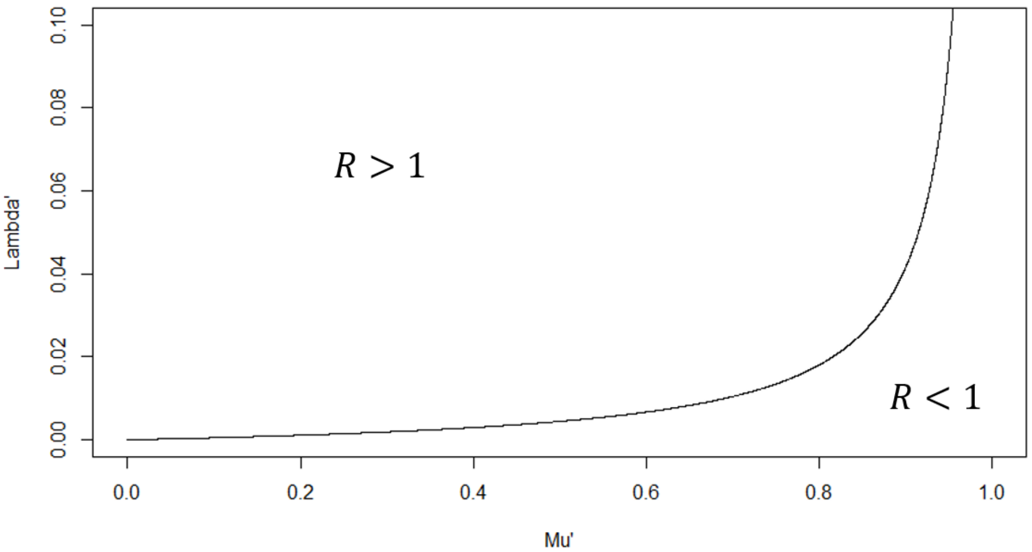
The probability of a systemic event under the cloud scenario is given by:

[8] $F(t) \coloneqq Prob(s \leq t)$, where $t = 1, 2, \ldots.$ is a given time period and s is the period at which a given outage finishes. The expression for the cloud scenario is closely analogous and so is omitted for brevity.

$$\begin{aligned}\text{Prob}[m_2(t) \geq 4] = 1 - \text{Prob}[m_2(t) \leq 3] = 1 - \left(1 - 0.0055^i\right)^4 \\ = \ 2.2\% \end{aligned} \qquad (24)$$

In contrast to the previous case, where the probability of being in cloud outage (at any given time period) of $\tau' = 2.8\%$ was needed to equalize systemic risk across the two scenarios, a lower cloud outage probability of $\tau' = 2\%$ now equalizes the values given the incident rate and repair rates. However, now that systemic events require consecutive outages, this value of τ' is not a unique solution, as the systemic outage probability is no longer only a function of τ'. Chart 5 plots the solutions to $(1 - \mu')\left(\frac{\lambda'}{\lambda' + \mu'}\right) = 0.44\%$, which equalizes the systemic risk rates between the two scenarios.

Chart 5
Values of cloud incident rate (λ') and cloud repair rate (μ') that equalize systemic event risk with baseline in stylized example (2-period threshold)



Note: The line plots values of cloud incident rate λ' and cloud repair rate μ' that equalize systemic outage probabilities between the no-cloud and cloud scenarios in the stylized application. Assuming that τ = 10% in the no-cloud scenario, the odds ratio R > 1 in the region above this line, i.e. the probability of simultaneous outage of at least 4 firms for at least 2 periods is greater when outsourcing takes place in the no-cloud baseline. The y-axis is truncated at λ' = 10% for clarity.

At high values of the cloud repair rate, there is a steep relation between the values of λ' and μ' that equalize systemic risk across the two scenarios, since in this region cloud outages become likely to be repaired within a single period.

General relation between λ' and μ' under minimum time conditions

To develop intuition on the role of the minimum-time condition in determining systemic risk, we now consider the more general case where a systemic event requires the same S firms to be in outage for $k \geq 1$ consecutive periods. The odds ratio $R = 1$ whenever:

$$\text{Prob}[m_k(t) \geq S \mid Cloud] = \text{Prob}[m_k(t) \geq S \mid No\ cloud] \qquad (25)$$

First let us denote $\text{A} \coloneqq \text{Prob}[m_k(t) \geq S \mid No\ cloud]$ to be the probability of a systemic event in the no-cloud baseline.

Then define B to be the probability that a single cloud provider has been in outage for at least k periods, i.e. that a systemic event takes place in the cloud scenario. $R = 1$ then implies that:

$$1 - (1 - B)^{n'} = A \tag{26}$$

i.e.

$$B = 1 - (1 - A)^{\frac{1}{n'}} \tag{27}$$

By analogy with equation (21), we know B also satisfies:

$$B = (1 - \mu')^{k-1}\tau' = \frac{(1 - \mu')^{k-1}\lambda'}{\lambda' + \mu'} \tag{28}$$

which rearranges to:

$$\lambda' B = (1 - \mu')^{k-1}\lambda' - B\mu' \tag{29}$$

i.e.

$$\lambda' = \frac{B\mu'}{(1 - \mu')^{k-1} - B} \tag{30}$$

for $\lambda' \in (0,1]$.

This functional form imposes strong convexity for high k, meaning that progressive improvements in μ' allow for larger increases in λ' under the constraint of equal systemic risk across scenarios ($R = 1$). In general, for sufficiently high μ', equation (29) cannot be satisfied for $\lambda' < 1$. In other words, if the CSP has a high enough repair rate, even very frequent outages will not create additional systemic risk, because they will be immediately resolved.

For further concrete illustration of this relationship, suppose that a systemic event requires the same four (or more) firms to be in outage for three consecutive periods.

Using the same parameters as before, we have that $(1 - \mu)^2\tau = 55\% \times 55\% \times 10\% = 3.025\%$, while $(1 - \mu')\tau' = 22\% \times 2.5\% = 0.121\%$.

Systemic risk in the no-cloud baseline is given by:

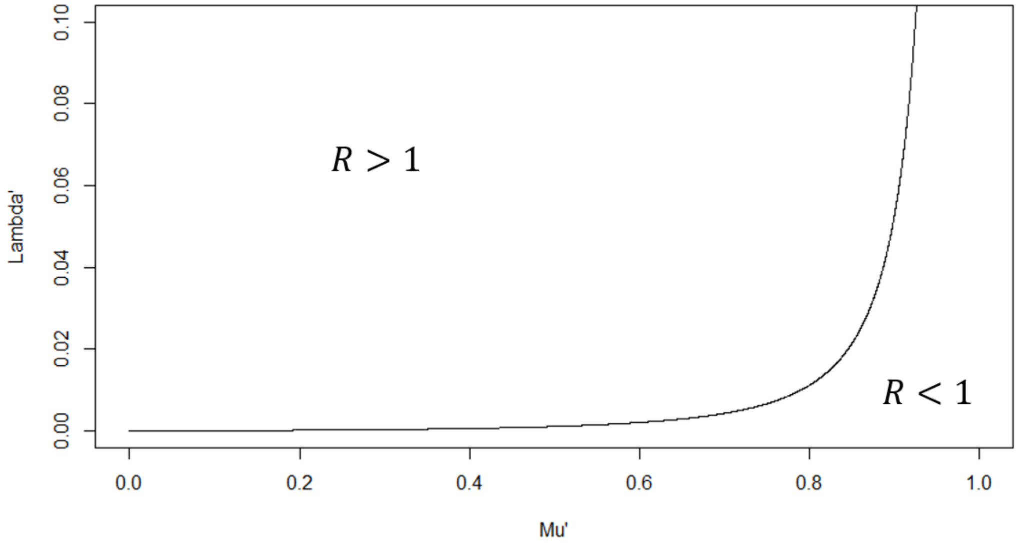
$$\text{Prob}[m_3(t) \geq 4 |\, No\ cloud] = 0.28\% \tag{31}$$

while in the cloud scenario, it is:

$$\text{Prob}[m_3(t) \geq 4 |\, Cloud] = 0.48\% \tag{32}$$

The solutions of (λ', μ') that yield $R = 1$ are even more strongly convex than in the previous case, as shown in Chart 6.

Chart 6
Values of cloud incident rate (λ') and cloud repair rate (μ') that equalize systemic event risk with baseline in stylized example (3-period threshold)



Note: The line plots values of cloud incident rate λ' and cloud repair rate $\bar{\mu}'$ that equalize systemic outage probabilities between the no-cloud and cloud scenarios in the stylized application. Assuming that τ' = 10% in the no-cloud scenario, the odds ratio R > 1 in the region above the line, i.e. the probability of simultaneous outage of at least 4 firms for at least 3 periods is greater when outsourcing takes place in the no-cloud baseline. The y-axis is truncated at λ' = 10% for clarity.

4 Application of the model to a network of clearing members

4.1 Motivation

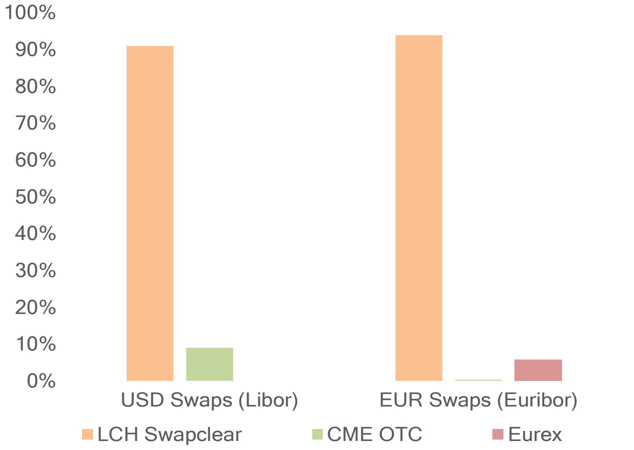
Systemic risk can arise if a large number of financial institutions become dependent on one or few third party providers (FSB, 2020). A major disruption of those providers could result in a single point of failure for the entire financial system. A range of financial institutions and infrastructures, which are critical to the function of the financial system could be considered exposed to this risk: payment and settlement systems (Eisenbach, Kovner and Lee, 2020), trading platforms or Global Systemically Important Banks, for example. Within financial market infrastructures, the clearing members that allow Centralized Counterparties (CCPs) to function constitute a possible real-world application of our model.

CCPs are key financial market infrastructures that clear transactions between market participants. Such clearing services rely on clearing members – firms that directly clear the transactions facilitated by CCPs. This arrangement is designed to reduce counterparty risk in the financial system, and the move to central clearing for some class of derivatives was one of the key reforms introduced after the Global Financial Crisis. CCPs and their clearing members tend to specialize in the clearing of specific instruments (e.g., interest rate derivatives or CDS), resulting in a concentrated landscape with a few dominant CCPs (see Chart 7**Error! Reference source not found.**). Clearing members' transactions also tend to be concentrated, with a small number of large institutions accounting for most of the activity (ESMA, 2020b; see Chart 8).

Chart 7

CCP market landscape

High concentration in CCP market



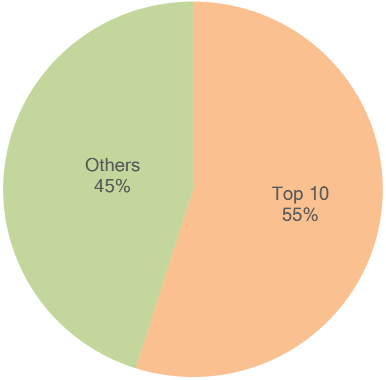
Note: Market share of CCPs for selected interest rate derivatives in percentages of single sided gross nationals, as of December 2020. Sources: Clarus, ESMA.

Sources: IOSCO – CPMI Quantitative Disclosures, ESMA.

Chart 8

Clearing members' transactions landscape

High concentration of collateral among clearing members of example CCP



Note: Share of total collateral posted at Eurex by different groups of clearing members. 'Top 10' = share of collateral posted by 10 clearing members with largest amounts of collateral posted. 'Others' = share of collateral posted by all other clearing members (n=114)

Sources: Eurex, ESMA.

In that context, the network of CCP clearing members offers a tractable application of the model outlined in the previous section. In particular, we consider the risk that of one or several clearing members are unable to clear transactions due to an outage in a CSP on which they rely[9].

We focus on of the risk of operational outages for clearing members, rather than an outage of a CCP itself, for a range of reasons. One is that while financial sector entities in general – including clearing members – increasingly are turning to cloud migration strategies, CCPs themselves have indicated they intend to retain the relevant operations in-house. Another reason to focus on clearing members is that the resulting setting fits naturally with the stylized model presented in section 2. In particular, the model is suited to studying the impact on the financial system of several institutions suffering an outage at the same time. In this way, the clearing members application offers a richer setting to analyse systemic risk across financial institutions, with CSPs representing a single point of failure (Danielsson and Macrae, 2019).

If clearing members outsource core services, and one or more CSPs suffers an outage, the impact on the financial system could be substantial. First, the failure of some clearing members to post collateral would lead to the liquidation of their positions according to the default management rules used by CCPs, implying potential losses due to fire sales and the consumption of some of the resources in the default fund. In addition, outages affecting clearing members could prevent some of their clients from clearing transactions with them. This, in turn, could result in additional costs – either in the form of frictional costs incurring by clients switching to other clearing members (where possible) or, worse, the cancellation of transactions where clearing cannot be executed. In its 2020 stress test, ESMA estimates that the failure of the two

[9] The outage of a CSP could also impact CCP critical functions not related to clearing and settlement such as risk management and risk monitoring.

largest counterparties to a CCP could lead to losses of around EUR 1bn each for the two largest EU CCPs (ESMA, 2020).

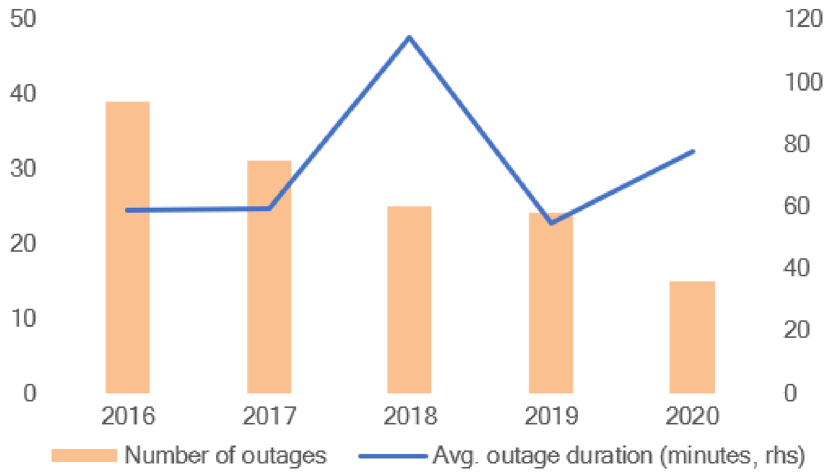
4.2 Data and calibration

Frequency and duration of outages for clearing members

To apply the model to the network of clearing members, the different parameters need to be calibrated. Calibration is particularly challenging due to the lack of public data on the frequency and duration of outages for clearing members. However, partial data are in fact available for CCPs themselves, which we use as a proxy. These partial data are sourced from the voluntary public quantitative disclosures set up by CPMI and IOSCO (CPMI-IOSCO, 2015). Our estimation assumes that in the no-cloud baseline, clearing members suffer outages as frequently and with the same average duration as CCPs. As noted above, we use CCP outage data as a proxy for clearing member outages given the lack of available data on outages among the clearing members themselves. However, the frequency and severity of clearing members outages might be different from CCPs due to a range of factors (use of different IT systems, intragroup exposures, different reliance on CSPs etc.).

The quarterly disclosures include information on operational risk (Principle 17 of the CPMI-IOSCO framework) such as the actual operational availability of the core system(s) over the previous twelve months and the total duration and number of failures affecting the core system(s) involved in clearing. We retrieve quarterly data for 10 CCPs over the 2016-2020 periods. Chart 9 provides summary statistics.

Chart 9
Reported outages and average outage duration (in minutes) used as a proxy for outages of clearing members in the no-cloud baseline



Note: Number of outages (trailing 12-month sum) and outage duration (in minutes, secondary axis) from 10 CCPs through the period 2016-2020. Observations that reported 0 outages during a period have been excluded from the analysis. CCP outage data are used as a proxy for outages by clearing members in the model presented in this paper.
Sources: IOSCO – CPMI Quantitative Disclosures from 10 CCPs : CME, DTCC, Eurex, ICC_CDS, IC NGX, ICUS_F&O, JSCC OTC-JGB, LCH.Clearnet.LTC, LCH.Clearnet.SA.

Table 2: Outage data used to model clearing member outages

Year	2016	2017	2018	2019	2020	*All years*
Total outages	39	31	25	24	15	*134*
Avg. outages per year per CCP	3.9	3.1	2.5	2.4	1.5	*2.7*
Avg. mean outage duration (hours)	1.0	1.0	1.9	0.9	1.3	*1.2*

Note: Outages are reported in the dataset on a 12-month rolling basis each quarter, so total outages per year are estimated using Q4 reported values by CCPs, for a given year. Average number of outages and mean outage duration (in minutes, secondary axis) from 10 CCPs (CME, DTCC, Eurex, ICC_CDS, ICE NGX, ICEU, ICUS_F&O, JSCC OTC-JGB, LCH.Clearnet.Ltd, LCH.Clearnet.SA) through the period 2016-2020. Observations that reported zero outages have been excluded from the analysis. Data reported by CCPs only include total outage time and number of outages; this implies that for periods for which there is more than one outage we can only infer the mean outage time. Distribution of outage times and other descriptive statistics such as standard deviation, maximum and minimum outage times cannot be inferred from the data provided.

We use the reported number of outages and outage duration to set values for λ and μ. Denoting total time in outage as *X*, total time in the observation period as *T*,[10] and number of outages as *N*, we calculate the following:

$$Mean\ time\ between\ outages\ per\ firm = \frac{T\ -\ X}{\left(\frac{N}{n}\right)}\ =\ 704\ hours \tag{33}$$

$$Mean\ time\ per\ outage\ = \frac{X}{N}\ =\ 1.2\ hours \tag{34}$$

It will be convenient to work in a basis of hours. Using equations (2) and (3) to estimate the values of mu and lambda in hours we obtain:

$$\lambda\ = \frac{1}{Mean\ time\ between\ outages\ in\ hours}\ =\ 0.14\%\ per\ hour \tag{35}$$

$$\mu\ = \frac{1}{Mean\ time\ of\ an\ outage\ in\ hours}\ =\ 83\%\ per\ hour \tag{36}$$

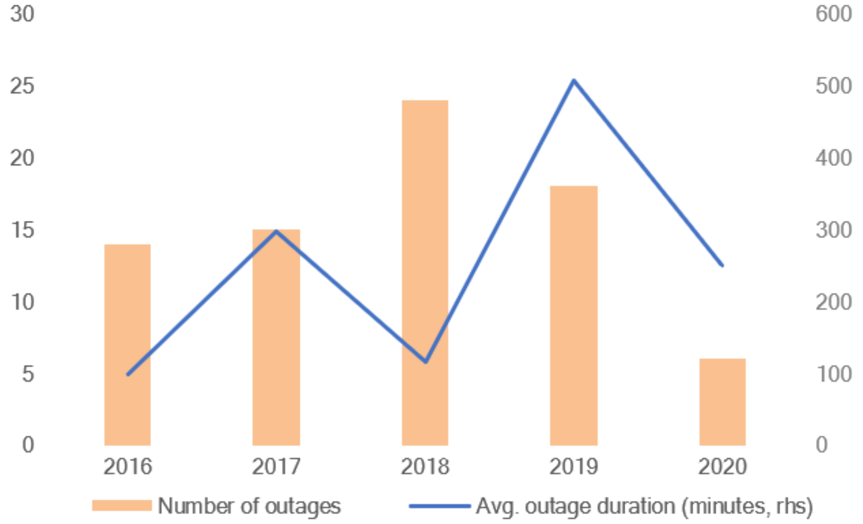
Frequency and duration of outages for CSPs

To apply the model to the CSP network, the different parameters need to be calibrated. Calibration in this context is made easier given that CSPs publicly disclose data on frequency and duration of outages. However, data are not reported in a consistent manner across CSPs. When looking at the three biggest cloud providers there are major differences on the amount and the quality of incident data published. Microsoft Azure only reports data after November 2019 and does not distinguish between ‘outage’ and ‘disruption’. AWS only reports data of what they consider ‘major’ incidents, which amounts to a total of just 5 incidents since 2015. Google Cloud on the other hand, reports detailed service outage and disruption data since Q2 2015.

[10] We consider trading hours only. We assume there are 8 hours per trading day and 20 trading days in a given month, which implies a total of 60 trading days per quarter. This amounts to 480 trading hours per quarter, or equivalently 28,800 trading minutes. In our sample we have 18 quarters, which implies a total time of 518,400 minutes, or 8,640 hours (*T = 8,640).*

Based on this, to calibrate our model we use the outage data reported by Google Cloud, which is publicly available on their website. The disclosures include information on service outages and disruptions to Google Cloud services; we focus on the former, in keeping with the focus of this paper. Chart 10 provides summary statistics.

Chart 10
Average reported outages and average outage duration (in minutes) used as a proxy for outages of clearing members in the no-cloud baseline



Note: Number of outages and average duration of outages reported by Google Cloud for 2016-2020. Sources: Google Cloud, ESMA.

Table 3: CSP outage data

Year	2016	2017	2018	2019	2020	*All years*
Number of outages	14	15	24	18	6	*77*
Number of outages per service per year	0.9	0.9	1.5	1.1	0.4	*0.96*
Avg. duration (minutes)	99	298	116	507	250	*250*

A challenge in calibrating the cloud scenario in the model is to decide which of the 16 different cloud services listed in the dataset CMs would rely on for transactions with CCPs, if they were to outsource to the cloud as described by our model. There is no clear-cut answer, as the different cloud services can be used in different ways and in various combinations for a given area of business activity. Reflecting this uncertainty, our estimate for λ' is based on a simple average outage frequency across all cloud services.

Another major limitation is that our parameter estimates implicitly assume that if a CM were to rely on a particular cloud service, any outage reported as occurring with that service would directly affect the CM. In other words, the estimate does not consider the fact that some outages may be local, rather than global. This could severely bias upward our estimate for λ' in particular.

Following the same estimation procedure as in the case of the CCP data above, we arrive at the following parameter estimates:

$$\lambda' = \ 0.08\% \, per\ hour$$

$$\mu' \ = 24\% \ \ per\ hour$$

Recalling from Proposition 1 that the average time a firm spends in outage τ can be expressed as $\tau = \frac{\lambda}{\lambda+\mu}$, we can compare between the two scenarios:[11]

$$\tau = 0.17\%$$

$$\tau' = 0.34\%$$

In other words, under our calibration, when firms outsource their services to CSPs, their expected time in outage doubles.

Calibration of the systemic cost of outages

We define systemic events as outages that result in a substantial fraction of CMs being unable to operate. The intuition is that if large CMs or a multitude of smaller CMs are disrupted, then the CCP is unable to operate in an orderly manner since several counterparties would be unable to post and receive margins.

More precisely, we define a systemic event to occur whenever at least 3 of the same CMs are simultaneously unable to operate for at least 8 hours. This requirement is stricter than the one used for CCP stress tests, where CCPs should be able to withstand the simultaneous default of their two largest CMs. However, in our model and application we only focus on the number of firms suffering an outage irrespective of their size. Therefore, we counterbalance this effect by requiring 3 CMs to suffer an outage[12].

Regarding the duration of the outages, the Principles for Financial Market Infrastructures (FMIs) put forward by CPMI and IOSCO explicitly specify that FMIs should have a business continuity plan that ensure that critical IT systems are able to resume two hours after a disruptive event (CPMI-IOSCO, 2012). The analysis of section 3.3 first imposes a minimum-time condition of 2 hours, i.e. defines a systemic event to occur whenever 3 firms are simultaneously in outage for at least 2 hours.

The longer the duration of the outage, the higher the probability that the event will be systemic. Any event that prevents or impairs end-of-day settlement could then be considered systemic (Brauchle, Göbel and Seiler, 2020). Therefore, we also analyse what happens when we impose an 8-hour minimum for a simultaneous outage to count as systemic. This condition reflects the fact that clearing is on a T+1 basis, and 8 hours is the approximate length of a trading day.

Finally, alongside both sets of results we examine the effect of relaxing the minimum-time condition, i.e. suppose that a systemic event simply occurs whenever at least 3 of the same CMs are simultaneously out. The resulting comparison of results gives insight into the role played by the recovery rate parameter μ' in mitigating systemic risk.

4.3 Results

Given the definition of systemic event in the present application, we wish to calculate

[11] Although the estimated value of τ' is only an approximation given assumptions in the estimation strategy and limited data availability, it is close to Google's target of availability for zonal services of $\tau' = 0.1\%$. For more details on this availability target, see section 3.4.1.

[12] An extension of the model where systemic events are defined based on size could be analysed in future work.

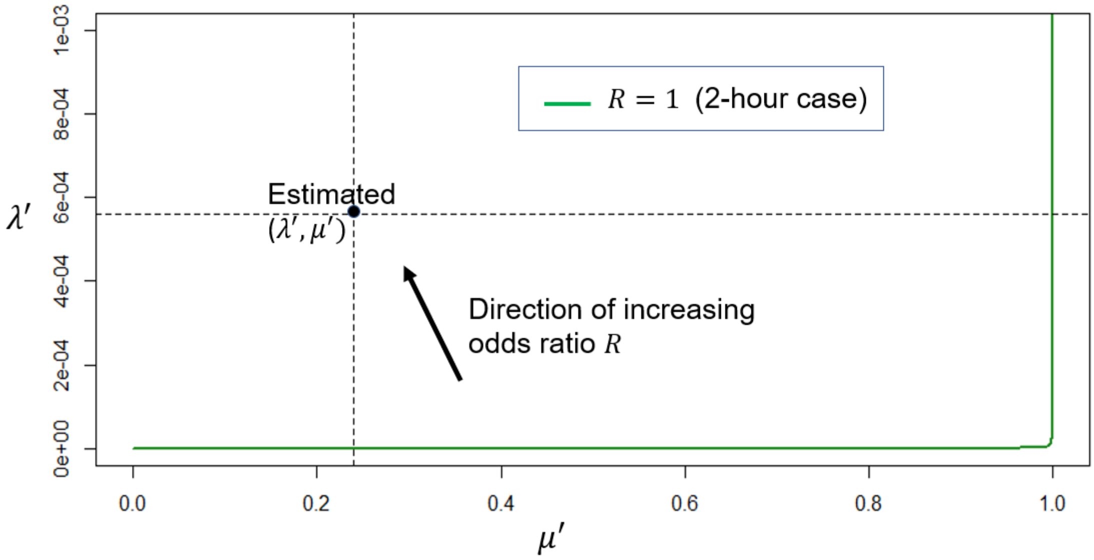
$$\text{Prob}[m_8(t) \geq 3 \mid No\ cloud] = 1 - \text{Prob}[m_8(t) \leq 2 \mid No\ cloud] = 1 - B_n(2, (1-\mu)\tau)$$

and

$$\text{Prob}[m_8(t) \geq 3 \mid Cloud] = 1 - \text{Prob}[m_8(t) \leq 2 \mid Cloud] = 1 - B_n(2, (1-\mu')^7\tau')$$

Using the parameter values for λ, λ', μ and μ' estimated in section 4.2 yields the solutions for $R = 1$ plotted in Chart 11, under the specification that a systemic event requires the same 3 firms to have a simultaneous outage for at least 2 hours.

Chart 11
Estimated incident and repair rate for cloud outsourcing compared with solutions for R = 1 using data-based calibration



Note: The lines plot values of cloud incident rate λ' and cloud repair rate μ', expressed as per-hour quantities, for which systemic events have the same probability in the no-cloud baseline and the cloud scenario, given the parameter estimates for λ and μ based on CCP outage data. A systemic event occurs whenever the same 3 firms are out simultaneously for at least 2 hours. The y-axis is truncated at λ' = 0.1% for clarity.

The purpose of the analysis is not to provide accurate point estimates of the relative risk of systemic events between the two scenarios, given the limitations in the data discussed above and the stylized features of the model such as independence of outages across firms (Assumption 1) and the specification that an outage affecting 3 firms is the threshold for a systemic event. However, the results give a useful framework for further analysis.

The precise parameter values of CSP outage probability λ' and recovery probability μ' that we infer from the available data (using the assumptions discussed in section 3.2) are approximate estimates only. Nonetheless, as order-of-magnitude estimates they appear to be *plausible*, in that they are close to the target values adopted by the CSP in question. Given that these plausible values of (λ', μ') lie far above the risk-equalization ($R = 1$) lines, we conclude that $R > 1$ in the present application. In other words, given the available data, our model suggests that outsourcing of core services by clearing members could create a new source of systemic risk, through simultaneous operational outages.

Consequently, as financial sector firms adopt cloud outsourcing for core functions, policymakers should investigate the possibility that additional systemic risk arises. They may do this by:

- seeking and collecting more comprehensive data on outages by clearing members (and indeed CSPs), or by other firms for whom simultaneous outages may have systemic effects; and
- investigating the extent to which the modelling assumptions hold in practice and adjusting the modelling accordingly

Chart 11 indicates that in the most time time-critical applications – where one hour of simultaneous outage represents a systemic event – then there is a linear trade-off between the cloud incident rate and cloud repair rate in equalising risk with the no-cloud baseline.

In plotting Chart 11, we have taken 2 hours to be the time threshold for a systemic event. However, it could be argued that the systemic effects of an outage are less time-critical than that. For instance, we could instead assume that CCP outages have systemic effects only after 8 trading hours, given the T+1 clearing cycle. Using an 8-hour minimum makes the probability of a systemic event in the no-cloud baseline vanishingly small in our model for the parameter estimates based on CCP outage data.[13] The implied probability of λ' for $R = 1$ would accordingly be vanishingly small – in effect requiring CSPs to prevent outages with perfect reliability.

In summary, where systemic events occur only after extended periods of simultaneous outages among firms, our modelling suggests that CSPs would need *perfect* service availability so as not to introduce additional systemic risk compared to the no-cloud baseline. Achieving equality of systemic risk with the no-cloud baseline (the $R = 1$ line in Charts 10 and 11) is therefore effectively unattainable for CSPs in the case of an 8-hour minimum for systemic events. This finding illustrates certain limitations with the modelling, however:

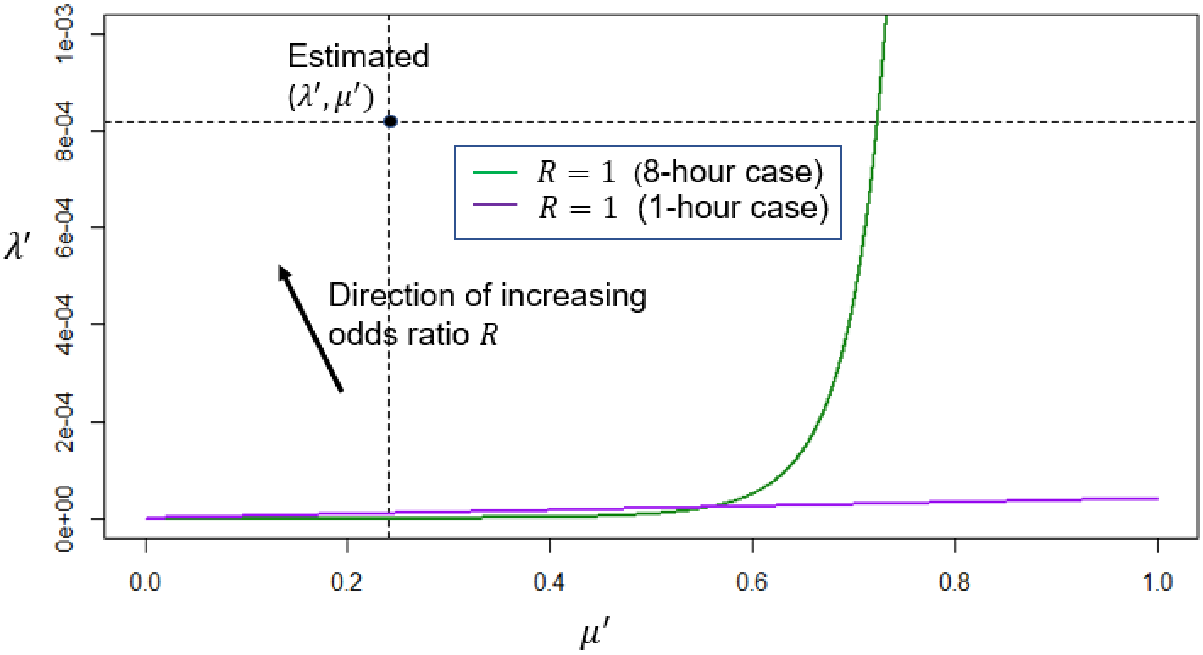
- Policymakers may wish to tolerate more than the level of vanishingly small risk implied by the no-cloud baseline, given other benefits of the cloud computing paradigm.
- The no-cloud baseline risk is based on simplifying assumptions, as set out above.
- The CCP outage data may not provide a true guide to firm-level outage duration. One issue is that the data report only total outage length per firm per quarter, rather than the length of each outage. This makes is hard to test the goodness-of-fit of the geometric decay implied by our modelling (as opposed to a fat-tailed distribution). In particular, the data do not identify the number of day-long outages among CCPs.

One way to address these limitations is to consider the values of λ' and μ' that are required to achieve a less extreme mitigation of systemic risk, while retaining the 8-hour minimum for systemic events. To do this, Chart 12 plots the $R = 1$ line while specifying that the repair rate in the no-cloud baseline is now equal to that implied by the CSP data. In other words, we now set $\mu = 0.24$, rather than $\mu = 0.83$. Under this calibration, we have that $\tau = 0.59\% > \tau'$. In other words, moving to the cloud now *reduces* the expected time a firm spends in outage. This models an incentive for firms to outsource, as explored in an equilibrium framework in Annex 1.

The hourly probability of systemic of a 3-firm outage in the no-cloud baseline is now of the order of 2 in 10,000, or roughly once every 2.5 years. Even with this more modest target for systemic risk, however, our model indicates that CSPs have a greater risk than the baseline of a simultaneous outage (whether for one hour or 8 hours).

[13] The hourly probability of a 3-firms outage would be around 2 in a million according to our model, based on the assumptions and calibration used, notably including Assumption 1 (independence of firm-level outages in the baseline). This implies it would happen less often than once a century.

Chart 12
Estimated incident and repair rate for cloud outsourcing compared with solutions for R = 1 assuming higher baseline risk



Note: The lines plot values of cloud incident rate λ' and cloud repair rate μ', expressed as per-hour quantities, for which systemic events have the same probability in the no-cloud baseline and the cloud scenario, given the parameter estimate for λ based on CCP outage data but a lower estimate of μ=24% (equal to that inferred from CSP data). Systemic event occurs whenever the same 3 firms are out simultaneously for at least 8 hours. The y-axis is truncated at λ = 0.1% for clarity.

Chart 12 suggests that starting with the estimates of λ' and μ' from the CSP data, systemic risk (relating to simultaneous outages of several hours) will be most effectively addressed by improving the cloud repair rate μ'. Doubling μ' will enable the systemic risk target to be met, while halving the incident rate λ' will not. If μ' is increased to around two-thirds, then a far higher outage frequency λ' can be tolerated without introducing systemic risk. Put simply, if cloud outages are almost always repaired within an hour or two, then even if they are relatively frequent, they will not introduce systemic risks that only emerge on a timescale of many hours. This qualitative result appears especially relevant from the perspective of policymaking and risk management.

4.4 Mitigating risk through back-up: multi-cloud outsourcing

A simple extension of the analysis examines the possibility that firms may have access to a backup cloud service – either from a different provider, or by the same provider such that the back-up version of a given service operates fully independently of the primary version (known as a ‘multi-cloud’ approach). The assumption of full independence is idealised, as in practice a single provider (or even separate providers) could be subject to a given cyberattack or financial crisis undermining their ability to do business, for instance. There may also be common, external dependencies among given infrastructures, for example via chain outsourcing.

Our focus is on a multi-cloud approach for a given core service, to address risk arising from concentration at system level, in contrast to a multi-cloud approach across services to address risk arising to the operations of a single firm from concentration within the firm.[14]

This risk mitigation strategy is already offered to some extent by some CSPs by constructing separate groups of cloud computing resources designed to be largely independent of each other, often known as 'zones'. Zones may be connected to each other within a geographical region. Services can be provided at regional level, meaning that even if one zone suffers an outage, the services are likely to remain in operation. For example, Google Cloud (2021) aims for each zone to achieve 99.9% availability (i.e. $\tau' = 0.1\%$) but aims for each region to achieve 99.99% availability (i.e. $\tau' = 0.01\%$).

We assume *perfect portability* in this scenario. This means that if one provider suffers an outage and a firm has a back-up provider, the service can switch frictionlessly between the two. In reality however, back-up strategies, especially those involving different cloud providers, may face major frictions, in contrast to our idealised model. In other words, cloud services may not be fully or instantly portable between different providers, for reasons including the different configurations and technological solutions used. The same considerations are likely to apply to a lesser extent when considering availability zones within a single provider.

Another pertinent issue not modelled here is that multi-cloud solutions may have some impact on the recovery rate parameter μ.

General approach to modelling multi-cloud for a given core operation

For simplicity, we consider the case of back-up via a different cloud provider. In the general n-firm case, for any two providers A and B, A is primary provider to $\frac{n}{n'}$ firms and B is secondary provider to a share $\frac{1}{n'-1}$ of these firms. Likewise, A is secondary provider to $\frac{1}{n'-1} \times \frac{n}{n'}$ firms whose primary provider is B.

The number of firms suffering an outage at any given time is then:

$$m(t) = \frac{n}{(n'-1)n'} \sum_{j=1}^{n'} \sum_{j' \neq k}^{n'-1} \omega_j(t)\, \omega_{j'}(t) \tag{37}$$

Remark: Given that

$$\sum_{k' \neq k}^{n'-1} \omega_{k'}(t) \leq n' - 1 \tag{38}$$

it follows that $m(t)$ is never larger in scenario 3 than scenario 2. There are always fewer firms in outage under the back-up scenario than under the initial cloud scenario.

Multi-cloud in the securities markets application

Suppose that each of the 20 firms in the application of section 3.2 now uses a multi-cloud model – specifically, using a back-up service from a different provider to seamlessly enable them to carry out CM functions if their primary CSP suffers an outage. As set out below, a key feature

[14] ESMA (2020c) includes guidelines for firms to assess concentration risk both at firm level and at sectoral level, and for competent authorities to monitor such risks once they are identified.

of this new scenario is that a systemic event (again triggered when 3 firms suffer simultaneous outage) now requires simultaneous outage among two providers, rather than one.

For simplicity, as in the general n-firm case we assume that providers' clients are equally shared by the other firms. This implies that just as in the primary market, the 4 cloud providers have equal market shares in the market for back-up services.

If just one cloud provider suffers an outage, then its client firms are instantly able to switch to the back-up service, and their operations are interrupted. If, two cloud providers suffer a simultaneous outage, then a third of the 5 client firms of each provider suffer an outage (since each is backs up one third of the market for the other firms), making a total of $\frac{10}{3}$ firms. Since the threshold for a systemic outage is $S = 3$, a systemic event now requires simultaneous outage by two cloud providers.

Assuming a 2-hour minimum for systemic outage, the odds ratio of scenario 2 (cloud outsourcing without back-up) compared with the no-cloud baseline is $R{\sim}10^3$. In other words, systemic risk is around a thousand times higher in the presence of cloud outsourcing.

In contrast, the odds ratio of scenario 3 (cloud outsourcing with back-up) is $R{\sim}1$, i.e. risk is reduced to around the level of the no-cloud baseline.

In summary, if firms back up their cloud services, the odds ratio decreases by several orders of magnitude. A multi-cloud model is a successful mitigant in the stylized model, based on the parameter calibration examined.

An important caveat to this result is that CSP outages are (like firm outages) assumed to be independent. Introducing positive correlation between CSP outages (stemming from example from common vulnerabilities) would weaken the effectiveness of a multi-cloud policy. Nonetheless, discussion with market participants suggests that CSPs are likely to have different cybersecurity strategies and measures, which limits the scope for common vulnerabilities from malicious actions. Additionally, the scope for common vulnerabilities to natural disasters is limited by geography, in a similar manner to Assumption 1 (independence of firm-level outages in the baseline).

4.5 Possible model extensions

There are several ways the model could be extended. However, as the precise results are dependent on the calibration, which is based on limited available data, at this stage we have restricted attention to studying a heavily stylized version, discussing key features of the results and dependencies between parameters. Further work is likely to be most fruitful if and when more comprehensive data become available.

A key simplification in the model is that firms are equally important from a systemic risk perspective. Future work could relax this constraint.

In the model, outage status follows a Markov chain, implying a geometric decay process for outages (where the constant per-period probability of outage resolution is μ). This feature of the model simplifies the analysis considerably. However, there is some evidence that outage duration follows a fat-tailed distribution (Naldi, 2017).[15] Further work could investigate this possibility through numerical simulations.

[15] This possibility is one reason we adjust downwards the baseline (in-house) repair rate μ in the analysis presented in Chart 12.

Another simplification is that systemic costs are assumed to be binary once a threshold number of firms suffer simultaneous outage. A natural alternative would be to consider a (weakly convex) systemic cost function, strictly increasing in the number of firms in outage. Our initial analysis of this problem is that such a cost function increases the odds ratio R.

A related extension would be to examine in detail how changes in the assumed market structure change the systemic risk profile of cloud outsourcing. For example, the effect increasing market concentration by CSPs should, intuitively, increase systemic risk. To capture this effect properly, it is likely that we would need to relax the assumption that systemic costs are binary and introduce a systemic cost function strictly increasing in the number of firms in outage.[16]

Finally, important simplifications in the model are Assumption 1 (independence of firm-level outages in the no-cloud baseline), and the assumption that CSP outages are likewise independent. Relaxing the former assumption would, other things equal, reduce the odds ratio R, while relaxing the latter would increase it. Further work could study these relationships in greater detail.

5 Conclusions

In a stylized framework, we have shown that individual firms have an incentive to outsource some of their IT infrastructure to CSPs. However, due to concentration risk, the likelihood of simultaneous outages might increase, thereby leading to higher systemic risk for the financial system.

We discuss several options that can be pursued to mitigate this risk. First, if CSPs are substantially more resilient than individual firms, systemic risk could decline as the additional resilience of using CSPs more than compensate concentration risk. Finally, multi-cloud solutions, where firms use one CSP and another one as backup – or alternatively, the successful provision of cloud services via independent groups of resources by the same provider – may significantly reduce systemic risk. This will only happen, however, if the different CSPs or groups of resources have low common vulnerabilities (i.e. can reasonably be treated as independent) *and* if the services in question are rapidly portable between them. In reality, the first of these assumptions (independence of CSP outages) may not hold in certain circumstances, especially within a single cloud provider, while the second assumption (back-up portability) may not hold especially for back-up strategies that use different providers.

Our work also shows the need for detailed data on outages by financial institutions and CSPs. Having consistent data reported by firms and CSPs would allow a better calibration of the model and improve the assessment of trade-offs between different uses of CSPs by firms.

Given the ubiquity of CSPs and continuing migration to use of their services – a trend accelerated by the COVID-19 pandemic – it is crucial for policymakers and market participants to assess benefits and risks of outsourcing to CSPs. An important example in the EU is the proposed Digital Operational Resilience Act that envisages a mandate for the European Supervisory Authorities, working with other authorities, to oversee third party providers of critical financial services to address related systemic risks (European Commission, 2020).

[16] In the presence of binary systemic cost, we can briefly consider the effect of decreasing the number of CSPs covering the market. Once a single CSP outage is enough to trigger a systemic event, then further decreasing the number of CSPs would in fact decrease systemic risk in the model. Unless the specification of a binary systemic cost is strongly motivated by a particular setting, a strictly increasing systemic cost function would be a better basis for investigating the relationship between market concentration and systemic risk.

References

Aldasoro, I. *et al.* (2020) ‘The Drivers of Cyber Risk’, *BIS Working Papers*, 865. Available at: https://papers.ssrn.com/abstract=3613173 (Accessed: 15 February 2021).

Bouveret, A. (2019) ‘Estimation of losses due to cyber risk for financial institutions’, *Journal of Operational Risk*, 14(2), pp. 1–20. doi: 10.21314/JOP.2019.224.

Brauchle, P., Göbel, M. and Seiler, J. (2020) ‘Cyber Mapping the Financial System’, *Cyber Policy Initiative Working Paper Series*, 6, p. 28.

CPMI-IOSCO (2012) ‘Principles for financial market infrastructures’. Available at: https://www.bis.org/cpmi/publ/d101.htm (Accessed: 2 March 2021).

CPMI-IOSCO (2015) *Public quantitative disclosure standards for central counterparties*. Basel: Bank for Internat. Settlements. Available at: http://www.bis.org/cpmi/publ/d125.htm (Accessed: 2 March 2021).

Danielsson, J. and Macrae, R. (2019) ‘Systemic consequences of outsourcing to the cloud’, *VoxEU.org*, 2 December. Available at: https://voxeu.org/article/systemic-consequences-outsourcing-cloud-0 (Accessed: 15 February 2021).

Demirer, M. *et al.* (2018) ‘Estimating global bank network connectedness’, *Journal of Applied Econometrics*, 33(1), pp. 1–15. doi: https://doi.org/10.1002/jae.2585.

Eisenbach, T. M., Kovner, A. and Lee, M. (2020) ‘Cyber Risk and the U.S. Financial System: A Pre-Mortem Analysis’, *Federal Reserve Bank of New York Staff Reports*, 909. doi: 10.2139/ssrn.3522710.

European Commission (2020) 'Proposal for a Regulation of the European Parliament and of the Council on digital operational resilience for the financial sector and amending Regulations (EC) No 1060/2009, (EU) No 648/2012, (EU) No 600/2014 and (EU) No 909/2014.' COM/2020/595 final.

ESMA (2020a) *3rd EU-wide CCP Stress Test*. Report. Paris, France. Available at: https://www.esma.europa.eu/sites/default/files/library/esma70-151-3186_3rd_eu-wide_ccp_stress_test_report.pdf.

ESMA (2020b) *EU Derivatives Markets*. Annual Statistical Report. Paris: European Securities and Markets Authority. Available at: https://www.esma.europa.eu/sites/default/files/library/esma50-165-1362_asr_derivatives_2020.pdf.

ESMA (2020c) *Final Report: Guidelines on Outsourcing to Cloud Service Providers*. Available at: https://www.esma.europa.eu/sites/default/files/library/esma50-157-2403_cloud_guidelines.pdf

FSB (2019) *Third-party dependencies in cloud services - Considerations on financial stability implications*. Available at: https://www.fsb.org/wp-content/uploads/P091219-2.pdf.

FSB (2020) *Regulatory and Supervisory Issues Relating to Outsourcing and Third-Party Relationships*. Discussion paper. Available at: https://www.fsb.org/wp-content/uploads/P091120.pdf.

Google Cloud (2021), *Architecting Disaster Recovery for Cloud Infrastructure Outages.* Available at https://cloud.google.com/architecture/disaster-recovery#introduction

Harmon, R., Vytelingum, P. and Babaie-Harmon, J. (2020) 'Cloud Concentration Risk: A Framework Agent Based Model For Systemic Risk Analysis'. doi: 10.13140/RG.2.2.17912.67841.

Lloyd's (2017) *Counting the cost Cyber exposure decoded*. Emerging Risk Reports. Available at: https://assets.lloyds.com/media/09c41bb0-d73b-4ae6-a2f8-d5cc41429998/pdf-emerging-risk-report-2017-counting-the-cost.pdf.

Lloyd's (2018) *Cloud down Impacts on the US economy*. Emerging Risk Reports. Available at: https://assets.lloyds.com/assets/pdf-air-cyber-lloyds-public-2018-final/1/pdf-air-cyber-lloyds-public-2018-final.pdf.

Naldi, M. (2017) 'Evaluation of Customer's Losses and Value-at-Risk under Cloud Outages', in. doi: 10.1109/TSP.2017.8075927.

Annex 1: Equilibrium where firms outsource but do not fully back up

To complement our discussion of options to lower concentration risk, we present a simple equilibrium framework that (i) accounts for the existence of the market for cloud-based outsourcing and (ii) illustrates how, nonetheless, firms may not find it individually rational to back up their services, even if doing so would be optimal from a system-wide perspective. This suggests that policy intervention may be needed to correct a form of market failure.

In this framework, firms each pay a fee to outsource cloud services to a single market-wide provider. This fee is paid in each period and equals $K' > 0$ per unit of time.

We also allow for the fact that cloud outsourcing saves costs of maintaining in-house services. To the extent that cloud services are scalable on demand, they may also reduce excess capacity costs, which are another form of in-house costs. Denote K the sum of all in-house costs.

While it is simplest to interpret K and K' as costs, they could alternatively be interpreted more holistically to reflect relative incentives that are not immediately financial in nature, e.g. the business or technological capabilities offered by cloud services versus in-house systems,

Assumption 4: Firms are risk neutral and suffer individual costs due to outages equal to the total amount of time of the outage.

Let $\delta \in (0,1)$ denote the common discount rate, so that firms discount future expected costs by a factor of $e^{\delta t}$. Assumption 4 implies that it is ex-ante optimal for any given firm to outsource services to a cloud provider when the following inequality holds.[1]

$$\int_{t=0}^{\infty} (\tau + K) e^{-\delta t}\, dt \geq \int_{t=0}^{\infty} (\tau' + K') e^{-\delta t}\, dt \qquad (A1)$$

By inspection, this is true whenever the following inequality holds:

$$K \leq \tau - \tau' + K \qquad (A2)$$

We suppose that the cloud provider is a price setter, selecting a value of K to maximise total revenue. Firms then choose whether to pay the fee for the outsourced services. By symmetry, the cloud provider needs to select a price that all firms are willing to pay. Proposition 3 characterises the resulting equilibrium.

Proposition 3. In scenario 2, there exists a unique value of K' in which firms and the cloud provider are in a (unique, subgame-perfect) Nash equilibrium: $\boldsymbol{K' = \tau - \tau' + K}$

Proposition 3 implies that cloud migration is (weakly) optimal for each firm. The reason is that the cloud provider provides the outsourced services with less average outage time than the firm would be if the services were instead in-house.

Among the right-hand side variables in the expression for K' in Proposition 2, while τ and K_{in} are exogenous to the provider, the value of τ' can be reduced if the provider is able to reduce λ' or increase μ' Intuitively, firms will pay a premium for service reliability.

However, as discussed in section 2, a systemic cost may arise in section 2 if all firms suffer an outage simultaneously. For simplicity, assume that this cost is proportional to the number of firms in the market, so can be denoted. nC_{system}, where $C_{system} > 0$ is a constant. The systemic costs are binary in nature: if the number of firms suffering a simultaneous outage reaches some threshold S, then the systemic cost arises.

We assume that C_{system} represents spill-over costs for firms. For example, if several firms suffer an outage, the impact on liquidity for all firms creates costs beyond those directly incurred by the firms subject to outage as they cease their business activities. As such, K_{system} does not appear in the equation in Proposition 2.

Scenario 2 is less efficient (more costly) than the baseline no-cloud scenario from a system-wide perspective if the following inequality holds.

$$qnC_{system} + n\tau' > pnC_{system} + n\tau \tag{A3}$$

where q is the probability of systemic outage in scenario 2 and p is the probability of systemic outage in scenario 1. Rearranging gives

$$(q{-}p)C_{system} > \tau - \tau' \tag{A4}$$

Proposition 3 showed that it is individually rational for all firms to outsource to the cloud, assuming a provider is available. The reason is that $\tau - \tau' > 0$, i.e. cloud providers offer a more secure and resilient service. The right-hand side of the inequality (A4) is therefore positive. Expected total net costs (excluding K', which is a payment rather than an economic cost to the system) can nonetheless be higher in scenario 1 than scenario 2 if (i) the systemic cost parameter C_{system} is large enough and (ii) cloud provision yields a large enough increase $(q-p)$ in the probability of a mass outage (of at least S firms) compared with the no-cloud case.

Proposition 3 can be used to guide risk analysis in the context of large firms gaining market share in the cloud services sector, which is especially relevant given the market concentration highlighted above in section 2. Specifically, it relates to risks analysis around large providers serving critical financial sector firms). As a provider grows, we may expect its value of τ' to decrease due to economies of scale, thereby increasing the right-hand side of the inequality. However, at the same time the left-hand side of the inequality can also be expected to increase, as the firm's value of α will rise if it serves more of the population of firms. Consequently, the effect of a firm gaining market share is *a priori* ambiguous.

We now turn to analysis of the impact of cloud outsourcing on expected system-wide net costs. Let $q_{multi} < q$ denote the probability of systemic outage in scenario 3.

For firms seeking back-up, the marginal cost compared with scenario 2 is K' (as opposed to $K' - K$, the marginal cost of entering scenario 2 versus scenario 1), assuming that the same price is offered to firms whether or not a cloud service is used as back-up. This would be necessitated by assuming for instance that CSPs provide their services at cost.

Now suppose that $\tau < 2\tau'$, i.e. cloud-based outsourcing less than halves the average outage time of in-house provision. This is a sufficient condition for the marginal benefit individual marginal benefit to the firm from obtaining back-up to be less than $\tau - \tau'$, the marginal individual benefit of seeking simple cloud (scenario 2) compared to in-house provision (scenario 1).

Substituting these bounds on marginal costs and benefits into the equality in Proposition 3 yields

$$\tau < K' \tag{A5}$$

i.e. the marginal benefit to a firm of obtaining back-up is less than the marginal cost, and hence firms choose not to back up their services.

In conclusion, there exist parameter values for which scenario 2 is in equilibrium, but scenario 3 is not.

esma European Securities and Markets Authority