CYBERSECURITY AND FINANCIAL STABILITY

KARTIK ANAND, CHANELLE DULEY, AND PRASANNA GAI

ABSTRACT. Cyberattacks can trigger depositor runs when vulnerabilities in a bank's IT systems are exploited to impair its ability to service withdrawals. A fundamental trade-off between protection and resilience shapes a bank's private incentives to invest in cybersecurity. The socially optimal choice of cybersecurity depends critically on the severity of the threat posed and the nature of bank fragility, and regulators need to be alert to these considerations when designing ex ante and ex post regulation. We show how operational resilience standards, red-teaming, subsidies for enhancing cyber-capability, and negligence rules with penalties may facilitate socially desirable investment in cybersecurity. Our analysis is robust to the introduction of ambiguity-averse investors, a lender-of-last-resort, multiple banks sharing common IT platforms, and technology vendors.

Keywords: Cybersecurity, bank runs, global games, public good provision.

JEL classifications: G01, G21, G28, H41.

We thank Toni Ahnert, Mei Dong, Thomas Eisenbach, Sayantan Ghosal, Tetsuya Hoshino, Charlie Kahn, Philipp J. König, Stephen Morris, Silvio Petriconi, Hyun Shin, Xavier Vives and participants at the Bank for International Settlements Research Seminar Series, CAFRAL Webinar, the Deutsche Bundesbank Research Brown Bag, the European Banking Theory Brown Bag Webinar 2021, IFABS Conference, Oxford, 2021, Ridge Virtual Forum 2021 Workshop on Financial Stability, Montevideo, the 2021 European Winter Meeting of the Econometric Society, Barcelona, RiskLab/BoF/ESRB Conference on Systemic Risk Analytics 2022, Helsinki, ESRB Mini Workshop of Cyber risks 2022, Frankfurt, Workshop on Financial Stability 2022, Halle, CEMLA 2022 Workshop on Financial Stability, Mexico City, and ESRB Task Force on Stress Testing Workshop 2023, Luxembourg for helpful comments. All remaining errors are our own. This paper represents the authors' personal opinions and does not necessarily reflect the views of the Deutsche Bundesbank or the Eurosystem. Anand: Deutsche Bundesbank, Research Department, Wilhelm-Epstein-Strasse 14, 60431 Frankfurt, Germany. Email: kartik.anand@bundesbank.de. Duley and Gai: University of Auckland, 12 Grafton Rd, Auckland 1010, New Zealand. Email: chanelle.duley@auckland.ac.nz and p.gai@auckland.ac.nz.

1. INTRODUCTION

Overview. From mobile phone-based solutions for customers to virtual data rooms automating due diligence processes, modern banks bear little resemblance to the brick and mortar institutions of yesteryear. But as the digital transformation in banking has gathered pace, so too have cyber risks.[1] Cyberattacks have the potential to trigger bank runs (Duffie and Younger, 2019) and disrupt wholesale funding markets (Eisenbach et al., 2022). In light of these threats to financial stability, some policymakers advocate stress-tests to assess banks' resilience in the event of cyberattacks (European Systemic Risk Board, 2022). Others endorse threat-led penetration or "red team testing" to find and eliminate vulnerabilities in banks' IT systems to boost protection against cyber violation (G7, 2018). But despite policymaker interest, our theoretical understanding of how cyberattacks influence banks' security choices and the risk of bank runs,, and shape regulation is limited. Our paper seeks to fill this gap.

Cyberattacks occur when a malicious agent takes deliberate action to exploit vulnerabilities in a bank's IT system to either gain unauthorised access to data or disrupt critical processes. Malign attackers vary in the sophistication and threat that they pose – from state-sponsored actors with deep-pockets to opportunistic hackers with less know-how and few resources. A bank can protect itself from a cyberattack by investing in cybersecurity, i.e., taking measures to detect and address vulnerabilities. Our analysis highlights that a bank's investment in cybersecurity is subject to a fundamental trade-off – between increasing protection against a cyberattack versus remaining resilient in the face of such an event. Central to this trade-off is the idea that cybersecurity is an intangible asset that generates future economic value by thwarting cyberattacks but is itself illiquid and cannot be sold or traded (Dell'Ariccia et al., 2021; Gatzert and Schubert, 2022; Lev and Radhakrishnan, 2005). Thus, while diverting more resources towards cybersecurity lowers the probability of a successful attack, it comes at the expense of reduced investment in profitable

[1]ENISA (2023) reports a sharp increase (22%) in the volume of distributed-denial-of-service (DDoS) attacks on financial institutions during 2022, and an industry survey from IT security company Sophos reports that the rate of ransomware attacks in financial services has risen consistently from 34% in 2021, to 55% in 2022, and up to 64% in 2023 (Sophos, 2023).

activities.[2] With fewer assets available to service depositors and a lower equity value, a bank is less able to honour its claims to depositors in the event of a successful attack.

The protection-resilience trade-off sheds light on the optimal policy response to financial stability threats posed by cyberattacks. Optimal policy is shaped by the sophistication of the attacker as well as the nature of bank fragility. When the risk of a bank run is high and cyberattacks precipitate funding illiquidity, the bank fails to properly account for the social benefits from greater protection against cyberattacks, leading to under-investment in cybersecurity. Moreover, the extent of the under-investment grows as the attacker becomes less sophisticated. In such circumstances, policymakers should place greater emphasis on enhancing the cybersecurity capabilities of the bank ex ante. This might include, for example, promoting red team tests to identify potential weaknesses or subsidising a bank's investment in cybersecurity. Ex post negligence rules that penalise banks for failing their duty of care may further encourage the adoption of preventative measures.

When the risk of bank runs is low and bank failure is insolvency-driven, the relationship between optimal policy and attacker sophistication is more subtle. If the attacker is highly capable in finding and exploiting vulnerabilities, the bank over-invests in cybersecurity as it fails to internalise the social benefits of greater resilience. Ex ante measures that shore up a bank's operational resilience in the event of a cyberattack are optimal. The 2024 cyber stress-test by the ECB – conducted during a period of heightened geo-political tensions with state-sponsored actors being accused of cybercrimes – is one such example.[3] The aim is to assess how supervised banks will respond and recover from a cyberattack (European Central Bank, 2024). But when the attacker's capabilities are poor, a bank's investment in cybersecurity is more likely to succeed in identifying and mitigating vulnerabilities. Since the social benefits of greater protection are not internalised, there is under-investment in cybersecurity. This, in turn, calls for measures to shore up cybersecurity capabilities.

Approach and Main Results. Our baseline analysis incorporates cyberattacks and the provision of cybersecurity into a global game model of bank runs. A representative bank obtains funding

[2]JPMorgan Chase considers cybersecurity practices "run-the-bank priorities" and "change-the-bank" investments. The bank allocates an annual budget of $12 billion, or one third of the value of its annual investments, to technological updates, including cybersecurity protection (Warren-Kachelein, 2022).
[3]Business continuity plans and cyber hygine notices also fall into this category.

from risk-neutral depositors by issuing uninsured demandable debt claims for profitable, risk-free, liquid investments. The bank uses IT systems to manage investments and is solely responsible for its security. But these systems have vulnerabilities that are unknown to the bank (Perlroth, 2021). A malicious agent can expend costly effort to discover and exploit these vulnerabilities. The bank, in turn, chooses how much to invest in protection to uncover and address vulnerabilities. The attacker and bank, are, therefore, engaged in a strategic contest (Dixit, 1987).

A successful cyberattack causes a temporary outage to the bank's IT systems, impairing its recourse to liquidity. This can lead to deposits being withdrawn early, precipitating a bank run. And, even after the resumption of services, the bank may suffer permanent losses that further undermines its ability to service deposits. Bank failure may, thus, be illiquidity- or insolvency-based. We pin down a unique equilibrium where a run occurs if the outage exceeds a certain threshold which depends on the bank's investment decisions (Morris and Shin, 2003).

Central to the bank's ex ante choice of cybersecurity is the trade-off between protection and resilience. The more that the bank contributes to cybersecurity, the more likely it is that vulnerabilities in its IT systems are discovered and mitigated. But this comes at the expense of profitable investments. Conditional on the cyberattack being successful, the bank has fewer resources at its disposal to service depositors, which can precipitate a bank run. Our model, thus, shows how ex ante cyber-risk management is shaped by the shadow of rollover risk ex post.

The comparative static results of our model define the 'cyber threat landscape' in terms of the sophistication of the attacker, the long-term disruption caused, and the nature of bank fragility. As the attacker becomes less sophisticated, the bank's chances of finding and mitigating vulnerabilities improves. This has two distinct implications. First, the bank contributes more towards its protection at the expense of greater fragility in the event that the attack is successful. And second, since the bank is deemed to be less risky, the face value of debt sought by investors decreases, which induces an opposing effect in equilibrium. We show that the direct effect dominates as long as the investor is highly sophisticated. The more enduring is the disruption of a successful cyber-attack, the lower are the benefits from greater financial resilience. Moreover, the difference in the

bank's expected equity value when attacks are unsuccessful and successful becomes larger. These features reinforce the benefits to protection leading to an increase in the bank's investment in cybersecurity. And, increased rollover risk implies that a bank is more likely to fail due to illiquidity in the event of a successful attack. The marginal benefits to protection are higher relative to the marginal benefits from greater resilience.

Socially optimal investment in cybersecurity depends critically on the nature of bank fragility, i.e., whether the attack triggers illiquidity or insolvency. In making its private security choice, the bank does not internalise the social cost of failing. These social costs include, for example, the loss of payment services or the contraction in credit to the real economy. Consequently, when failure is insolvency-driven, the bank over-invests in cybersecurity because it does not take into account the social benefit of greater resilience in the event of a cyberattack. But when failure is driven by illiquidity, the bank under-invests in cybersecurity. While the threat of the run strengthens incentives to invest in protection, the social benefits of greater protection are not taken into account, inducing under-investment in cybersecurity.

Although our main analysis is couched in terms of risk-neutral investors and a single, representative, bank that is responsible for its protection, the core insights are preserved once these assumptions are relaxed. Specifically, we extend the model to allow for (a) ambiguity-averse investors facing Knightian uncertainty about "zero-day" vulnerabilities – security flaws previously unknown to exist; (b) asset illiquidity and lender-of-last-resort support by the authorities; (c) multiple banks and common IT solutions; and (d) the introduction of a technology vendor who contributes to bank cybersecurity. Our results shed light on how regulatory interventions may need to be modified in such circumstances to achieve socially optimal outcomes. Specifically, we highlight the importance of regulators mapping out the contours of common IT infrastructure across banks, sharing information about vulnerabilities, and setting minimum cybersecurity standards for technology vendors.

Related literature. The theoretical literature on cyber risks and financial stability is yet to take hold, and our paper is an early contribution to this topic. The existing literature in the area is largely

empirical in nature. Duffie and Younger (2019) describe cyber-runs and conduct a stress test to understand the resilience of systemically important US banks to wholesale depositor withdrawals following a cyberattack.[4] Eisenbach et al. (2022) examine how cyberattacks impair a bank's ability to repay withdrawing creditors and discuss how this influences creditors' incentives to run. They suggest that, since cyberattacks impair the ability of the bank to repay early withdrawals, the first-mover advantage of creditors is weakened. The sequential service constraint means that creditors who withdraw face a lower probability of being repaid in full. Our analysis complements this argument. In our model, whenever rollover risk is low, bank failure following a cyberattack is not driven by coordination failure but instead by deadweight losses. In this case, since the impairment to banks' ability to repay is permanent there is no scope for an inefficient run. But when rollover risk is large, the impairment suffered by a bank following a cyberattack is more transient, and inefficient runs are a source of bank failure.

Our paper informs the growing policy literature in this area (Kashyap and Wetherilt, 2019; Adelmann et al., 2020; Elestedt et al., 2021; Fell et al., 2022). These papers recognise that cyberattacks can impair financial stability and note more should be done to bolster banks' resilience in the face of the attacks. Fell et al. (2022) argue that policymakers should improve their monitoring of cyberattacks and expand their macroprudential toolkits to cover cyber risks. Elestedt et al. (2021) suggest that better coordination and cooperation between the financial sector and relevant agencies responsible for cybersecurity is vital to financial stability. Our analysis provides a formal framework from which to explore such policy concerns and identifies conditions in which certain tools may be more appropriate than others.

[4] Jamilov et al. (2021) is a related contribution cyber risk and contagion. They construct a text-based measure of cyber risk, which they use to test the link between balance sheet and income statement information from publicly-listed companies and their exposure to cyber risk. They also explore whether cyber risk exposure influences asset pricing, and whether the effects of this exposure propagate to unaffected peer firms. They demonstrate that cyber risk exposure has a negative and significant effect on stock returns of affected firms and find evidence that cyber risk is a source of systematic risk in financial markets due to contagion effects or firm-to-firm networks. See also Kamiya et al. (2021), Woods et al. (2021) and Florackis et al. (2020).

There are also points of contact with the economics of security.[5] Gordon and Loeb (2002) consider a one-period model of a firm choosing how much to invest in IT security, given an exogenous threat probability. They argue that, since the firm's investment depends on the marginal product of security investment, it may be optimal for the firm to invest very little or nothing at all. Varian (2004) and Grossklags et al. (2008) extend this analysis to the case of multiple firms with network externalities where cybersecurity exhibits properties of public goods. Varian (2004) shows that underprovision of cybersecurity at the system level can be rectified using negligence rules. Grossklags et al. (2008) examine how the option to invest in insurance against damages following a cyberattack influences decisions to contribute to cybersecurity. Our contribution to this literature shows how ex post coordination failures in the form of bank runs and the cyber-threat landscape shapes banks' incentives to contribute to the public good of cybersecurity.

Finally, we add to the large literature on bank runs and global games (Morris and Shin, 2003; Goldstein and Pauzner, 2005). We specifically build on Rochet and Vives (2004), where unsecured debt holders delegate their rollover decisions to professional managers, so the decisions to rollover are global strategic complements. Our contribution shows how cyber risk management interacts with run risk in such a setting.

2. Model

Agents, preferences, and endowments. There are three dates, $t = 0, 1, 2$. A single good economy comprises a representative bank, a unit mass of investors, and an "attacker" intent on causing harm to the bank. All agents are risk-neutral. Whereas investors are indifferent between consumption at $t = 1$ and $t = 2$, the bank and the attacker care only about consumption at $t = 2$. We suppose that the attacker is deep-pocketed and that each investor is endowed with a unit of the consumption

[5] In this context, our analysis also relates to attack and defender games in the economics literature. This literature focuses on how the incentives of attackers and defenders depends on the network structure (Bier et al., 2007; Dziubinski and Goyal, 2013; Goyal and Vigner, 2014; Acemoglu et al., 2016). Our work abstracts from the network perspective and highlight the financial stability and cybersecurity aspects that arise from the strategic interplay between an attacker and a defender.

good.[6] Investors also have access to a risk-free storage technology that yields $r > 1$ per unit of investment. Without loss of generality, we normalise the bank's endowment to zero and assume that it is subject to limited liability.

Bank balance sheet. The bank has access to a safe and liquid asset at $t = 0$ that provides a return $R > r$ per unit of investment at $t = 2$.[7] To finance this investment, the bank issues $D > 0$ of uninsurable demandable debt claims to investors. Let $F > 0$ be the face value of debt and suppose that it is independent of the withdrawal date. The bank invests $I \leq D$ in the asset and allocates the remainder, $S \equiv D - I$, to its *cybersecurity*. Table 1 summarises the balance sheet at $t = 0$.

Assets	Liabilities
$I = D - S$	D

TABLE 1. Balance sheet at $t = 0$.

IT systems and cybersecurity. Banks use various IT systems to manage investments. These include software to manage liquidity and hardware solutions to securely store confidential data. In our core model, the bank is solely responsible for the security of these systems, i.e., finding vulnerabilities and mitigating them.[8] But if the attacker discovers these vulnerabilities, it can exploit them for private gain and disrupt the bank. We therefore treat the interaction between the attacker and the bank as a strategic contest (Dixit, 1987).

Cybersecurity bears the hallmarks of an intangible asset. As we discuss below, investing in cybersecurity improves the bank's chances of thwarting cyberattacks, which has economic benefits for the bank.[9] Moreover, these investments can readily be accounted for on the bank's balance

[6] Recent analysis suggests that cybercriminals have amassed considerable wealth in recent years to support their cybercrime capabilities (MIT Technology Review, 2022). State-sponsored hackers can also be assumed to have deep pockets.

[7] The introduction of asset illiquidity does not materially change our results but comes at the cost of expositional clarity. We extend our analysis to consider asset illiquidity in Section 5.2.

[8] In Section 5.4, we explicitly consider a third-party vendor that provides IT systems and also contributes to cybersecurity.

[9] Gatzert and Schubert (2022) find a positive and significant impact of cyber risk management on banks' firm value as measured by Tobin's Q, with banks that actively manage cyber risk being valued almost 11% higher than those without formal cybersecurity practices.

sheet. But, the bank cannot trade or sell its cybersecurity, which is ingrained in, and inalienable from, the institution (Crouzet et al., 2022).

cyberattacks and bank failure. The outcome of the contest between the bank and the attacker is determined at the start of $t = 1$. With probability

$$p(A, S) \equiv \frac{S}{A + S}, \tag{1}$$

the bank is successful in identifying and resolving the vulnerability. While with probability, $1 - p(A, S)$, the attacker is successful and inflicts a successful cyberattack. We consider the payoffs in both scenarios in turn.

If the bank is successfully able to address the vulnerability, the cyberattack is thwarted. In this case the attacker receives nothing, while the bank obtains the equity value, $RI - FD$, at $t = 2$ after investors are repaid. But if the attacker discovers the vulnerability, it deploys malicious code that causes glitches to the bank's IT systems and temporarily impairs its recourse to liquidity.[10] Specifically, the bank is subject to an impairment shock, $\alpha \in [0, 1]$, which is a uniformly distributed random variable drawn at $t = 1$. When $\alpha = 0$, the impairment does not impact on the bank's recourse to liquidity, while $\alpha = 1$ entails a complete denial to liquidity. The impairment takes place at $t = 1$ and is resolved by $t = 2$. The attacker obtains a payoff V at $t = 2$ from a successful attack.

By impairing the bank's recourse to liquidity, the shock can precipitate bank runs (Duffie and Younger, 2019). If a fraction $\ell \in [0, 1]$ of debt is withdrawn at $t = 1$, the bank fails due to *illiquidity* whenever

$$(1 - \alpha)RI - \ell FD < 0, \tag{2}$$

[10] The cyberattack on the New Zealand stock exchange in December 2020, which prevented the posting of market announcements and led to trading being suspended for several days is an exemplar (Tarabay, 2021).

i.e., the liquidation value of available assets is insufficient to service withdrawals. So whenever $\alpha > \alpha^{IL}(\ell) \equiv 1 - \ell\frac{FD}{RI}$, the bank fails and its equity value is wiped out.[11] Depositors are assumed to face a zero recovery rate upon bank failure.

Cyberattacks can also have longer lasting repercussions. These include, for example, the loss of secret information pivotal to the bank's role as a financial intermediary, losses from paying ransom demands, and even physical damage to IT systems. The credit downgrading of the Maltese bank, Valletta PLC, following a cyberattack and concerns over the bank's operational risk management illustrate how cyberattacks can threaten bank solvency (S&P Global Market Intelligence, 2019).

Our model reflects this possibility by assuming that the bank is subject to a deadweight loss proportional to the shock. After a successful attack subsides and the bank regains access to its IT systems at $t = 2$, its investments yield $(1 - \delta\alpha)RI$, where $\delta < 1$ reflects deadweight losses incurred due to the cyberattack. The larger the deadweight loss, the greater are the bank's losses and the strain on its solvency.

The bank fails due to *insolvency* at $t = 2$ when

$$(1 - \delta\alpha)RI - \ell FD < (1 - \ell)FD, \tag{3}$$

i.e., the gross returns from the asset are insufficient to repay the total debt claims against the bank. The bank fails whenever $\alpha > \alpha^{IN}(\ell) \equiv \frac{1}{\delta}\left(1 - \frac{FD}{RI}\right)$, which is independent of the fraction of withdrawals at the interim date. If the bank does not fail from either illiquidity or insolvency, it receives $EV(\alpha) = (1 - \alpha\delta)RI - FD$ as its equity value. Table 2 summarises the bank's balance sheet at $t = 2$ following a successful attack and rollover of bank deposits.

[11] In many standard bank run models (e.g., Rochet and Vives, 2004; Ahnert et al., 2019), the cost of runs at $t = 1$ stems from the fire-sale or costly liquidation of assets. We abstract from such fire sales and costly liquidation to highlight the core tension between ex ante contribution to cybersecurity – to improve the surveillance of IT systems for vulnerabilities – and ex post resilience in the event of a cyberattacks. Instead, the cost of the run at $t = 1$ is that if the bank is unable to service even a single withdrawal, this triggers a bankruptcy that completely erodes the value of its assets.

Assets	Liabilities
$(1-\alpha\delta)RI$	FD
	$EV(\alpha)$

TABLE 2. Balance sheet at $t = 2$ following a successful cyberattack by the attacker and rollover of bank debt at $t = 1$.

Rollover decisions. Investors delegate rollover decisions to professional fund managers who are rewarded for making the right decision—if the bank does not fail, a fund manager's payoff difference between withdrawing and rolling over is $-o < 0$; if the bank fails, the differential payoff is $w - o > 0$. The *conservatism ratio*, $\gamma \equiv \frac{w-o}{w}$, summarises these payoffs. More conservative managers (i.e., higher γ) are less inclined to rollover since the cost of withdrawing is low. When $\gamma > 0$, fund managers' actions are strategic complements and the bank is subject to rollover risk.

Investors are often are not perfectly informed about the details of a cyberattack because banks are reluctant to publicly broadcast details due to fear of further attacks and reputational concerns (Biener et al., 2015; Pretty, 2018). To this end, we suppose that the continuum of fund managers have incomplete information about the impairment shock on which they base their rollover decisions. Each fund manager, indexed $k \in [0, 1]$, receives a noisy signal

$$x_k = \alpha + \epsilon_k, \tag{4}$$

where ϵ_k is a zero-mean noise term that is independent of the shock, α, and is independently and identically distributed across fund managers according to a continuous distribution H with support $[-\epsilon, \epsilon]$, where $\epsilon > 0$.

Table 3 illustrates the timing of events in the model.

3. ANALYSIS

The unique symmetric, pure-strategy, perfect Bayesian equilibrium comprises attacker effort, A^*, critical thresholds for shock and signal, x^* and α^*, respectively, the bank's debt issuance, D^*, investment and cybersecurity, I^* and S^*, and the face value of debt, F^*, such that

$t = 0$	$t = 1$	$t = 2$
1. Bank chooses the amount of debt, D, to issue, investment, I, and cybersecurity, S,	1. Attacker succeeds with probability $p(A, S)$ and launches cyberattack	1. Bank's investments mature
2. Attacker chooses A to discovering and exploiting vulnerabilities	2. Fund managers receive private signals on shocks and roll over or withdraw	2. Attacker receives the prize if successful
		3. Bank and depositors consume

TABLE 3. Timeline of events.

(1) at $t = 1$, fund managers' rollover decisions, x^*, are optimal and the run threshold leads to bank failure whenever $\alpha > \alpha^*$, given A^*, D^*, I^*, S^* and F^*;

(2) at $t = 0$, the attacker's effort, A^*, maximises the expected prize from the contest, given the bank's contribution to cybersecurity, S^*;

(3) at $t = 0$, the bank's choices, D^*, I^*, and S^* maximise expected equity value given the critical thresholds, x^* and α^*, the attacker's effort, $A^* \equiv A^*(S)$, and the face value of debt, F^*;

(4) at $t = 0$, the face value of debt, F^*, makes investors indifferent between lending to the bank and using the storage technology, given D^*, I^*, and S^*, and the thresholds, x^* and α^*.

We construct the equilibrium backwards, solving first for the optimal rollover decision of fund managers before establishing the optimal effort by the attacker and bank, and finally the face value of debt.

Rollover risk. For a given mass of early withdrawals, ℓ, the bank does not fail provided the impairment shock is sufficiently small. The criteria determining the largest shock that the bank can withstand depends on whether failure is driven by illiquidity or insolvency.

Lemma 1. *There exists a critical level of withdrawals,*

$$\widehat{\gamma} \equiv \frac{1}{\delta} - \left(\frac{1}{\delta} - 1\right)\frac{RI}{FD}, \tag{5}$$

such that the bank fails due to illiquidity if and only if the mass of withdrawals is large, i.e., $\ell > \widehat{\gamma}$.

Figure 1 illustrates Lemma 1 by plotting the illiquidity and insolvency conditions together with their "envelope"—the red line encapsulating the region where the bank does not fail. In Region 1, the fraction of withdrawals is small, $\ell < \widehat{\gamma}$, so the bank is able to service them following the cyberattack as long as the impairment shock is not too large, $\alpha < \alpha^{IN}$. But, for a larger shock, as in Region 2, the losses borne by the bank are so high that it has too few resources to repay claims that are rolled over. So although the bank can meet interim liquidity needs, the cyberattack renders it insolvent at $t = 2$.

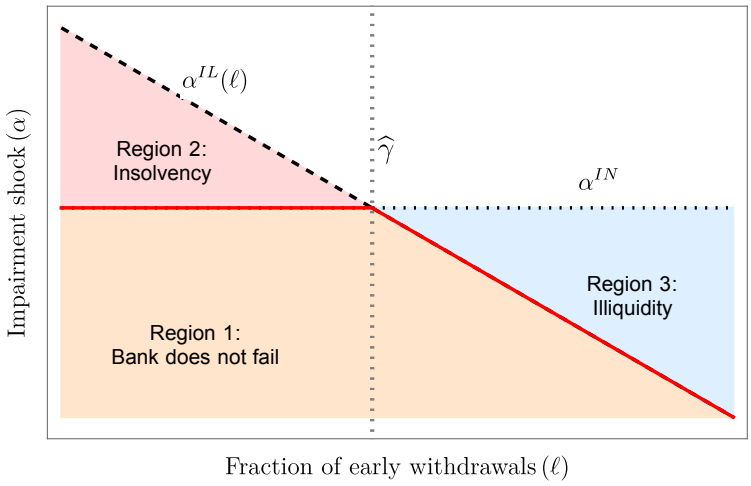


FIGURE 1. Failure conditions for a bank following a successful cyberattack.

Finally, in Region 3, the fraction of withdrawals, $\ell > \widehat{\gamma}$, given the impairment shock, $\alpha > \alpha^{IL}(\ell)$, is such that the bank is unable to service withdrawals at $t = 1$. This is despite the bank having sufficient resources at $t = 2$, allowing for cyberattack losses, to satisfy all its claims, i.e., $\alpha < \alpha^{IN}$. In this case, the bank fails due to illiquidity even though it is solvent.

Assumption 1. $(1 - \delta)RI < FD;\ RI > FD$.

Assumption 1 ensures a strict tripartite classification over values of the impairment shock (Morris and Shin, 1998). The first part of the assumption implies that, in the worst case where $\alpha = 1$, if all debt claims are rolled over, $\ell = 0$, the losses that the bank incurs are so large that the bank cannot service its debts at $t = 2$ and fails due to insolvency. A sufficiently large deadweight loss, δ, is required to ensure this region – where it is a dominant action for fund managers to withdraw – is well defined. In what follows, $\delta > 1 - \frac{r}{R}$ so that, irrespective of the bank's choices, this upper dominance region is always well defined.

The second part of the assumption ensures that if $\alpha = 0$ and $\ell = 1$, i.e., there is no impairment and all fund managers withdraw at $t = 1$, there are sufficient resources to repay all investors in full at $t = 1$. Since $I = D - S$, Assumption 1 places an upper bound on the bank's contribution to cybersecurity, i.e., $S < \overline{S} \equiv D\left(1 - \frac{F}{R}\right)$.

In the limit of vanishing private noise, $\epsilon \to 0$, the unique run threshold converges to the failure threshold, $x^* \to \alpha^*$, and fund managers face only strategic uncertainty about the behaviour of other managers.[12]

Proposition 1. *There exists a unique failure threshold,*

$$\alpha^* = \begin{cases} \alpha^{IN} \equiv \frac{1}{\delta}\left(1 - \frac{FD}{RI}\right) & \text{if } \gamma < \widehat{\gamma} \\ \alpha^{IL}(\gamma) \equiv 1 - \frac{\gamma FD}{RI} & \text{if } \gamma \geq \widehat{\gamma}, \end{cases} \tag{6}$$

such that the bank fails whenever $\alpha > \alpha^*$. Moreover, greater investment increases resilience, i.e, $\frac{\partial \alpha^*}{\partial I} > 0$.

Bank failure is driven by illiquidity when rollover risk is sufficiently large, i.e., $\gamma \geq \widehat{\gamma}$. Following an impairment shock, the bank is only able to service debt claims if few fund managers withdraw. But if sufficient withdrawals occur, it is in each fund manager's best interest to also withdraw. Under vanishing private noise, each fund manager observes the impairment shock almost perfectly

[12] Our results follow as long as ϵ is sufficiently small. We take the limit of vanishing private noise, which is conventional in the global games literature for tractability (Morris and Shin, 1998).

and knows all other fund managers do too. There is a unique impairment shock, $\alpha^{IL}(\gamma)$, such that the bank fails due to illiquidity whenever $\alpha > \alpha^{IL}(\gamma)$.

By contrast, when rollover risk is low, $\gamma < \widehat{\gamma}$, bank failure is driven by insolvency. Concerns over the losses borne by the bank due to the deadweight loss, δ, incentivises fund managers to withdraw – each fund manager has a strictly dominant strategy to withdraw at $t = 1$, since the bank is sure to fail at $t = 2$. We consider α^* to be a measure of *resilience* in the event of a cyberattack.

The result in Proposition 1 is conditional on a cyberattack taking place and does not account for the trade-off faced by the bank between resilience in the event of a cyberattack, and protection against cyberattacks altogether. We next examine this trade-off in the context of the bank's ex ante investment decision.

Attacker effort and optimal cybersecurity. At $t = 0$, the attacker chooses how much effort to expend in uncovering vulnerabilities to maximise its expected prize, taking as given the bank's level of cybersecurity. At the time of choosing A, the attacker does not know what vulnerabilities it will find and how disruptive their exploitation might be for the bank. Vulnerabilities are thus "known unknowns" (Rumsfeld, 2002). The attacker's problem can be expressed as

$$A^* \equiv \arg\max_{A} \Big(1 - p(A, S)\Big)V - cA. \tag{7}$$

Lemma 2. *The attacker expends effort*

$$A^*(S) = \begin{cases} \sqrt{\frac{SV}{c}} - S & \text{if } S < \frac{V}{c} \\ 0 & \text{otherwise}, \end{cases} \tag{8}$$

at $t = 0$ to discover and exploit vulnerabilities. Effort is increasing in the prize but decreasing in the marginal cost of effort.

Lemma 2 is intuitive. The attacker expends strictly positive effort as long as the prize is sufficiently valuable. Equilibrium effort is larger, the greater the prize. And, as the marginal cost, c, increases, the potential gain to effort is lower and there is less effort on the part of the attacker.

Banks typically commit to their IT budgets at the start of the financial year. Consistent with this, we follow Dixit (1987) and suppose that the bank pre-commits to its debt issuance, investment and security choices and, as such, has a first-mover advantage over the attacker. If $A^* > 0$, the probability that the bank will identify vulnerabilities and put mitigating measures in place is

$$p(S) \equiv p\left(A^*(S), S\right) = \sqrt{\frac{cS}{V}}, \tag{9}$$

which satisfies the Inada conditions.

The bank's expected profit is the sum of two terms: its profit in the event that it successfully mitigates vulnerabilities and its residual profits if the attacker is successful, i.e.,

$$\pi(D, I, S) = p(S)(RI - FD) + \Big(1 - p(S)\Big) \int_0^{\alpha^*(D,I)} EV(\alpha)\, d\alpha\,, \tag{10}$$

The optimisation problem for the bank is thus

$$\begin{aligned} \max_{\{D,I,S\}} \quad & \pi(D, I, S) \\ \text{subject to} \quad & D = I + S \\ & D \leq 1 \end{aligned}$$

The constraints reflect: (i) the bank's initial balance sheet, and (ii) the resources available for the amount of debt that can be issued.

Proposition 2. *For any given I and S, the bank issues $D^* = 1$ of uninsured demandable debt.*

The net marginal benefit to the bank from issuing debt is strictly positive, even after accounting for the potential losses in the event of a cyberattack. This marginal benefit is the bank's *intermediation margin*, i.e., the difference between what it earns from investing the debt it has raised and the additional repayment to investors. If $R - F > 0$, the intermediation margin is always strictly positive and the bank is incentivised to issue as much as debt as possible. In what follows, we exploit that $D^* = 1$ to simplify the remainder of our analysis. An immediate consequence is that, due to limited liability, the participation constraint, $\pi(D^*, I, S) \geq 0$, is satisfied for any I and S.

To ensure that the profit function is well defined and concave, we assume that the odds ratio for avoiding a cyberattack is sufficiently large.

Assumption 2. $\frac{p(S)}{1-p(S)} > \frac{2S}{1-S}$.

The odds ratio is a sufficient statistic that captures the relative likelihood that the bank finds and resolves vulnerabilities before the attacker is able to exploit them. Assumption 2 requires that these odds be high relative to the ratio of the bank's ex ante contribution to cybersecurity and its investment.

Proposition 3. *There exists a lower bound on returns, $\underline{R}$, and an upper bound on the deadweight loss, $\overline{\delta}$. For $R > \underline{R}$, there is a unique S^* that maximises the bank's expected equity value. The optimal S^* is given by the solution to*

$$\frac{R(1-S) - F - \int_0^{\alpha^*(S)} EV(\alpha)d\alpha}{p(S)R + \Big(1-p(S)\Big)\left(\int_0^{\alpha_b^*(S)} R(1-\delta\alpha)d\alpha - EV(\alpha^*)\frac{\partial\alpha^*}{\partial S}\mathbb{1}_{\gamma>\widehat{\gamma}}\right)} = \frac{1}{\partial p/\partial S}, \tag{11}$$

such that the threshold $\widehat{\gamma} \equiv \widehat{\gamma}(S^*)$ is unique and well defined. For $\delta < \overline{\delta}$, the bank's contribution to cybersecurity is increasing in the face value of debt i.e., $\frac{\partial S^*}{\partial F} > 0$.

The bank's choice of cybersecurity trades-off protection from cyberattacks against resilience in the event of a cyberattack. By contributing more to cybersecurity, the bank improves its chances of uncovering vulnerabilities, thereby lowering the probability of a successful attack. But this comes at the cost of reduced investment in the safe asset, which has two implications. First, irrespective of the outcome of the contest, the value of the bank's investments are lower, thereby reducing its equity value. And second, in the event of a successful attack, the bank has fewer assets to service withdrawals for any given disruption. The bank is, therefore, less able to weather a run in the event of an attack.

Equation (11) formalises this trade-off. The left-hand side is the ratio of the marginal impact on expected profits from an increase in protection (numerator), to the marginal impact on expected profits from an increase in investments (denominator). The right-hand side is the marginal rate

of transformation between the bank's contribution to cybersecurity and protection. So the bank chooses an allocation that equates the quantity of investment it is willing to forgo for a unit of extra protection with the quantity of investment needed to produce that extra unit.

The protection-resilience trade-off depends on the nature of bank failure. For $\gamma \leq \widehat{\gamma}$, the disruption caused by the cyberattack does not precipitate funding liquidity risk. Failure is, thus, driven by solvency concerns, i.e., $\alpha^* = \alpha^{IN}$. At the point of failure, bank equity value is wiped out, i.e., $EV(\alpha^*) = 0$. When rollover risk is large, $\gamma > \widehat{\gamma}$, the disruption caused by the cyberattack leads to a run on the bank. The bank fails due to illiquidity and $\alpha^* = \alpha^{IL}(\gamma)$. Since the run is inefficient, the bank fails even though it has positive equity value, $EV(\alpha^*) > 0$. In this case, the marginal impact to the bank from additional investment in the safe asset induce an additional effect. There is a benefit to preserving equity value in the event of an attack, following a marginal shift in the failure threshold.

An increase in the face value of debt generates two countervailing effects on the bank's contribution to cybersecurity. First, the relative gains from finding and mitigating vulnerabilities are lower. This is because the equity value that accrues to the bank in either state is lower and because resilience is reduced, i.e., $\frac{\partial \alpha^*}{\partial F} < 0$. Second, the bank has increased incentives to contribute to cybersecurity. Insofar that bank resilience is lower, the bank faces a lower cost of contributing to cybersecurity as it is more likely to fail when the attacker is successful. If the asset return, R, is large enough, the second effect dominates and a higher face value of debt increases the bank's contribution to cybersecurity.

Pricing of unsecured debt. Investors are repaid in full whenever the bank does not fail. This occurs if the bank is either successful in mitigating vulnerabilities, or if it remains resilient through a successful cyberattack. Since investors receive nothing in the event of bank failure, the value of a debt claim is

$$\mathcal{V}(F) \equiv \Big(p(S) + \Big(1 - p(S)\Big)\alpha^*(F)\Big)F\,. \tag{12}$$

Under perfect competition, the equilibrium face value of debt, F^*, ensures investors are indifferent between lending to the bank and investing in the storage technology, i.e., $\mathcal{V}(F^*) = r$, given the bank's contribution to cybersecurity.

Lemma 3. *The face value of debt is increasing in the bank's contribution to cybersecurity,* $\frac{\partial F^*}{\partial S} > 0$ *whenever bank failure is driven by insolvency risk,* $\gamma \leq \widehat{\gamma}$*. But for* $\gamma > \widehat{\gamma}$*,* $\frac{\partial F^*}{\partial S} < 0$*.*

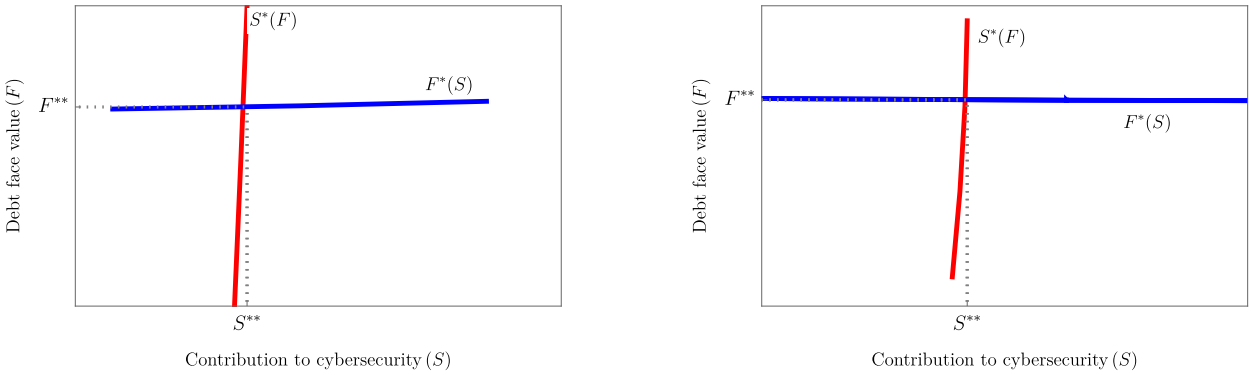
The ex ante behaviour of investors depends on whether bank failure is driven by illiquidity or insolvency. If $\gamma > \widehat{\gamma}$ and illiquidity is the root cause of financial distress, it follows from Assumption 2 that the value of a debt claim increases as the bank contributes more to cybersecurity. In these circumstances, investors prefer that the bank takes measures to forestall cyberattacks and, hence, runs. So the face value of debt is decreasing in S. But, when $\gamma \leq \widehat{\gamma}$, and failure is insolvency-driven, investors look upon bank contributions to cybersecurity less favourably. As S increases, the bank has fewer resources to repay investors at $t = 2$ following an attack. The increased likelihood of the bank being unable to repay investors prompts them to seek greater compensation, and so $\frac{\partial F^*}{\partial S} > 0$.

Private equilibrium. The joint equilibrium for the bank's contribution to cybersecurity and the face value of debt is summarised in Proposition 4 and depicted in Figures 2a and 2b.

Proposition 4. *There exist a unique equilibrium, which is jointly characterised by* $\mathcal{V}(F^{**}, S^{**}) = r$*, where the bank's contribution to cybersecurity is an interior,* $S^{**} \equiv S^*(F^{**})$*.*

Figure 2a depicts the private equilibrium under insolvency risk. In this case, both best-response correspondences are upward sloping. When investors expect the bank to contribute to cybersecurity, thereby reducing its resilience ex post in the event of a cyberattack, they require a higher face value of debt in return for lending. The bank, in turn, expects a higher face value of debt and is incentivised to contribute more to cybersecurity since the marginal costs of doing so are lower. Through a simple iterative process, the unique equilibrium obtains.

The equilibrium when bank failure is illiquidity-driven is shown in Figure 2b where the best-response correspondence for the face value of debt is downward sloping. When investors expect



(A) $\gamma = 0.1$ and $\widehat{\gamma}(S^{**}, F^{**}) = 0.55$.

(B) $\gamma = 0.8$ and $\widehat{\gamma}(S^{**}, F^{**}) = 0.68$.

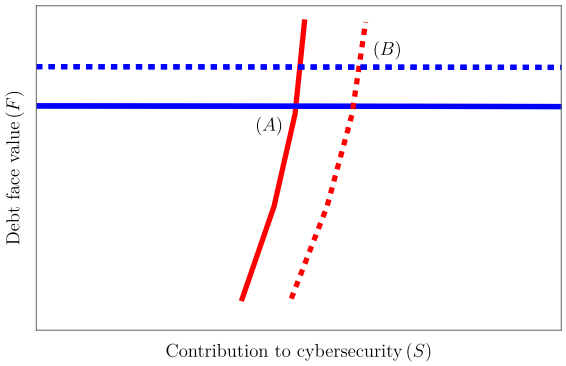
FIGURE 2. In producing the figure, we assume $R = 3.5$, $r = 1$, $\delta = 0.8$, $c = 0.1$ and $V = 1$.

a higher contribution to cybersecurity by the bank, they require a lower face value of debt to lend. Expecting a low face value of debt, the bank is dis-incentivised from contributing to cybersecurity. But anticipating this, in turn, investors seek a high face value of debt, and so on. Following this iterative process, a unique fixed point and equilibrium is obtained.

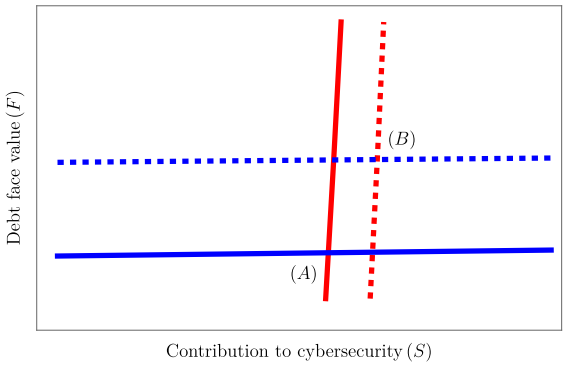
Cyber threat landscape. In the context of our model, the cyber threat landscape can be summarised by the sophistication of the attacker, c, and the risks posed by an attack as reflected in the deadweight loss, δ, and the nature of bank fragility, γ.

Proposition 5. *The bank's contribution to cybersecurity, S^{**}, is increasing in rollover risk, γ, and the deadweight loss, δ. There exists a critical level for the attacker's sophistication, $\tilde{c}^{**}$, such that the bank's contribution to cybersecurity is decreasing in the sophistication of the attacker if and only if $c > \tilde{c}^{**}$.*

To the extent that $\gamma > \widehat{\gamma}$, an increase in rollover risk is more likely to result in bank failure due to a run in the event of a cyberattack. So the marginal benefit of cybersecurity contribution is high relative to the benefits of bolstering resilience. The bank is better off improving cybersecurity to ward off an attack entirely, even though it results in lower resilience in the event of a successful attack. This results in an upward shift to the bank's cybersecurity contribution schedule. At the same time, the increase in γ heightens investors' concerns about repayment and they seek a higher



(A) Rollover risk increases from $\gamma = 0.8$ to $\gamma = 0.82$.



(B) Deadweight loss increases from $\delta = 0.8$ to $\delta = 0.82$. Rollover risk is $\gamma = 0.1$.

FIGURE 3. In producing the figure, we assume for the other parameters that $R = 3.5$, $r = 1$, and $V = 1$.

face value of debt as compensation. The face value of debt schedule also moves outwards. This further amplifies the bank's incentives to contribute to cybersecurity. Figure 3a illustrates how an increase in γ shifts the equilibrium from (A) to (B).

An increase in the deadweight loss, δ, has two effects. First, the difference in expected equity value earned by the bank between an unsuccessful and successful cyberattack is larger. This increases the marginal rate of substitution for the bank. Second, the benefits from greater resilience are lower for any level of impairment, α. The bank is better off trying to avoid becoming exposed to impairment shocks altogether. The two effects reinforce the benefits to protection via greater contribution to cybersecurity. When $\gamma < \widehat{\gamma}$, investors' face value of debt also responds to a change in the deadwight loss. As δ increases, investors require a higher face value of debt as compensation to break even. The $S^*(F)$ schedule also moves outwards, which further increases the bank's incentives to contribute more to cybersecurity. Figure 3b illustrates how an increase in δ shifts the equilibrium from (A) to (B).

An increase in c implies that it is more costly for the attacker to invest in finding vulnerabilities to exploit. Lemma 2 shows that this causes the attacker to invest less, which improves banks' chances of uncovering vulnerabilities first and taking sufficient mitigating measures. This incentivises banks to contribute more to cybersecurity which also increases fragility in the event a successful cyberattack is launched by the attacker. And so the $S^*(F)$ schedule moves outwards. At

the same time, however, the increase in the likelihood of the bank succeeding in protecting itself leads investors to judge that lending to the bank is marginally safer. So they require a lower face value of debt for lending to the bank. The $F^*(S)$ schedule shifts downwards, which counters the direct effect on $S^*(F)$. In general, the direct effect outweighs the indirect effect, $\frac{\partial S^*}{\partial c} + \frac{\partial S^*}{\partial F}\frac{\partial F^*}{\partial c} > 0$, when the attacker is highly sophisticated, i.e., $c < \tilde{c}^{**}$. In this case, a marginal increase in c elicits only a small effect on the face value of debt and $\frac{dS^{**}}{dc} > 0$. But if $c > \tilde{c}^{**}$, then the indirect effect dominates and $\frac{dS^{**}}{dc} < 0$.

4. Normative implications

Social cost of bank failure. Bank failures are socially costly. Examples include the loss of payment services when the bank enters financial distress and the macroeconomic costs of a sharp credit contraction. We treat these social costs as exogenous and capture them by the parameter $\lambda > 0$. Social costs are incurred if the bank fails following the cyberattack with probability $(1-p)(1-\alpha^*)$. Although the bank does not account for the social cost of failure when making its choices, these are taken into account by a social planner.

The planner takes as given the incomplete information structure of the game and chooses a cybersecurity contribution schedule for a given face value of debt, namely

$$S^P(F) = \arg\max_S \quad \pi(S) - \lambda\Big(1 - p(S)\Big)\Big(1 - \alpha^*(S)\Big).$$

Although the cybersecurity contribution schedules of the planner and bank differ, the participation constraint of investors remains the same.

Proposition 6. *There exists a critical threshold for the attacker's sophistication, $\widehat{c}$, such that the bank over-invests in cybersecurity, $S^{**} > S^P$, if and only if $\gamma \leq \widehat{\gamma}$, i.e., its failure is driven by insolvency concerns, and $c < \widehat{c}$. Else, the bank under-invests in cybersecurity, i.e., $S^{**} < S^P$.*

The bank over-invests in cybersecurity when its failure is driven by solvency concerns and it faces a sophisticated attacker, but underinvests otherwise. To see this, consider the difference in

marginal benefit from greater cybersecurity for the planner and bank. Upto the multiplicative factor λ, this is given by

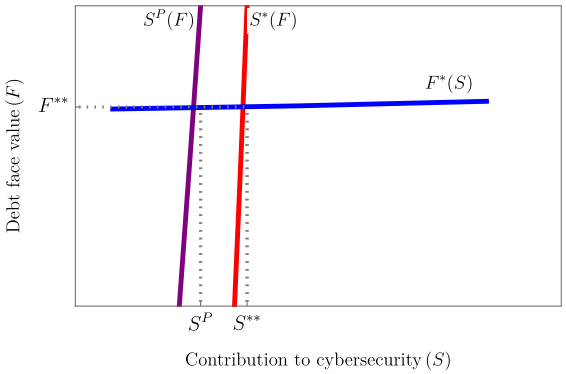
$$\frac{\partial p}{\partial S}\Big(1-\alpha^*(S)\Big)+\Big(1-p(S)\Big)\frac{\partial \alpha^*}{\partial S}, \tag{13}$$

which has two opposing effects. On one hand, by contributing more to cybersecurity, the ex ante likelihood of bank failure is reduced. This *social protection effect* is given by the first term in Equation (13). On the other hand, conditional on suffering a cyberattack, the bank is more likely to fail since it has fewer resources to meet its obligations. This *social resilience effect* is given by the second term in Equation (13).

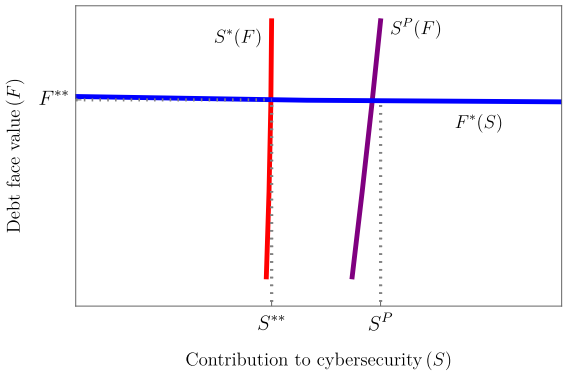
When insolvency concerns dominate and $\gamma < \widehat{\gamma}$, the failure threshold is relatively large. As a result, the ex ante likelihood of the bank failing is low and scales down the social protection effect. At the same time, where the attacker is highly sophisticated, $c < \widehat{c}$, the ex ante likelihood of the attacker winning the contest is low, $p(S)$ is small, which scales up the social resilience effect. Accordingly, $S^P < S^{**}$, and the planner allocates less to cybersecurity than the bank. But as the attacker's capabilities diminish and $c > \widehat{c}$, the chances of the bank winning the contest are large, which diminishes the social protection effect leading to the bank under-investing in cybersecurity.

When bank failure is driven by illiquidity, $\gamma > \widehat{\gamma}$, two forces come into play. First, the critical threshold is lower than when failure is insolvency-driven. This implies that the conditional likelihood of bank failure following an attack is higher when rollover risk is high. The social benefit from greater protection is, therefore, greater. Second, following a marginal increase in S, the critical threshold responds by less, i.e., $\frac{\partial \alpha^*}{\partial S}\Big|_{\gamma<\widehat{\gamma}} < \frac{\partial \alpha^*}{\partial S}\Big|_{\gamma>\widehat{\gamma}} < 0$. In other words, the social opportunity cost from allocating more to cybersecurity and reducing resilience is smaller. So the social resilience effect is tempered. The combination of the two effects means that the social protection effect dominates, irrespective of the attacker's sophistication, and the planner allocates more to cybersecurity than the bank, $S^P > S^{**}$. Figure 4 illustrates both cases.

Proposition 7. *The difference, $|S^P - S^{**}|$ is increasing in c, i.e., the attacker becomes less sophisticated, if either: (i) $\gamma > \widehat{\gamma}(S^{**}, F^{**})$, or (ii) $\gamma < \widehat{\gamma}(S^{**}, F^{**})$ and $c > \widehat{c}$.*

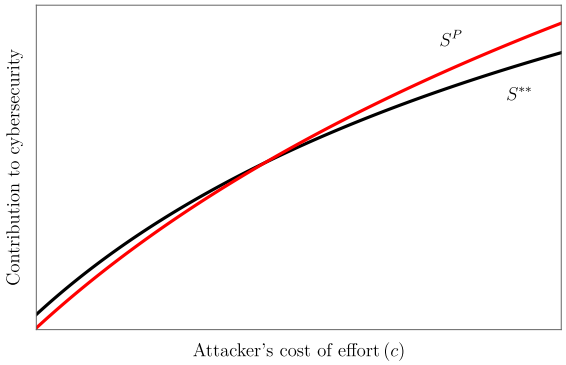


(A) $\gamma = 0.1$ and $\widehat{\gamma}(S^{**}, F^{**}) = 0.55$.

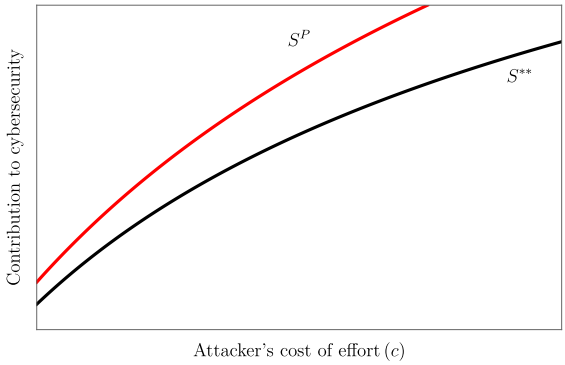


(B) $\gamma = 0.8$ and $\widehat{\gamma}(S^{**}, F^{**}) = 0.68$.

FIGURE 4. Joint equilibrium for the bank's contribution to cybersecurity as well as the social planner's allocation and the face value of debt. In producing the figure, we assume $R = 3.5$, $r = 1$, $\delta = 0.8$, $c = 0.1$, $V = 1$ and $\lambda = 1$.



(A) $\gamma = 0.1$.



(B) $\gamma = 0.8$.

FIGURE 5. Equilibrium investment in cybersecurity as a function of the attacker's cost of effort. In producing the figure, we assume $R = 3.5$, $r = 1$, $\delta = 0.8$, $\gamma = 0.8$, $\lambda = 1$, and vary c from 0.5 to 2.5.

Proposition 6 established that when bank failure is illiquidity-driven, or if it is insolvency-driven and attacker sophistication is low, the bank contributes too little to cybersecurity relative to the planner. Proposition 7 shows, moreover, that as the attacker becomes less sophisticated, the extent of this underinvestment increases. As c increases and the bank's chances of mitigating vulnerabilities increases, the planner places greater emphasis on social resilience and allocates more to cybersecurity. Figure 5 illustrates.

4.1. **Operational resilience standards.** Our analysis highlights that the correct regulatory prescription crucially depends on the nature of the attacker and whether or not cyberattacks precipitate

insolvency or illiquidity, i.e., the nature of the bank's fragility. When attackers are sophisticated and insolvency risk is central, policies that limit the bank's contribution to cybersecurity and prioritise shoring up resilience may be more effective.

Proposition 8. *When bank failure stems from insolvency, $\gamma < \widehat{\gamma}$, the constrained efficient solution is obtained by introducing the constraint, $S \leq S^P(F)$, into the bank's optimisation problem.*

Absent the constraint, which imposes a *minimum operational resilience standard* on the bank, Proposition 6 suggests that the bank over-invests in protection relative to resilience in the event of an attack. This is because the social resilience effect outweighs the social protection effect in Equation (13) and this social benefit is absent from the bank's private incentives to shore up resources to meet its obligations. The introduction of the constraint eliminates the wedge between the bank and the planner's solution.

Such constraints can be implemented in practice. Cyber stress-tests, cyber hygiene notices and business continuity planning are examples.[13] By assessing the resilience of banks to a range of cyberattacks, stress-tests subject banks to scenarios that assume disruption and can help assess deficiencies in banks' resilience. Supervisory authorities increasingly use cyber hygiene notices to set minimum IT standards on banks, while industry initiatives such as Sheltered Harbor offer certification on resilience planning.[14] Detailed business continuity planning requiring emergency payment systems, data back-ups, and account reconciliations also serve to shape minimum operational resilience standards (Duffie and Younger, 2019).[15]

4.2. **Red-team testing.** When bank failure from a cyberattack is driven by liquidity risk, our analysis points to a greater contribution to cybersecurity to mitigate vulnerabilities. The planner's constrained efficient solution can be achieved by improving the efficacy of the bank's contribution

[13] See, for example, Goh et al. (2020); Kashyap and Wetherilt (2019); Adelmann et al. (2020); Crisanto et al. (2023) and Englund and Sosa (2022).

[14] The Federal Reserve's standards are jointly set with the OCC and FDIC for global systemically important banks, and the Uniform Rating System for Information Technology(URSIT) is applied to regional banks (Federal Reserve Board of Governors, 2023).

[15] See, for example, guidance published by the Federal Financial Institutions Examination Council (FFIEC) which assists examiners in evaluating financial institution business continuity processes (Federal Financial Institutions Examination Council, 2019).

to cybersecurity. By contributing S, the level of cybersecurity can be bolstered to $S(1+\rho)$, where $\rho > 0$ is the boost to cybersecurity provided by the planner. So, for any given level of attack effort, A, , the level of protection obtained by the bank is

$$p(A, S; \rho) \equiv \frac{S(1+\rho)}{A + S(1+\rho)}, \tag{14}$$

which is increasing in ρ.

Proposition 9. *When failure is driven by illiquidity, $\gamma \geq \widehat{\gamma}$, the constrained efficient equilibrium is obtained by bolstering the bank's contribution to cybersecurity by ρ^*, such that $p(S^{**}(\rho^*)) = p(S^P)$.*

The boost by the planner induces an increase in the ratio between the marginal benefit from greater cybersecurity to the marginal benefit from greater investment. In other words, the bank is more efficient at allocating resources towards cybersecurity in a way that improves its changes of winning the contest against the attacker.

The planner's boost to the bank's security can be viewed as a form of assisted attack simulation exercises, penetration testing, and red-teaming. These involve third-party experts performing sanctioned attacks on the bank to uncover vulnerabilities in systems, and promoting information sharing across services and institutions that share common hardware and software solutions (Association of Banks in Singapore, 2018).

Red-team testing alters the nature of the contest between the bank and the attacker. By subjecting the bank to simulated attacks, they are better equipped to guard against intrusions in the first instance.[16] These tests may be funded by the banks themselves, by authorities, or even crowd-sourced through initiatives that promote information sharing and vulnerability reporting.[17] Table 4 summarises red team practices in various jurisdictions.

[16]To account for the varying nature of attacks, a combination of so-called blackbox, in which no knowledge of the environment is assumed, and greybox, in which the red team is authenticated using permissions granted to a normal customer, tests are carried out on bank systems (Monetary Authority of Singapore, 2021).

[17]The Cybersecurity & Infrastructure Security Agency in the US promotes free, open-source software, such as command-line vulnerability scanners, that promote best practice in cybersecurity.

Jurisdiction	Framework	Year Launched	Institutions Covered
Germany	TIBER-DE	2019	Participation is voluntary
European Union	Threat Intelligence Based Ethical Red Teaming (TIBER-EU)	2018	At the discretion of relevant national authorities
Hong Kong SAR	Intelligence led Cyber Attack Simulation Testing (iCAST)	2016	All banks
Iceland	TIBER-IS	2023	Systematically important banks
Netherlands	TIBER-NL	2016	Institutions that are part of the core financial infrastructure
Saudi Arabia	Financial Entities Ethical Red Teaming (FEET)	2019	Domestic systemically important institutions
Singapore	Adversarial Attack Simulation Exercises(AASE)	2018	Large financial institutions
United Kingdom	CBEST	2014	Critical financial institutions

TABLE 4. Summary of red-teaming frameworks across jurisdictions. Adapted and extended from Kleijmeer et al. (2019).

4.3. **Subsidising cyber capability.** Whereas red-team testing entails the involvement of third-parties to help alter the contest between bank and attacker, a direct subsidy can also help free up resources to enable the bank to meet cybersecurity requirements. The planner's solution can be obtained using revenue-neutral, targeted Pigovian subsidies that are funded through lump-sum taxation.

Let σ as the dollar value of the subsidy, and τ the lump sum tax used to finance the subsidy. Proposition 10 demonstrates that Pigouvian subsidies elicit higher investment in cybersecurity to bridge the gap between the level of protection offered by the bank and planner.

Proposition 10. *When failure is driven by illiquidity, $\gamma \geq \widehat{\gamma}$, the planner can achieve the constrained efficient outcome using a Pigouvian subsidy, $\sigma(S)$, at $t = 2$. The subsidy is funded by taxes, $\tau = \sigma(S^P)$, imposed as a lump-sum on the bank.*

The subsidy rewards the bank for contributing additional units towards cybersecurity while offsetting the private cost that would be incurred from heightened fragility in absence of the scheme. The amount granted on each unit allocated to cybersecurity is set to the marginal social value of

protection. This induces the bank to forego investment and set cybersecurity to the level S^P. The lump sum tax is then set to the level of the subsidy that is paid to the bank. The optimal subsidy is proportional to the degree of rollover risk, $\frac{\partial \sigma}{\partial \gamma} > 0$. This reflects the wedge that funding liquidity drives between the bank's incentives to shore up its defences and the planner's objective.

Regulators, such as the Dutch central bank, offer direct, targeted cybersecurity subsidies (*subsidie cyberweerbarheid*). Such policies are aligned with calls from the IMF and BIS to provide banks with financial support to bolster cybersecurity practices (Mauer and Nelson, 2020; Doerr et al., 2022).

4.4. **Duty of care.** The social planer can also penalise the bank in the event of a successful breach to ensure the ex ante constrained efficient outcome. Such an approach requires a diagnostic assessment of the bank's compliance with regulatory requirements if a successful attack takes place. A negligence rule establishes a minimum level of *due care* for the bank (Brown, 1973; Shavell, 2009). In the event of a cyberattack, there is no further liability for the bank as long as it exercises the due care standard, (i.e. contributes S^P). Otherwise, a penalty, $\kappa(S)$, is imposed proportional to the level of under-investment.

Proposition 11. *When bank failure is driven by illiquidity, $\gamma \geq \widehat{\gamma}$, the planner can achieve the constrained efficient outcome by introducing a negligence rule at $t = 2$ with a penalty, $\kappa(S)$, in the event of a successful attack, that is proportional to the bank's investment, S.*

The negligence rule is an ex post intervention that disciplines bank behaviour, in contrast to the ex ante intervention imposed by cyber stress testing and red-team testing. The penalty, κ, represents the distance between the planner's optimum and what the bank chooses in absence of intervention. It helps ensure that the constrained efficient allocation provides the best private outcome for the bank. And since incentives to invest are increasing in rollover risk, $\frac{\partial \kappa}{\partial \gamma} < 0$. The penalty required to enforce the negligence rule is smaller when rollover risk prompts the bank to contribute more in cybersecurity. The mere threat of this penalty induces the bank to substitute

towards cyber protection until the constrained efficient allocation is obtained, meeting the standard of due care.

A number of regulators implement such negligence rules in practice. For example, in 2021, the SEC fined First American almost $500,000 for poor cybersecurity practices that resulted in the disclosure of customer information. And in June 2023, the Australian Prudential Regulation Authority (APRA) penalised a health insurance provider with a requirement to hold an extra $250 million in capital following a cyber breach in 2022.

The suite of regulatory tools described above are complementary. Table 5 summarises the use case for each tool as well as considerations for their interactions with other initiatives.

Regulatory tool	**Use case**	**Timing of implementation**	**Additional considerations**
Operational resilience standards	Insolvency risk	Ex ante	Identify extent to which continuity plans are jointly realistic at system level. Care in mandating specific recovery time (Crisanto et al., 2023). Adequacy of capital requirements for cyber vs. financial risks.
Red-team testing	Illiquidity risk	Ex ante	Requires test infrastructure and expertise to enable "no holds barred" exercises. Initial costs likely substantial but builds "cybersecurity capital" over time.
Cyber capability subsidy	Illiquidity risk	Ex ante	Assists banks with getting capability off the ground. Model suggests risk of misallocation of resources if not accompanied by appropriate guidelines and standards.
Duty of care	Illiquidity risk	Ex post	Threat of penalty encourages adoption of preventative measures such as red-team testing. Threat of penalties from deficient practices may disincentivise incident reporting and information sharing.

TABLE 5. Proposed regulatory toolkit for addressing cyber threat landscape.

5. Discussion

In this section we consider four extensions to our baseline model. First, we investigate how ex ante Knightian uncertainty may shape the behaviour of depositors. Second, we extend the model to consider asset illiquidity and the role of a lender of last resort. Third, we consider multiple banks using common IT systems. And finally, we consider the role of technology vendors.

5.1. **Knightian uncertainty.** For a given face value, F, the value of a debt claim, $\mathcal{V}(F)$, depends on investors' belief that the bank will be hit by a cyberattack as well as the loss distribution conditional on a cyberattack. Assessing the likelihood that the bank will be subject to a cyberattack is, however, often beset by Knightian uncertainty. This is especially true for zero-day vulnerabilities, which are security flaws that were previously unknown to exist (Perlroth, 2021). So if even security experts are unaware of what security flaws might be exploited, assessing the risk of a cyberattack for investors – who have limited technical expertise – is much more challenging. But, conditional on a cyberattack taking place, the loss distribution, and thereby the disruptions faced by the bank, are measurable and constitute a quantifiable risk.

Formally, suppose that investors have an initial reference belief B for the likelihood of the bank being hit by a cyberattack. Drawing on Hansen and Sargent (2001), the value of the debt claim for the representative investors is given by

$$\mathcal{V}(F) = \min_{K\in[0,1]} \left(KF + (1-K)\left(\int_0^{\alpha^*(F)} d\alpha \right) F + \theta\, KL(K||B) \right), \tag{15}$$

where $KL(K||B)$ is the Kullback-Leibler divergence of the probability K from the reference probability B. The parameter $\theta \geq 0$ captures the degree of trust that investors have in their reference probability, B, i.e., the concern for model misspecification (Strzalecki, 2011). In the limit $\theta \to \infty$, the investor fully trusts the reference model, B, and the debt claim is valued using a standard expected utility criterion. While, in the limit $\theta \to 0$, the investor has no confidence in their reference model and the debt claim is valued as per the maxmin criterion, i.e., by considering the worst state of the world (Gilboa and Schmeidler, 1989).

Solving for the probability, K^*, from the first-order condition, $\frac{\partial \mathcal{V}}{\partial K}\Big|_{K=K^*} = 0$, and evaluating Equation (15) at K^*, we obtain

$$\mathcal{V}(F) = \alpha^*(F)F - \theta \log\left[1 - B + B\,\mathrm{e}^{-\frac{1}{\theta}\left(1-\alpha^*(F)\right)F}\right]. \tag{16}$$

The result in Equation (16) can be viewed as interpolating between the extreme obtained when we consider the limits for both θ and B. For $B \to 0$, where investors anticipate that the cyberattack is inevitable, the debt claim is valued according to the maxmin criterion and so $\mathcal{V}(F) = \alpha^*(F)F$. While for $B \to 1$, where investors anticipate there to be no cyberattack, we obtain $\mathcal{V}(F) = F$. Similarly, for $\theta \to 0$, the debt claim is valued using the maxmin criterion. While for $\theta \to \infty$, the standard expected utility criterion, with belief B, is employed by investors.

Under perfect competition amongst investors, the equilibrium face value of debt, F^*, solves $\mathcal{V}(F^*) = r$, where the bank's contribution to cybersecurity is taken as given.

Proposition 12. *The face value of debt, F^*, is increasing in the bank's contribution to cybersecurity, $\frac{\partial F^*}{\partial S} > 0$, but is decreasing the investors' reference model, $\frac{\partial F^*}{\partial B} < 0$, and the degree of trust in the reference model, $\frac{\partial F^*}{\partial \theta} < 0$.*

The results in Proposition 12 are largely intuitive. First, as the bank contributes more to cybersecurity, this reduces its resilience in the event of a cyberattack. So to compensate for the greater risk of not being repaid in the event of a cyberattack, investors require a higher face value of debt as compensation for lending to the bank, implying that $\frac{\partial F^*}{\partial S} > 0$. As θ increases, investors place greater emphasis on their reference model, B. If $B > 0$, greater emphasis is placed on states of the world where the bank is successful in finding and patching the IT vulnerabilities. This, in turn, increases the expected valuation of the debt claim for any given F. So to break even, a lower face value of debt is required. By similar logic, as B increases, the investors' reference model becomes more optimistic with regards to the bank addressing its vulnerabilities. This increases the expected valuation of debt claims and, thus, a lower equilibrium face value of debt to break-even.

To the extent that $F^*(S)$ is increasing in S, irrespective of γ, and this does not impact the bank's choice of cybersecurity, all our normative results are qualitatively robust to the introduction of Knightian uncertainty.

5.2. **Asset illiquidity and the lender of last resort.** Our analysis has assumed that the bank's project is perfectly liquid. But very often investments are costly to liquidate. A simple way of capturing asset illiquidity is to assume that the liquidation value at $t = 1$ for a unit of the bank's investment is ψRI, where $\psi < 1$ is a fire-sale discount.

Our analysis of the run dynamics is robust to the introduction of asset illiquidity. The $t = 1$ failure condition is

$$(1-\alpha)RI - \frac{\ell F}{\psi} < 0,$$

and the $t = 2$ failure condition is

$$(1-\delta\alpha)RI - F\left[1+\ell\left(\frac{1}{\psi}-1\right)\right] < 0,$$

which now also depend on the fraction of withdrawals, ℓ, at $t = 1$. It follows that for $\ell < \widehat{\gamma}_\psi \equiv \frac{(1-\delta)\psi RI - F}{1-\psi-\delta}$, the insolvency condition drives bank failure, while for $\ell > \widehat{\gamma}_\psi$, the illiquidity condition is the relevant one. Figure 6 illustrates.

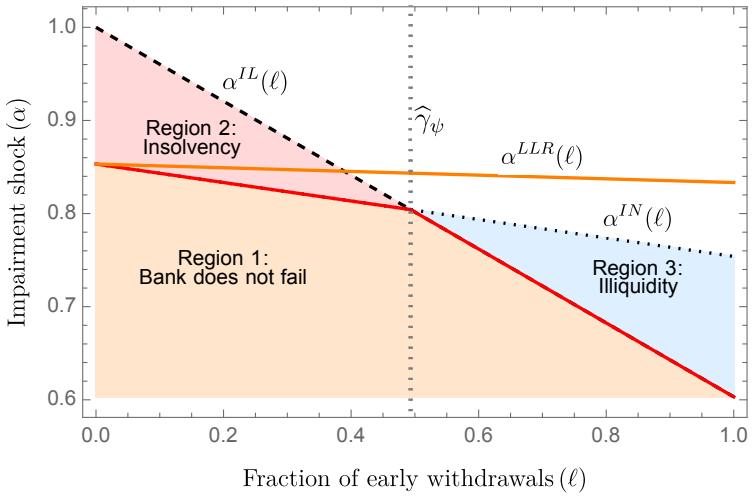


FIGURE 6. Failure conditions with asset illiquidity, $\psi < 1$, for the bank following a successful cyberattack. We also depict the failure threshold with the LLR.

Suppose bank failure is liquidity-driven, $\ell > \widehat{\gamma}_{\psi}$. An alternative to liquidating assets to service withdrawals is for the bank to approach the lender of last resort (LLR) for a bridge loan to service the withdrawals. This loan is at $t = 2$ at a penalty rate $\zeta \geq 1$. It is optimal for the bank to accept LLR support if $(1 - \delta\alpha)RI - \zeta\ell F - (1 - \ell)F > 0$, i.e., at $t = 2$, after the cyberattack, the bank has enough resources to repay the LLR and depositors who did not withdraw at $t = 1$. As depicted in Figure 6, as long as the penalty rate is not too high, $\zeta < 1 + \frac{\delta}{\psi} + \frac{(1-\delta)RI-F}{\ell F}$, the LLR facility leads to an upward shift in the bank's failure threshold, and thereby improving the bank's resilience.

Solving for the equilibrium in this sub-game using the global games approach, we obtain that for small but positive noise, the equilibrium is characterised by the tuple for the bank failure threshold and the run threshold, (α^*, x^*), such that $\alpha^* \equiv \alpha^{IN} - \frac{\zeta-1}{\delta}\frac{\ell(\alpha^*,x^*)F}{RI}$ and $\ell(\alpha^*, x^*) = \gamma$. In the limit $\zeta \to 1$, i.e., the LLR lends without a penalty rate, the failure threshold converges to the α^{IN} and the prospect of coordination failure and runs are completely eliminated. For $\zeta > 1$, the bank's ex ante contribution to cybersecurity is given by the solution to $\lim_{\epsilon\to 0} \frac{\partial\pi}{\partial S}\Big|_{S=S^*} = 0$. We must therefore take the limit of vanishing private noise after deriving the first-order condition.

Proposition 13. *The bank's contribution to cybersecurity is increasing in the penalty rate, i.e.,* $\frac{\partial S^*}{\partial \zeta} > 0$.

An increase in the penalty rate has two effects on the bank's incentives to contribute to cybersecurity. First, the marginal benefit from greater protection is higher since, in the event of a cyberattack, the bank is more likely to fail. And second, the marginal benefit from greater resilience is reduced, which lowers the opportunity cost for contributing to cybersecurity. The two effects reinforce each other leading to the bank increasing its contribution to cybersecurity.

5.3. **Common IT solutions, multiple banks and cybersecurity as a public good.** Our analysis has focused on a single bank using proprietary IT systems. However, in practice, many banks use common IT systems provided by third-party vendors. Oracle Corporation, for example, sells

FLEXCUBE to banks such as Wells Fargo, Citigroup, and HSBC. FLEXCUBE is a suite of software solutions to help streamline and optimise core bank operations including the management of accounts and loans, transfer of funds, and reporting tools.[18]

We extend our baseline model to include $N \geq 2$ identical banks sharing common IT systems. Specifically, each bank runs the same software suite or uses a particular hardware component provided by a third-party vendor. Although vendors also have incentives to contribute to cybersecurity, to identify and patch vulnerabilities, we focus on the role of banks. We seperately consider the role of the vendor in Section 5.3.

An important aspect of cybersecurity is that it also involves the sharing of information once weak points have been detected. In practice, this means publicly releasing information on vulnerabilities to communities such as the National Vulnerability Database (NVD) or the Common Vulnerabilities and Exposures (CVE) system.[19] By eliciting and sharing information, a bank's expenditure on cybersecurity benefits all banks who can then take mitigating actions, making cybersecurity a *public good* (Mester, 2019).

We follow Hirshleifer (1983) and model cybersecurity as a *best-shot* public good,

$$X(\vec{S}) = \max\{S_1, \ldots, S_N\}, \tag{17}$$

where X is the level of cybersecurity available in the economy, given the contributions of the individual banks, $\vec{S} = (S_1, \ldots, S_N)$.[20] Thus, the bank that contributes the most to finding vulnerabilities defines the level of cybersecurity in the system.[21]

[18] Other prominent technology vendors include Finestra and Amazon Web Services (AWS).

[19] While database administrators admit that public vulnerability disclosure can assist attackers, it has been argued that the benefits of information sharing outweigh the costs (Johnson et al., 2016). Tony Sager, Senior Vice President of the Center for Internet Security, is a proponent of this approach, encouraging "one organization's detection to become another's prevention".

[20] The best-shot public good can be derived as a special case of the more general symmetric constant elasticity of substitution function, $X(\vec{S}) = \left[\frac{1}{N}\sum_{i=1}^{N} S_i^{\nu}\right]^{1/\nu}$, by taking the limit $\nu \to \infty$ (Cornes, 1993). The parameter ν characterises the extent to which banks' contributions are substitutes. If $\nu = 1$ then banks' contributions to cybersecurity are perfect substitutes. In the limit $\nu \to -\infty$, they become perfect complements and $X(\vec{S})$ converges to a weakest-link public good (Hirshleifer, 1983).

[21] To motivate the best-shot public good, Hirshleifer (1983) considers the example of anti-missile batteries surrounding a city and defending it from an incoming missile. The consequences of the missile hitting the city is destruction for all. Thus, all that matters is whether the single best defensive shot is good enough to destroy the incoming missile.

In the game between the attacker and banks, given the attacker's effort, A, to find and exploit vulnerabilities, at $t = 1$, banks are successful in uncovering vulnerabilities before the attacker and applying mitigating measures with probability

$$p(A, X(\vec{S})) = \frac{X(\vec{S})}{A + X(\vec{S})}. \tag{18}$$

If, however, with probability $1-p(A, X(\vec{S}))$ the attacker is successful, then a cyberattack is launched on *all* banks where the prize earned by the attacker remains $V > 0$. The common disruption shock suffered by the banks is $\alpha \in [0, 1]$, which is uniformly distributed. If a fraction ℓ_b of the fund managers affiliated with bank $b = 1, \ldots, N$ withdraw, then the bank fails due to illiquidity at $t = 1$ whenever $\alpha > \alpha^{IL}(\ell_b)$. While, if bank b was able to service its debts at $t = 1$, then it may nevertheless fail at $t = 2$ due to insolvency whenever $\alpha > \alpha^{IN}$, where we assume the same dead-weight loss, δ, for all banks.

Assuming there is no overlap in sets of fund managers across the different banks, and that, in the event of a cyberattack, fund managers receive noisy signals regarding the disruption to their bank, we can extend the result of Proposition 1 to define the failure threshold, $\alpha_b^* = \alpha^*$, for each bank, $b = 1, \ldots, N$.

System-wide equilibrium. Turning to the banks' decisions to contribute to cybersecurity at $t = 0$, we can express bank b's expected equity value as a function of its investment, I_b and the level of protection, p, i.e., as $\pi_b(I_b, p) = \pi(I_b, p)$, i.e., which is given in Equation (10). Bank b's optimisation problem may be expressed as

$$\begin{aligned} \max_{\{I_b, p, S_b\}} \quad & \pi_b(I_b, p) \\ \text{subject to} \quad & 1 = I_b + S_b, \\ \text{and} \quad & p = \sqrt{\frac{cX(\vec{S})}{V}}. \end{aligned}$$

Accounting for the incentives of all banks, and assuming a fixed face value of debt, Proposition 14 describes the system-wide equilibrium.

Proposition 14. *In equilibrium, an arbitrarily chosen bank, labelled* $b = 1$*, contributes* $S_1^* > 0$ *to cybersecurity, which solves*

$$\left.\frac{\partial \pi_1 / \partial p}{\partial \pi_1 / \partial I_1}\right|_{S_1 = S_1^*} = \frac{2S_1^*}{p(S_1^*)}. \tag{19}$$

The remaining banks, $b = 2, \ldots, N$*, free-ride on bank 1's contribution to cybersecurity, and so* $S_2^* = \ldots = S_N^* = 0$.

The trade-off faced by bank 1 is identical to the one previously described in Proposition 3 between achieving greater protection against being subject to cyberattacks and being more resilient in the face of cyberattacks. However, at the system-wide level and given the best-shot nature of the public good, all other banks find it optimal to shirk entirely on contributing to cybersecurity. In particular, given bank 1's contribution, S_1^*, any level of contribution $S_b \leq S_1^*$ would not influence the level of cybersecurity at the system-level, given how banks' contributions are amalgamated. So no bank would ever choose a positive level of contribution to cybersecurity that is less than S_1^*. And, no bank would ever choose $S_b > S_1^*$ since all banks are identical and S_1^* is the private optimal. Thus, in equilibrium, no other bank has an incentive to contribute anything to cybersecurity leading them to free-ride on bank 1's contribution.

Social planner's solution. The social planner seeks to maximise the sum of banks' expected equity values minus the social cost of bank failure, which we assume is linearly separable in individual bank failure. While from the perspective of individual banks, cybersecurity is a best-shot public good, the planner considers the *total contribution* of all banks (Varian, 2004). Formally, the

planner's problem may be expressed as

$$\begin{aligned}
\max_{\{\{I_b\},X,\{S_b\}\}} \quad & \sum_{b=1}^{N} \pi_b(I_b,p) - \lambda\Big(1-p\Big)\sum_{b=1}^{N}\Big(1-\alpha_b^*(I_b)\Big) \\
\text{subject to} \quad & 1 = I_b + S_b \quad \forall\, b = 1,...,N\,, \\
& p = \sqrt{\frac{cX}{V}}\,, \\
\text{and} \quad & X = \sum_{b=1}^{N} S_b\,.
\end{aligned}$$

The social planner sets banks' allocations, taking into account their balance sheet conditions and how banks' contributions to cybersecurity influence the level of public good provision.

The solution to the planner's problem, $(S_1^P, S_2^P, \ldots, S_N^P)$, is given by the following system of equations:

$$\frac{\partial\pi_b/\partial X + \lambda(1-\alpha_b^*)}{\partial\pi_b/\partial I_b + \lambda\partial\alpha_b^*/\partial I_b} + \sum_{k\neq b}^{N} \frac{\partial\pi_k/\partial p + \lambda(1-\alpha_k^*)}{\partial\pi_b/\partial I_b + \lambda\partial\alpha_b^*/\partial I_b} = \frac{2X}{p(X)} \qquad \forall\ b = 1,\ldots,N\,. \tag{20}$$

Under symmetry, $S_b^P = S_C^P$ for all banks, $b = 1,\ldots,N$, where the subscript C is used to highlight the common IT infrastructure .

The solution presented in Equation (20) is a version of the Samuelson rule (Samuelson, 1954). The left-hand side of (20) is the planner's marginal rate of substitution. It is the sum of ratios of the marginal impact for each bank from an increase in cybersecurity, to the marginal impact on bank b's profits from increasing its investments, I_b. Embedded in the ratios are the marginal impacts on the social cost of bank failure. This ratio captures the planner's willingness to forego a unit of bank b's private investment for a unit increase in the level of the public good. The planner chooses an allocation such that the level of investment that is given by b in favour of contributing more to cybersecurity is equal to the contribution needed to produce the additional unit of cybersecurity i.e., the right-hand side of Equation (20).

Our analysis thus suggests that it is crucial to map out the sets of common IT infrastructure across banks in order to design policies. Such efforts are underway in several jurisdictions (Brauchle et al., 2020; Deutsche Bundesbank, 2022).

5.4. **Contributions to cybersecurity by vendors.** Our analysis has so far been silent about the vendor's role in contributing to cybersecurity. This assumption stems from the view that software developers are incentivised to prioritise fast product delivery over security. Such incentives are particularly perverse when developers use open-source software to build their solutions (The White House, 2023). However, as exemplified by Microsoft following a series of high-profit security lapses in 2002, concerns over reputation can incentivise vendors to prioritise security. As then CEO Bill Gates wrote in a memo to all Microsoft employees, "*when we face a choice between adding features and resolving security issues, we need to choose security*" (Microsoft, 2012).

While a detailed analysis of such issues is beyond the scope of this paper, we consider the following reduced-form approach. Suppose that the vendor contributes $Y > 0$ to finding vulnerabilities in the IT systems that it has sold to the banks where the marginal cost is normalized to unity. The level of cybersecurity is, thus, given by $X(\vec{S}, Y) = \max\{S_1, \ldots, S_N, Y\}$. Denoting the loss to its reputation in the event of an attack by $\Omega > 0$, the vendor's objective function is

$$U(Y) = -\Big(1 - p(X(\vec{S}, Y))\Big)\Omega - Y\,. \tag{21}$$

The private provision of a best-shot public good comes from the one with the highest benefit-to-cost ratio of doing so, while all others free-ride and contribute nothing (Varian, 2004). For the vendor, this ratio is just Ω, while for the banks, it is given by their marginal rates of substitution between protection and resilience. Thus, if

$$\Omega > \max\left\{ \left.\frac{\partial \pi_1/\partial p}{\partial \pi_1/\partial I_1}\right|_{S_1=0}, \ldots, \left.\frac{\partial \pi_N/\partial p}{\partial \pi_N/\partial I_N}\right|_{S_N=0} \right\}$$

then the vendor has the highest benefit-to-cost ratio and contributes, Y^*, which is given by the solution to

$$\Omega = \frac{2Y^*}{p(Y^*)}, \tag{22}$$

where $p(Y^*) \equiv p(X(\vec{0}, Y^*))$. The banks, in turn, contribute nothing to cybersecurity and so $S_1^* = \ldots, S_N^* = 0$. If, however, this condition is violated, then the vendor no longer has the highest benefit-to-cost ratio and, as Proposition 14 suggests, it instead falls to one of the banks to contribute to cybersecurity while all others, including the vendor, free-ride .

Including the vendor in the normative analysis, the corresponding Samuelson condition for the socially optimal investment in cybersecurity by the vendor is given by the solution to

$$\Omega + \sum_{k=1}^{N} \frac{\partial \pi_k / \partial p + \lambda(1 - \alpha_k^*)}{\partial \pi_k / \partial I_b + \lambda \partial \alpha_b^* / \partial I_b} = \frac{2X(\vec{S}, Y)}{p\Big(X(\vec{S}, Y)\Big)}. \tag{23}$$

Relative to its private solution, the vendor also fails to internalise how its contribution to cybersecurity influences the contributions of other banks and the social cost of bank failure. This suggests that the vendor would also benefit from regulatory interventions that promote greater contributions to cybersecurity. A crucial first step in this direction is for regulators to have oversight over critical technology vendors and set minimum standards for them. In the United Kingdom, for example, the government issued a proposal in 2022 where it outlined giving financial regulators the powers to designate technology vendors as 'critical' and thereby set minimum cybersecurity standards for them (HM Treasury, 2022).

6. Conclusion

We provide an analytical framework to show how cyberattacks may morph into bank runs, and which clarifies the protection-resilience trade-off faced by a bank when deciding on its cybersecurity investment. Since the bank does not take into account the social costs of its distress in the event of a successful attack, there is scope for a planner to correct sub-optimal bank investment in cybersecurity. The socially optimal choice of cybersecurity depends critically on the severity of the threat posed and the nature of bank fragility, and regulators need to be alert to these considerations

when designing ex ante and ex post cyber regulations. We show how operational resilience standards, red-teaming, subsidies for enhancing cyber-capability, and negligence rules with penalties may facilitate socially desirable investment in cybersecurity. Financial regulators are increasingly focused on these measures to address cyber concerns, and our work provides a formal basis to understand the implications of such initiatives.

The core insights of our analysis carry over into more general settings. The introduction of Knightian uncertainty, the presence of a lender of last resort, multiple banks sharing common IT platforms, and the presence of a technology vendor do not obviate our findings. Future work might consider how industry initiatives such as Sheltered Harbor and cyber insurance markets shape the provision of cybersecurity. Deeper analysis of the drivers of the deadweight losses from cyberattacks is also warranted. Arguably, cyberattacks compromise the ability of a bank to both make and keep secret information and it is the loss of such information that is, ultimately, most devastating for the integrity of the financial system.

Appendix A. Proofs

A.1. **Proof of Lemma 1.** The bank fails due to insolvency whenever

$$\alpha > \alpha^{IN} \equiv \frac{1}{\delta}\left(1 - \frac{F}{RI}\right), \tag{24}$$

and it fails due to illiquidity whenever

$$\alpha > \alpha_{IL}(\ell) \equiv 1 - \frac{\ell F}{RI}. \tag{25}$$

While α^{IN} is invariant to the proportion of withdrawals, the threshold $\alpha^{IL}(\ell)$ is decreasing in ℓ. The proportion of withdrawals, $\widehat{\gamma}$, for which the two failure conditions intersect is given by $\alpha^{IL}(\hat{\gamma}) = \alpha^{IN}$, i.e.,

$$\frac{1}{\delta}\left(1 - \frac{F}{RI}\right) = 1 - \frac{\widehat{\gamma} F}{RI}, \tag{26}$$

which on rearranging yields Equation (5).

A.2. **Proof of Proposition 1.** The proof is in three steps. First, we show that the dominance regions at $t = 1$ are well defined. If all fund managers withdraw early, $\ell = 1$, then the illiquidity failure threshold is given by $\alpha^{IL}(1) = 1 - \frac{F}{RI}$, where $\alpha^{IL}(1) < \alpha^{IN}$ since $\delta < 1$. If, however, no fund manager withdraws at $t = 1$, then $\ell = 0$. In this case, the bank never fails due to illiquidity since $\alpha^{IL}(0) > 1$. But the bank can, nevertheless, fail at $t = 2$ due to insolvency whenever $\alpha > \alpha^{IN}$, since, under Assumption 1, $\alpha^{IN} < 1$.

It also follows from Assumption 1 that $\underline{\alpha} \equiv \alpha^{IL}(1) > 0$ is the largest shock that bank b can withstand even if all fund managers withdraw early. When $\alpha \in [0, \underline{\alpha}]$, fund managers have a dominant strategy to roll over their claims. Next, let $\overline{\alpha} \equiv \alpha^{IN} < 1$ denote the upper dominance bound, beyond which the bank fails regardless of the number of fund managers who withdraw early. When $\alpha \in [\overline{\alpha}, 1]$, fund managers have a dominant strategy to withdraw early. Finally, for $\alpha \in (\underline{\alpha}, \overline{\alpha})$, if the shock were common knowledge, the run dynamics of fund managers would be characterised by multiple, self-fulfilling equilibria, as shown in Figure 7.

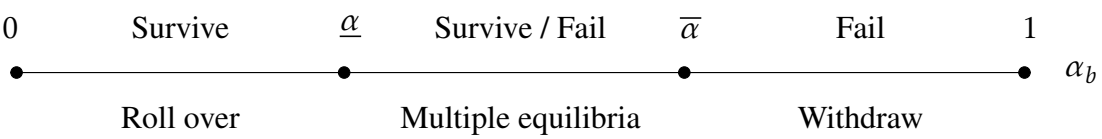


FIGURE 7. Tripartite classification of the accessibility shock.

Second, we define a threshold strategy which, for well defined dominance bounds and sufficiently precise private information, survives iterated deletion of strictly dominated strategies (Morris and Shin, 2003; Frankel et al., 2003). Denote this threshold point x^*, with corresponding strategy

$$s(x_k) = \begin{cases} \text{withdraw} & \text{if } x_k > x^*, \\ \text{roll over} & \text{if } x_k \leq x^*. \end{cases} \tag{27}$$

Finally, we characterise this equilibrium. With switching point x^*, the proportion of fund managers who withdraw at $t = 1$, given some realisation of the shock α is

$$\ell(\alpha, x^*) = \Pr(x_k > x^* | \alpha) = 1 - H\Big(x^* - \alpha\Big), \tag{28}$$

For $\gamma > \widehat{\gamma}$, the failure threshold, α^*, solves

$$\alpha^* = 1 - \ell^*(\alpha^*, x^*)\frac{F}{RI}. \tag{29}$$

For a given x^*, the left-hand side of Equation (29) is strictly increasing in α^* and is unbounded over the entire unit interval, while the right-hand side is decreasing in α^* and is bounded between 1 and $1 - \frac{F}{RI}$. Hence, there exists a unique failure threshold, α^*.

The posterior distribution of the shock conditional on the private signal can be derived using Bayes' rule. At the threshold signal x^*, fund managers are indifferent between withdrawing and rolling over, so that

$$\gamma = \Pr(\alpha \leq \alpha^* | x_k = x^*). \tag{30}$$

For small ϵ, this can be written as $\gamma = 1 - H\Big(x^* - \alpha^*\Big)$. The indifference condition therefore implies $x^* - \alpha^* = H^{-1}\Big(1 - \gamma\Big)$. Inserting this into $\ell\Big(\alpha^*, x^*\Big)$, the withdrawal proportion at the threshold α^*

becomes $\ell\Big(\alpha^*, x^*\Big) = 1 - H\Big(x^* - \alpha^*\Big) = 1 - H\Big(H^{-1}\Big(1-\gamma\Big)\Big) = \gamma$. Thus, for $\gamma \geq \widehat{\gamma}$, we have that $\alpha^* = \alpha^{IL}(\gamma)$. While for $\gamma < \widehat{\gamma}$, the bank fails only due to insolvency and so $\alpha^* = \alpha^{IN}$.

A.3. **Proof of Proposition 2.** The marginal benefit from issuing debt is given by

$$\frac{\partial \pi}{\partial D} = p[R-F] + (1-p)\left\{\int_0^{\alpha^*}\Big[(1-\delta\alpha)R - F\Big]d\alpha + EV(\alpha^*)\frac{\partial \alpha^*}{\partial D}\mathbf{1}_{\gamma>\widehat{\gamma}}\right\}.$$

For $\gamma > \widehat{\gamma}$, we have that $\frac{\partial \alpha^*}{\partial D} = -\frac{\gamma F}{R}\frac{S}{(D-S)^2} > 0$. And so the expression multiplying into $(1-p)\mathbf{1}_{\gamma>\widehat{\gamma}}$ is strictly positive. Moreover, we have that

$$\int_0^{\alpha^*}\Big[(1-\delta\alpha)R - F\Big]d\alpha > (1-\delta\alpha^*)R - F.$$

For $\alpha^* = \alpha^{IN}$, we get that $(1-\delta\alpha^{IN})R - F = F\left(\frac{D}{I} - 1\right) > 0$. And so the expected intermediation margin in the event of cyberattack is strictly positive. And since $\alpha^{IL}(\gamma) < \alpha^{IN}$, the expected intermediation margin in the event of a cyberattack is also positive when bank failure is driven by illiquidity. And so, in sum, for $R - F > 0$, we have that $\frac{\partial \pi}{\partial D} > 0$, implying that it is optimal for the bank to issue as much debt as possible, i.e., $D^* = 1$.

A.4. **Proof of Proposition 3.** The Lagrange equation for the bank's problem is given by

$$\mathcal{L} = \pi(I,p) + \mu(1-I-S) + \phi\left(p - \sqrt{\frac{cS}{V}}\right), \tag{31}$$

where μ and ϕ are the Lagrange multipliers for the balance sheet constraint and protection technology, respectively. The bank's problem is to maximise $\mathcal{L}$ by choosing the quantities I, p, and S. The first-order conditions yield

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial I} = 0 \quad &\Longrightarrow \quad \mu = \frac{\partial \pi}{\partial I} \\
\frac{\partial \mathcal{L}}{\partial p} = 0 \quad &\Longrightarrow \quad \phi = -\frac{\partial \pi}{\partial p} \\
\frac{\partial \mathcal{L}}{\partial S} = 0 \quad &\Longrightarrow \quad \mu = \phi\frac{1}{2}\sqrt{\frac{c}{SV}}
\end{aligned}$$

Putting these together, we obtain that the bank's contribution to cybersecurity, S^*, solves

$$\left.\frac{\partial\pi/\partial p}{\partial\pi/\partial I}\right|_{S=S^*} = \frac{2S^*}{p(S^*)}.$$

We can, thus, express the first-order condition solving for S^* as

$$\begin{aligned}\frac{\partial\pi}{\partial S} &= \frac{\partial p}{\partial S}\left[R(1-S)-F-\int_0^{\alpha^*(S)} EV(\alpha)d\alpha\right] \\ &- R\left[p(S)+\Big(1-p(S)\Big)\int_0^{\alpha^*(S)}(1-\delta\alpha)d\alpha\right]+\Big(1-p(S)\Big)EV(\alpha^*)\frac{\partial\alpha^*}{\partial S}\mathbf{1}_{\gamma>\widehat{\gamma}}. \qquad (32)\end{aligned}$$

Next, we need to argue that S^* is a maximum. To this end, note that since $p(S)$ satisfies the Inada conditions, it follows that $\lim_{S\to 0}\frac{\partial\pi}{\partial S} = +\infty > 0$ and $\pi(S=0) > 0$. Moreover, it follows from economic reasoning that the bank would never choose $S \geq \overline{S}$ in equilibrium. If, by contradiction, we suppose that the bank choose $S = \overline{S}$, then even if the vulnerabilities are uncovered and patched, the bank immediately defaults since it has generated zero asset returns. And whenever the attacker is successful, the bank always defaults, too. Thus, for $S = \overline{S}$, we have that $\pi(S=\overline{S}) \leq 0$ (with strict inequaility for $\gamma > \widehat{\gamma}$), implying that this level of contribution to cybersecurity is not feasible. And so by Rolle's theorem, at least one optimum exists within $(0, \overline{S})$.

Finally, we must show that S^* is unique. The second-order condition is given by

$$\begin{aligned}\frac{\partial^2\pi}{\partial S^2} &= \frac{\partial^2 p}{\partial S^2}\left[R(1-S)-F-\int_0^{\alpha^*(S)} EV(\alpha)d\alpha\right]-2\frac{\partial p}{\partial S}R\left[1-\int_0^{\alpha^*(S)}(1-\delta\alpha)d\alpha\right] \\ &- \Big(1-p(S)\Big)\frac{\partial\alpha^*}{\partial S}R\Big(1-\delta\alpha^*(S)\Big) \\ &- \frac{\partial\alpha^*}{\partial S}\left(EV(\alpha^*(S))\left[\frac{p(S)}{S}-\frac{1-p(S)}{1-S}\right]+\frac{F(1-\gamma\delta)}{1-S}\right)\times\mathbf{1}_{\gamma>\widehat{\gamma}}\end{aligned}$$

Under the assumption $\frac{p(S)}{1-p(S)} > \frac{2S}{1-S}$, the term on the last line above is positive and strictly decreasing in R. By a similar logic, as R increases, the term on the second line above becomes smaller. In contrast, the terms on the first line are negative and become more negative as R increases. It thus follows that for $R > \underline{R}_\pi$, which solves $\left.\frac{\partial^2\pi}{\partial S^2}\right|_{R=\underline{R}_\pi} = 0$, the bank's expected equity value is globally concave, implying that the optimum is unique and a maximum.

Finally, we argue that $\widehat{\gamma}$ is well defined. The critical threshold is implicitly defined as

$$\widehat{\gamma} = \frac{1}{\delta} - \left(\frac{1}{\delta} - 1\right) \frac{R\Big(1 - S^*(\widehat{\gamma})\Big)}{F}. \tag{33}$$

Suppose that a marginal increase in rollover risk increases banks' contributions to cybersecurity (we verify this below). And so the right-hand side of Equation (33) is increasing in $\widehat{\gamma}$. At $\widehat{\gamma} = 0$, the left-hand side of Equation (33) is smaller than the right-hand side. And at $\widehat{\gamma} = 1$, the right-hand side is less than one. To see this, note that $S^* < \overline{S}$. Thus, the highest possible value that the right-hand side can achieve is $\frac{1}{\delta} - \left(\frac{1}{\delta} - 1\right) \frac{R(1-\overline{S})}{F}$. To see that this is (weakly) less than one, suppose by way of contradiction that it is strictly greater than one, i.e.,

$$\frac{1}{\delta} - \left(\frac{1}{\delta} - 1\right) \frac{R\Big(1 - \overline{S}\Big)}{F} > 1 \quad \leftrightarrow \quad (1-\delta)\left(1 - \frac{R(1-\overline{S})}{F}\right) > 0.$$

However, given the definition of $\overline{S}$, the above expression is exactly equal to zero, implying a contradiction. And so the right-hand side of Equation (33) is never larger than one. Thus, by the intermediate value theorem, we have a well defined threshold, $\widehat{\gamma}$.

To obtain how a marginal change in the face value of debt impacts the bank's contribution to cybersecurity, note that

$$\begin{aligned}
\frac{\partial^2 \pi}{\partial S \partial F} &= \frac{p(S)}{2S}\left\{-1 + \alpha^*(S) - EV(\alpha^*)\frac{\partial \alpha^*}{\partial F}\mathbf{1}_{\gamma > \widehat{\gamma}}\right\} \\
&+ \Big(1 - p(S)\Big)\left\{\left(\frac{\partial EV(\alpha^*)}{\partial F}\frac{\partial \alpha^*}{\partial S} + EV(\alpha^*)\frac{\partial^2 \alpha^*}{\partial S \partial F}\right)\mathbf{1}_{\gamma > \widehat{\gamma}} - R\Big(1 - \delta\alpha^*(S)\Big)\frac{\partial \alpha^*}{\partial F}\right\}
\end{aligned}$$

The terms on the second line may be expressed as

$$\begin{aligned}
&\left\{\frac{(1-\gamma\delta)\gamma F}{R(1-S)^2} - \left[\Big(1 - \delta\alpha^*(S)\Big)R(1-S) - F\right]\frac{\gamma}{R(1-S)^2}\right\}\mathbf{1}_{\gamma > \widehat{\gamma}} + R\Big(1 - \delta\alpha^*(S)\Big)\left|\frac{\partial \alpha^*}{\partial F}\right| \\
= &\left\{\frac{(1-\gamma\delta)\gamma F}{R(1-S)^2} + \frac{\gamma F}{R(1-S)^2}\right\}\mathbf{1}_{\gamma > \widehat{\gamma}} + R\Big(1 - \delta\alpha^*(S)\Big)\left|\frac{\partial \alpha^*}{\partial F}\right|(1 - \mathbf{1}_{\gamma > \widehat{\gamma}}) > 0.
\end{aligned}$$

Combining the terms together again, in the case $\gamma \leq \widehat{\gamma}$, we obtain

$$\begin{aligned}\frac{\partial^2 \pi}{\partial S \partial F} &= -\frac{p(S)}{2S}\Big(1-\alpha^*(S)\Big)+\frac{1-p(S)}{1-S}\Big(1-\delta\alpha^*(S)\Big) \\ &= -\left\{\left[\frac{p(S)}{2S}-\frac{1-p(S)}{1-S}\right]\Big(1-\alpha^*(S)\Big)-\frac{1-p(S)}{1-S}\left(\frac{1}{\delta}-1\right)\right\},\end{aligned}$$

which is positive as long as

$$\delta < \frac{1}{1+\left[\frac{p(S)}{1-p(S)}\frac{1-S}{2S}-1\right]\Big(1-\alpha^*(S)\Big)}.$$

Since the right-hand side is decreasing in δ, while the left-hand side is increasing, it follows that there exists a unique $\overline{\delta}$ such that for $\delta < \overline{\delta}$, we obtain that $\frac{\partial^2 \pi}{\partial S \partial F} > 0$, and so $\frac{\partial S^*}{\partial F} > 0$.

In the case $\gamma > \widehat{\gamma}$,

$$\begin{aligned}\frac{\partial^2 \pi}{\partial S \partial F} &= \frac{p(S)}{2S}\frac{\gamma}{R(1-S)}\Big[EV(\alpha^*)-F\Big]+\Big(1-p(S)\Big)\frac{(2-\gamma\delta)\gamma F}{R(1-S)^2} \\ &= \frac{p(S)}{2S}(1-\delta)\gamma-\frac{(2-\gamma\delta)F}{R(1-S)}\left(\frac{p(S)}{2S}-\frac{1-p(S)}{1-S}\right),\end{aligned}$$

which is strictly increasing in R. Thus, for $R > \underline{R}_F$, where $\underline{R}_F$ solves $\frac{\partial^2 \pi}{\partial S \partial F}\Big|_{R=\underline{R}_F} = 0$, we obtain that $\frac{\partial S^*}{\partial F} > 0$. In what follows we define $\underline{R} \equiv \max\{\underline{R}_F, \underline{R}_\pi\}$

A.5. **Proof of Lemma 3.** Given F, the value of a debt claim is

$$\mathcal{V}(F) = \Big(p(S)+\Big(1-p(S)\Big)\alpha^*(F,S)\Big)F. \tag{34}$$

For $F = r$, we have $\mathcal{V}(r) < r$ since $\alpha^* < 1$. And $\lim_{F\to\infty}\mathcal{V}(F) > r$. And so by the intermediate value theorem, it follows that there exists an F^* that solves $\mathcal{V}(F^*) = r$. For this solution to be unique, we require that $\frac{\partial \mathcal{V}}{\partial F} > 0$. In general, we have that

$$\frac{\partial \mathcal{V}}{\partial F} = p(S)+\Big(1-p(S)\Big)\left[\alpha^*(S)+F\frac{\partial \alpha^*}{\partial F}\right],$$

which on rearranging implies that $\frac{p(S)}{1-p(S)} > -\left[\alpha^*(S) + F\frac{\partial \alpha^*}{\partial F}\right]$. For $\alpha^*(S) + F\frac{\partial \alpha^*}{\partial F} > 0$, the condition is trivially satisfied. This is guaranteed whenever $R > \underline{R}$. Next, note that

$$\frac{1}{F}\frac{\partial \mathcal{V}}{\partial S} = \frac{\partial p}{\partial S}\Big(1 - \alpha^*(S)\Big) + \Big(1 - p(S)\Big)\frac{\partial \alpha^*}{\partial S}. \tag{35}$$

For $\gamma > \widehat{\gamma}$, we have that $\frac{\partial \alpha^*}{\partial S} = \frac{\alpha^*(S)-1}{1-S}$, and so under the assumption $\frac{p(S)}{1-p(S)} > \frac{2S}{1-S}$, we obtain that $\frac{\partial \mathcal{V}}{\partial S} > 0$, and so $\frac{\partial F^*}{\partial S} < 0$. In contrast, for $\gamma < \widehat{\gamma}$, we have that

$$\frac{1}{F}\frac{\partial \mathcal{V}}{\partial S} \quad = \quad \left[\frac{p(S)}{2S} - \frac{1-p(S)}{1-S}\right]\Big(1 - \alpha^*(S)\Big) - \frac{1-p(S)}{1-S}\left(\frac{1}{\delta} - 1\right),$$

which is negative as long as $\delta < \overline{\delta}$. And so we obtain that $\frac{\partial \mathcal{V}}{\partial S} < 0$, and so $\frac{\partial F^*}{\partial S} > 0$.

A.6. **Proof of Proposition 4.** We now argue that there is a single crossing between the curves $S^*(F)$ and $F^*(S)$. To this end, note that $S^*(F = r) < 1 < r < F^*(S = 0)$. So, at $S^*(F = r) < F^*(S = S^*(F = r))$. Thus, the curve $F^*(S)$ starts off "above" $S^*(F)$. Next, in the limit $S \to 1$, we get that $F^* \to \frac{r}{p(1)}$. And insofar that $\frac{r}{p(1)} < R$, where R is the supremum of the domain for the bank's contribution to cybersecurity, then the curve $S^*(F)$ ends up "above" $F^*(S)$. Provided that the two curves are strictly monotone, there is a single crossing that defines our joint equilibrium, (S^{**}, F^{**}). As we have previously argued, at this equilibrium, the threshold $\widehat{\gamma}(F^{**}, S^{**})$ is well defined.

A.7. **Proof of Proposition 5.** In deriving the comparative statics for the joint equilibrium, we first look at how the bank's best-response correspondence shifts following a marginal change in each exogenous variable.

Attacker's effort cost. The cross derivative of the bank's expected profits with respect to its contribution to cybersecurity and the attacker's cost of effort is

$$\frac{\partial^2 \pi}{\partial S \partial c} \quad = \quad \frac{\partial p}{\partial c}\left\{\frac{\Delta EV}{2S} - R\left(1 - \int_0^{\alpha^*(S)} (1 - \delta\alpha) d\alpha\right) - EV_b(\alpha^*)\frac{\partial \alpha^*}{\partial S}\mathbf{1}_{\gamma > \widehat{\gamma}}\right\},$$

where $\Delta EV \equiv \left[R(1-S) - F - \int_0^{\alpha^*(S)} EV(\alpha)d\alpha\right]$. Evaluating the above expression at S^*, we get

$$\left.\frac{\partial^2 \pi}{\partial S \partial c}\right|_{S=S^*} = \frac{\partial p/\partial c}{p(S^*)}\left\{R\int_0^{\alpha^*(S^*)}(1-\delta\alpha)d\alpha - EV(\alpha^*)\frac{\partial \alpha^*}{\partial S}\mathbf{1}_{\gamma>\widehat{\gamma}}\right\} > 0\,.$$

Thus, by the implicit function theorem, we have $\frac{\partial S^*}{\partial c} > 0$.

Next, a marginal increase in c leads to a downward shift in the schedule for the face value of debt. To see this, note that $\frac{\partial F^*}{\partial c} = -\frac{\partial \mathcal{V}/\partial c}{\partial \mathcal{V}/\partial F}$, where $\frac{\partial \mathcal{V}}{\partial c} = F\frac{\partial p}{\partial c}\Big(1-\alpha^*(S)\Big) > 0$. And so $\frac{\partial F^*}{\partial c} < 0$, which is a countering effect to that on the $S^*(F)$ schedule. Insofar that $\frac{\partial S^*}{\partial c} + \frac{\partial S^*}{\partial F}\frac{\partial F^*}{\partial c} > 0$, then the direct effect on the bank's incentives to contribution to cybersecurity outweigh the indirect effects via the pricing of debt. Importantly, in this condition, the term $\frac{\partial p}{\partial c}$ drops out as it appears in both $\frac{\partial S^*}{\partial c}$ and $\frac{\partial F^*}{\partial c}$. Consequently, for $c < \tilde{c}^{**}$, where $\tilde{c}^{**}$ solves $\left.\frac{\frac{\partial S^*}{\partial c} + \frac{\partial S^*}{\partial F}\frac{\partial F^*}{\partial c}}{\frac{\partial p}{\partial c}}\right|_{S=S^{**},F=F^{**},c=\tilde{c}^{**}} = 0$, we have that the direct effect dominates and so $\frac{\partial S^{**}}{\partial c} > 0$.

Deadweight loss. The cross derivative of π with respect to S and δ is

$$\begin{aligned}\frac{\partial^2 \pi}{\partial S \partial \delta} &= \frac{\partial p}{\partial S}R(1-S)\frac{\alpha^*(S)^2}{2} - \Big(1-p(S)\Big)\Bigg[R\Big(1-\delta\alpha^*(S)\Big)\frac{\partial \alpha^*}{\partial \delta} - R\frac{\alpha^*(S)^2}{2} \\ &- \frac{\partial EV(\alpha^*)}{\partial \delta}\frac{\partial \alpha^*}{\partial S}\mathbf{1}_{\gamma>\widehat{\gamma}}\Bigg],\end{aligned}$$

where

$$\frac{\partial EV(\alpha^*)}{\partial \delta} = -\alpha^*(S)R(1-S) < 0\,.$$

Since $\frac{\partial \alpha^*}{\partial \delta} < 0$, it follows that $\frac{\partial^2 \pi}{\partial S \partial \delta} > 0$ and so by the implicit function theorem, $\frac{\partial S^*}{\partial \delta} > 0$.

Turning to the investors' schedule for the face value of debt, for $\gamma < \widehat{\gamma}$, we have that $\frac{\partial \mathcal{V}}{\partial \delta} = F\Big(1-p(S)\Big)\frac{\partial \alpha^*}{\partial \delta} < 0$, which implies that $\frac{\partial F^*}{\partial \delta} > 0$. Insofar that $F^*(S)$ is increasing, the upward shift in the schedule reinforces the effect of a marginal increase in δ on the $S^*(F)$ schedule. For $\gamma > \widehat{\gamma}$, note that $\mathcal{V}$ does not depend on δ, and so the only effect stems from that on the $S^*(F)$ schedule. Thus, in sum, we have that $\frac{dS^{**}}{d\gamma} > 0$.

Rollover risk. Assuming $\gamma > \widehat{\gamma}$, the cross derivative of π with respect to S and γ is given by

$$\begin{aligned}\frac{\partial^2 \pi}{\partial S \partial \gamma} &= -\frac{\partial p}{\partial S} EV(\alpha^*)\frac{\partial \alpha^*}{\partial \gamma} + \Big(1-p(S)\Big)\Bigg\{-R\Big(1-\delta\alpha^*(S)\Big)\frac{\partial \alpha^*}{\partial \gamma} + \frac{\partial EV(\alpha^*)}{\partial \gamma}\frac{\partial \alpha^*}{\partial S} \\ &+ EV(\alpha^*)\frac{\partial^2 \alpha^*}{\partial S \partial \gamma}\Bigg\}.\end{aligned}$$

Since $\partial\alpha^*/\partial\gamma < 0$, it follows that the first term is strictly positive. Next, we can re-write the terms that multiply into $1-p(S)$ as

$$\begin{aligned}&R\left(1-\delta\left(1-\frac{\gamma F}{R(1-S)}\right)\right)\frac{F}{R(1-S)} - \frac{(\gamma F)^2}{R(1-S)^2} - EV(\alpha^*)\frac{F}{R(1-S)^2} \\ =\ &\frac{(1-\delta)F}{R(1-S)} + \frac{\delta\gamma(F)^2}{R(1-S)^2} - \frac{\delta\gamma(F)^2}{R(1-S)^2} - \frac{(1-\delta)F}{R(1-S)} + \frac{(1-\gamma\delta)(F)^2}{R(1-S)^2} > 0.\end{aligned}$$

Thus, the cross derivative is strictly positive and so by the implicit function theorem, $\frac{\partial S^*}{\partial \gamma} > 0$.

Turning to the investors' schedule for the face value of debt, for $\gamma < \widehat{\gamma}$, this is independent of γ. And so the only effect stems from that on the $S^*(F)$ schedule. While for $\gamma > \widehat{\gamma}$, we obtain $\frac{\mathcal{V}}{\partial \gamma} = F\Big(1-p(S)\Big)\frac{\partial \alpha^*}{\partial \gamma} < 0$, which implies that $\frac{\partial F^*}{\partial \gamma} > 0$. This upward shift in the $F^*(S)$ schedule reinforces the effect induced on the $S^*(F)$ schedule. Thus, in sum, we obtain $\frac{dS^{**}}{d\gamma} > 0$.

A.8. **Proof of Propositions 6 and 7.** Let Δ denote the difference between the planner's and bank's objective functions, i.e.,

$$\Delta \equiv -\lambda\Big(1-p(S)\Big)\Big(1-\alpha^*(S)\Big).$$

It follows that

$$\frac{1}{\lambda}\frac{\partial \Delta}{\partial S} = \frac{p(S)}{2S}\Big(1-\alpha^*(S)\Big) + \Big(1-p(S)\Big)\frac{\partial \alpha^*}{\partial S}.$$

Insofar that $\frac{\partial \Delta}{\partial S} > 0$, it implies that the planner's incentives for contributions to cybersecurity are greater than those for the bank. To further our analysis, we consider the two cases where bank failure is driven by insolvency, $\gamma < \widehat{\gamma}$, and when it is driven by illiquidity, $\gamma > \widehat{\gamma}$.

Insolvency. Using the definition of $\alpha^*(S)$, we obtain

$$\frac{1}{\lambda}\frac{\partial \Delta}{\partial S} = \Big(1-\alpha^*(S)\Big)\left(\frac{p(S)}{2S} - \frac{1-p(S)}{1-S}\right) - \left(\frac{1}{\delta}-1\right)\frac{1-p(S)}{1-S}.$$

In the limit $c \to 0$, the probability with which the bank wins the contest converges to zero, i.e., $\lim_{c\to 0} p(S) = 0$. This, in turn, implies that $\lim_{c\to 0} \frac{\partial \Delta}{\partial S} < 0$, and so $S^P(F) < S^*(F)$, i.e., the bank contributes excessively to cybersecurity.

While in the limit $c \to \infty$, the bank is almost surely going to win the contest, i.e., $\lim_{c\to\infty} p(S) = 1$. And so $\lim_{c\to\infty} \frac{\partial \Delta}{\partial S} > 0$, implying that $S^P(F) > S^*(F)$ and there is under-investment in cybersecurity by the bank. Finally, since

$$\frac{1}{\lambda}\frac{\partial^2 \Delta}{\partial S \partial c} = \frac{1}{2}\frac{p(S)}{c}\left\{\Big(1-\alpha^*(S)\Big)\left(\frac{1}{2S} + \frac{1}{1-S}\right) + \left(\frac{1}{\delta}-1\right)\frac{1}{1-S}\right\} > 0,$$

it implies that there is a well defined threshold $\hat{c}(S,F)$ at which the derivative $\frac{1}{\lambda}\frac{\partial \Delta}{\partial S}$ switches sign. With a slight abuse of notation in what follows, we consider $\hat{c}^{**} \equiv \hat{c}(S^{**}, F^{**})$ to be an equilibrium quantity.

Illiquidity. In this case, we obtain

$$\frac{1}{\lambda}\frac{\partial \Delta}{\partial S} = \Big(1-\alpha^*(S)\Big)\left(\frac{p(S)}{2S} - \frac{1-p(S)}{1-S}\right) < 0,$$

and so $S^P(F) > S(F)$ implying that the bank contributes too little to cybersecurity. Moreover, note that

$$\frac{1}{\lambda}\frac{\partial^2 \Delta}{\partial S \partial c} = \frac{\partial p}{\partial c}\Big(1-\alpha^*(S)\Big)\frac{1+S}{2S(1-S)} > 0,$$

since $\frac{\partial p}{\partial c} > 0$. And so the difference $|S^P(F) - S^*(F)|$ is increasing in c.

A.9. **Proof of Proposition 8.** The planner's benchmark is derived in Proposition 6, which shows that the bank over-invests in cybersecurity when $\gamma < \widehat{\gamma}$. Proposition 3 suggests that $\frac{\partial \pi}{\partial S} > 0$ at $S^P < S^*$. Therefore, the introduction of a constraint $S \le S^P$ through the enforcement of minimum operational resilience standards is binding and the bank allocates exactly $S = S^P$ towards cybersecurity.

A.10. **Proof of Proposition 9.** By Proposition 6, when failure is driven by illiquidity, $\gamma \geq \widehat{\gamma}$, the level of protection offered by the bank in the laissez-faire equilibrium is too low. Let the red-team test perturb the contest between bank and attacker as follows so that the probability of a successful patch by the bank is:

$$p(A, S; \rho) = \frac{S(1+\rho)}{A + S(1+\rho)}. \tag{36}$$

The attacker's level of effort is

$$A^*(S;\rho) = \begin{cases} \sqrt{\frac{S(1+\rho)V}{c}} - S(1+\rho) & \text{if } S(1+\rho) < \frac{V}{c} \\ 0 & \text{otherwise.} \end{cases}$$

Substituting $A^*(S;\rho)$ into (36) determines the level of protection as a function of allocation towards cybersecurity by the bank:

$$p(A^*(S;\rho), S) = \sqrt{\frac{cS(1+\rho)}{V}}.$$

Under these conditions for the contest, the bank optimally allocates its cybersecurity contributions. The cross derivative of π with respect to S and ρ is

$$\frac{\partial^2 \pi}{\partial S \partial \rho} = \frac{\partial p}{\partial \rho}\left\{\frac{\Delta EV}{2S} - R\left(1 - \int_0^{\alpha^*(S)} (1-\delta\alpha)d\alpha\right) - EV(\alpha^*)\frac{\partial \alpha^*}{\partial S}\mathbb{1}_{\gamma > \widehat{\gamma}}\right\},$$

where $\Delta EV \equiv \left[R(1-S) - F - \int_0^{\alpha^*(S)} EV(\alpha)d\alpha\right]$ as before. Evaluating the above expression at S^*, we get

$$\left.\frac{\partial \pi}{\partial S \partial \rho}\right|_{S=S^*} = \frac{\partial p}{\partial \rho}\left\{\left(\frac{1-p}{p}\right)R + \frac{1}{p}\left(\int_0^{\alpha^*} R(1-\delta\alpha)d\alpha - EV(\alpha^*)\frac{\partial \alpha^*}{\partial S}\mathbb{1}_{\gamma > \widehat{\gamma}}\right)\right\} > 0.$$

Thus, the cross derivative is strictly positive and so by the implicit function theorem, $\frac{\partial S^*}{\partial \rho} > 0$. The final step is to set ρ such that $S^*(\rho) = S^P$.

A.11. **Proof of Proposition 10.** Fix $\gamma > \widehat{\gamma}$. With a subsidy scheme in place, the bank's expected profits include subsidy, $\sigma(S)$, and lump-sum tax, τ,

$$\pi(S) = p(S)(RI - F + \sigma(S) - \tau) + (1 - p(S))\int_0^{\alpha^*(S)} (EV(\alpha) + \sigma(S) - \tau)\, d\alpha. \tag{37}$$

For the subsidy to elicit the social optimum, the individual bank's marginal rate of substitution should equal the planner's. This requires the subsidy to be structured as follows

$$\sigma^* = \lambda . \tag{38}$$

To see this, note that substituting $\sigma = \lambda$ into the bank's expected profit function in (37) recovers the planner's objective function.

To fund these subsidies, a uniform lump-sum tax is set such that

$$\tau = \sigma^* . \tag{39}$$

The expected subsidy payout, $\Sigma \equiv \lambda[p(S^*) + (1 - p(S^*))\alpha^*(S^*)]$, is impacted by rollover risk. As we show in the proof of Proposition 6, the degree of under-investment is increasing in γ. Since the planner's problem is unaffected by the subsidy scheme, it suffices to show that

$$\frac{\partial^2 \Delta}{\partial S \partial \gamma} = \frac{\partial^2 \Sigma}{\partial S \partial \gamma} = -\lambda \frac{\partial \alpha^*}{\partial \gamma} \left[\frac{p(S)}{2S} - \frac{1 - p(S)}{1 - S} \right] > 0 .$$

The effect of a marginal increase in rollover risk on the distance between the planner's and private solution is equal to the marginal increase in the subsidy required to bring about the planner's solution.

A.12. **Proof of Proposition 11.** With a negligence rule in place, expected profits include a penalty, $\kappa(S)$, that is implemented conditional on a successful cyberattack

$$\pi(S) = p(S)(RI - F) \tag{40}$$

$$+ \quad (1 - p(S)) \int_0^{\alpha^*(I)} \Big[RI(1 - \alpha\,\delta) - F - \kappa(S) \Big] d\alpha . \tag{41}$$

The introduction of a negligence rule lowers expected profits in all events where a cyberattack is successful and the bank survives, increasing the relative benefits from investing more in cybersecurity. For the penalty to be effective in eliciting the social optimum, it must satisfy two conditions:

$$\frac{RI - F}{RI\delta} - \frac{\kappa(S)}{RI\delta} > \alpha^*(I) \tag{42}$$

i.e., the penalty is not so large that it leads the bank to fail for any successful attack, and

$$\pi(S^P) \geq \pi(S),$$

for all $S \leq S^P$. That is, the penalty should be large enough that profits subject to a negligence rule under the social optimum are preferable to those in the laissez faire optimum.

Suppose the penalty is structured in the following way

$$\kappa(S;\gamma) = \begin{cases} \kappa(S;\gamma) & \text{if } S^*(\gamma) < S^P(\gamma) \\ 0 & \text{otherwise}. \end{cases} \tag{43}$$

Then the penalty, $\kappa^*(S;\gamma)$, that delivers the social optimum is given by

$$\kappa^*(S;\gamma) = \frac{\lambda(1-\alpha^*(S))}{\alpha^*(S)}. \tag{44}$$

To see why, notice that substituting $\kappa^*(S;\gamma)$ from (44) into the bank's problem in (40) recovers the planner's problem.

The size of the optimal penalty is affected by the degree of rollover risk:

$$\frac{1}{\lambda}\frac{\partial \kappa^*}{\partial \gamma} = \frac{-1}{(\alpha^*(S))^2}\frac{\partial \alpha^*}{\partial \gamma} > 0.$$

This is consistent with Proposition 6 and 7, where we show that the degree of under-investment is increasing in γ (so the marginal effect of an increase in γ on the private marginal rate of substitution is smaller than on the planner's marginal rate of substitution).

A.13. **Proof of Proposition 12.** First we derive the expression in Equation (16) for the value of the debt claim and then derive the comparative static results.

The first step to derive $\mathcal{V}$ is to minimise the object in Equation (16) with respect to K. The first-order condition yields

$$\Big(1-\alpha^*(F)\Big)F+\theta\left\{\log\frac{K}{B}-\log\frac{1-K}{1-B}\right\}=0,$$

which implies that

$$\frac{K^*}{1-K^*}=\frac{B}{1-B}\exp\left\{-\frac{1}{\theta}\Big(1-\alpha^*(F)\Big)F\right\},$$

or

$$K^*=\frac{\frac{B}{1-B}\exp\left\{-\frac{1}{\theta}\Big(1-\alpha^*(F)\Big)F\right\}}{1+\frac{B}{1-B}\exp\left\{-\frac{1}{\theta}\Big(1-\alpha^*(F)\Big)F\right\}}=\frac{B}{Z(F)}\exp\left\{-\frac{1}{\theta}\Big(1-\alpha^*(F)\Big)F\right\},$$

where $Z(F)\equiv 1-B+B\exp\left\{-\frac{1}{\theta}\Big(1-\alpha^*(F)\Big)F\right\}$. Consequently, $1-K^*=\frac{1-B}{Z(F)}$. Next plugging K^* back into $\mathcal{V}(F)$, we obtain

$$\begin{aligned}
\mathcal{V}(F) &= \alpha^*(F)F+K^*(1-\alpha^*(F))F+\theta\left\{\frac{B}{Z(F)}\exp\left\{-\frac{1}{\theta}\Big(1-\alpha^*(F)\Big)F\right\}\log\left(\frac{\frac{B}{Z(F)}\exp\left\{-\frac{1}{\theta}\Big(1-\alpha^*(F)\Big)F\right\}}{B}\right)\right.\\
&+ \left.\frac{1-B}{Z(F)}\log\left(\frac{\frac{1-B}{Z(F)}}{1-B}\right)\right\}\\
&= \alpha^*(F)F+K^*(1-\alpha^*(F))F+\theta\left\{-\frac{1}{\theta}(1-\alpha^*(F))F\,\frac{B}{Z(F)}\exp\left\{-\frac{1}{\theta}\Big(1-\alpha^*(F)\Big)F\right\}\right.\\
&+ \left.\frac{B}{Z(F)}\exp\left\{-\frac{1}{\theta}\Big(1-\alpha^*(F)\Big)F\right\}\log\frac{1}{Z(F)}+\frac{1-B}{Z(F)}\log\frac{1}{Z(F)}\right\}\\
&= \alpha^*(F)F+K^*(1-\alpha^*(F))F-K^*(1-\alpha^*(F))F+\frac{\theta}{Z(F)}\left(1-B+B\exp\left\{-\frac{1}{\theta}\Big(1-\alpha^*(F)\Big)F\right\}\right)\log\frac{1}{Z(F)}\\
&= \alpha^*(F)F-\theta\log Z(F)
\end{aligned}$$

Next, for the comparative statics, note that

$$\frac{\partial\mathcal{V}}{\partial F}=\alpha^*(F)+F\frac{\partial\alpha^*}{\partial F}-\frac{\partial Z}{\partial F}\frac{\theta}{Z(F)},$$

where

$$\frac{\partial Z}{\partial F} = -\frac{B}{\theta}\left[1-\alpha^*(F)-F\frac{\partial \alpha^*}{\partial F}\right]\exp\left\{-\frac{1}{\theta}\Big(1-\alpha^*(F)\Big)F\right\}.$$

And so

$$\frac{\partial \mathcal{V}}{\partial F} = \left(\alpha^*(F)+F\frac{\partial \alpha^*}{\partial F}\right)\left(1-\frac{B}{\theta}\mathrm{e}^{-\frac{1}{\theta}\left(1-\alpha^*(F)\right)F}\right)+\frac{B}{\theta}\mathrm{e}^{-\frac{1}{\theta}\left(1-\alpha^*(F)\right)F}.$$

Thus, insofar that $\alpha^*(F)+F\frac{\partial \alpha^*}{\partial F}>0$, it follows that $\frac{\partial \mathcal{V}}{\partial F}>0$, which is guaranteed for $R>\underline{R}$.

And so for the comparative statics with respect to S, note that

$$\frac{\partial \mathcal{V}}{\partial S} = F\frac{\partial \alpha^*}{\partial S}-\frac{\partial Z}{\partial S}\frac{\theta}{Z(F)}<0,$$

since $\frac{\partial Z}{\partial S}=-\frac{B}{\theta}F\frac{\partial \alpha^*}{\partial S}\mathrm{e}^{-\frac{1}{\theta}\left(1-\alpha^*(F)\right)F}>0$. Thus, $\frac{\partial F^*}{\partial S}>0$.

Next, following a marginal increase in B,

$$\frac{\partial \mathcal{V}}{\partial B} = -\frac{\partial Z}{\partial B}\frac{\theta}{Z(F)} = \left(1-\mathrm{e}^{-\frac{1}{\theta}\left(1-\alpha^*(F)\right)F}\right)\frac{\theta}{Z(F)}>0,$$

implying that $\frac{\partial F^*}{\partial S}<0$.

Finally, following a marginal increase in θ, we have that

$$\frac{\partial \mathcal{V}}{\partial \theta} = -\log Z(F)-\frac{B\Big(1-\alpha^*(F)\Big)F\,\mathrm{e}^{-\frac{1}{\theta}\left(1-\alpha^*(F)\right)F}}{\theta\, Z(F)}.$$

Moreover, $\lim_{\theta\to\infty}\frac{\partial \mathcal{V}}{\partial \theta}=0$, $\lim_{\theta\to 0}\frac{\partial \mathcal{V}}{\partial \theta}=\infty>0$ and $\frac{\partial^2 \mathcal{V}}{\partial \theta^2}=-\frac{(1-B)B\left(1-\alpha^*(F)\right)^2F^2}{\theta^3 Z^2}<0$. In sum, this implies that $\frac{\partial \mathcal{V}}{\partial \theta}>0$ for all $\theta\in(0,\infty)$, and so $\frac{\partial F^*}{\partial \theta}<0$.

A.14. **Proof of Proposition 13.** For small but positive noise, the bank's expected profits are given by

$$\pi(S) = p(S)\left(R(1-S)-F\right)+\Big(1-p(S)\Big)\int_0^{\alpha^*(S)}\Big[(1-\delta\alpha)R(1-S)-F-(\rho-1)\ell(\alpha^*,x^*)F\Big]d\alpha, \quad (45)$$

where $\ell(\alpha^*, x^*) = \gamma$ and $\alpha^* = \alpha^{IN} - \frac{\rho-1}{\delta}\frac{\ell(\alpha^*,x^*)F}{R(1-S)}$. It thus follows that

$$\begin{aligned}\lim_{\epsilon\to 0}\frac{\partial \pi}{\partial S} &= \frac{\partial p}{\partial S}\left[R(1-S) - F - \int_0^{\alpha^*(S)}\Big[(1-\delta\alpha)R(1-S) - F\Big]d\alpha\right] \\ &- p(S)R - \Big(1-p(S)\Big)R\int_0^{\alpha^*(S)}(1-\delta\alpha)d\alpha\,,\end{aligned}$$

where $\alpha^*(S) \to \alpha^{IN} - \frac{\rho-1}{\delta}\frac{\gamma F}{R(1-S)}$. And so S^* is given by the solution to

$$\frac{[R(1-S^*) - F - \int_0^{\alpha^*(S^*)}\Big[(1-\delta\alpha)R(1-S^*) - F\Big]d\alpha}{p(S^*)R + \Big(1-p(S^*)\Big)R\int_0^{\alpha^*(S^*)}(1-\delta\alpha)d\alpha} = \frac{1}{\partial p/\partial S}\,. \tag{46}$$

Assuming that the second-order condition for S^* to be a maximum continues to hold, it follows that the derivative of the numerator on the left-hand side with respect to ρ is positive since $\frac{\partial \alpha^*}{\partial \rho} < 0$, while the derivative of the denominator on the left-hand side with respect to ρ is negative. And so, following a marginal increase in ρ, the marginal rate of substitution increases, and so S^* increases in equilibrium.

A.15. **Proof of Proposition 14 and derivation of the social planner's solution with multiple banks.** Modeled as a best-shot public good, the level of cybersecurity is given by $X(\vec{S}) = \max\{S_1, \ldots, S_N\}$. Suppose that bank 1 chooses $S_1^* = S^*(F)$, where $S^*(F)$ is the contribution to cybersecurity in the case $N = 1$. It immediately follows that no other bank, $j \neq 1$, would choose $S_j^* \leq S_1^*$ since this would have no material impact on the level of cybersecurity in the system. Moreover no bank, j, would individually choose $S_j^* > S^*(F)$ since, as we have previously shown, $S^*(F)$ is the optimum for the bank. Thus, all other banks invest nothing in cybersecurity, $S_2^* = \ldots, S_N^* = 0$.

Next, turning to the social planner's problem, the Lagrange equation is given by

$$\mathcal{L} = \sum_{b=1}^{N}\pi_b(I_b, X) - \lambda\left(1 - \sqrt{\frac{cX}{V}}\right)\sum_{b=1}^{N}\Big(1-\alpha_b^*(I_b)\Big) + \sum_{b=1}^{N}\phi_b(1-I_b-S_b) + \xi\left(\sum_{b=1}^{N}S_b - X\right), \tag{47}$$

where ϕ_b and ξ are Lagrange multipliers. For all $b = 1,\ldots,N$, the necessary and sufficient Kuhn-Tucker conditions are given by

$$(1)\quad \frac{\partial \mathcal{L}}{\partial I_b} = 0 : \frac{\partial \pi_b}{\partial I_b} + \lambda\left(1 - \sqrt{\frac{cX}{V}}\right)\frac{\partial \alpha_b^*}{\partial I_b} = \phi_b;$$

$$(2)\quad \frac{\partial \mathcal{L}}{\partial X} = 0 : \sum_{k=1}^{N}\left\{\frac{\partial \pi_k}{\partial X} + \frac{\lambda}{2}\sqrt{\frac{c}{VX}}\Big(1 - \alpha_k^*(I_k)\Big)\right\} = \xi;$$

$$(3)\quad \frac{\partial \mathcal{L}}{\partial S_b} = 0 : \phi_b = \xi\,.$$

From the above conditions, we obtain

$$\frac{\sum_{k=0}^{N}\left\{RI_k - F - \int_0^{\alpha_k^*(I_k)}\Big((1-\delta\alpha)RI_k - F\Big)d\alpha + \lambda\Big(1 - \alpha_k^*(I_k)\Big)\right\}}{\phi_b} = \frac{1}{\frac{1}{2}\sqrt{\frac{c}{VX}}},$$

for all $b = 1,\ldots,N$. Substituting for ϕ_b from condition (1), we have

$$\sum_{k=1}^{N}\frac{\left\{RI_k - F - \int_0^{\alpha_k^*(I_k)}\Big((1-\delta\alpha)RI_k - F\Big)d\alpha + \lambda\Big(1 - \alpha_k^*(I_k)\Big)\right\}}{R\sqrt{\frac{cX}{V}} + \left(1 - \sqrt{\frac{cX}{V}}\right)\left\{\int_0^{\alpha_b^*(I_b)}\frac{\partial EV_b}{\partial I_b}d\alpha + EV_b(\alpha_b^*)\frac{\partial \alpha_b^*}{\partial I_b}\mathbf{1}_{\gamma>\widehat{\gamma}} + \lambda\frac{\partial \alpha_b^*}{\partial I_b}\right\}} = \frac{1}{\partial X/\partial S_b}, \tag{48}$$

which is equivalent to

$$\frac{\partial \pi_b/\partial X + \lambda(1-\alpha_b^*)}{\partial \pi_b/\partial I_b + \lambda\partial\alpha_b^*/\partial I_b} + \sum_{k\neq b}^{N}\frac{\partial \pi_k/\partial p + \lambda(1-\alpha_k^*)}{\partial \pi_b/\partial I_b + \lambda\partial\alpha_b^*/\partial I_b} = \frac{2X}{p} \qquad \forall\ b = 1,\ldots,N\,. \tag{49}$$

REFERENCES

Acemoglu, D., A. Malekian, and A. Ozdaglar (2016). Network security and contagion. Journal of Economic Theory 116, 536–585.

Adelmann, F., I. Ergen, T. Gaidosch, N. Jenkinson, A. Morozova, N. Schwarz, and C. Wilson (2020). Cyber risk and financial stability: It's a small world after all. Staff Discussion Notes (007), International Monetary Fund, Washington, DC.

Ahnert, T., K. Anand, J. Chapman, and P. Gai (2019). Asset Encumbrance, Bank Funding and Fragility. Review of Financial Studies 32(6), 2422–2455.

Ardagna, C., S. Corbiaux, K. Van Impe, and R. Ostadal (2023). Enisa threat landscape 2023. The European Union Agency for Cybersecurity Report, October.

Association of Banks in Singapore (2018). Red team: Adverserial attack simulation exercises. Guidelines for Financial Industry in Singapore, Version 1.0.

Biener, C., M. Eling, and J. H. Wirfs (2015). Insurability of cyber risk: An empirical analysis. The Geneva Papers on Risk and Insurance-Issues and Practice 40(1), 131–158.

Bier, V., O. Santiago, and L. Samuelson (2007). Choosing what to protect: Strategic defensive allocation against an unknown attacker. Journal of Public Economic Theory 9, 563–587.

Brauchle, J.-P., M. Göbel, J. Seiler, and C. von Busekist (2020). Cyber mapping the financial system. Technical report, Carnegie Endowment for International Peace.

Brown, J. P. (1973). Toward an economic theory of liability. The Journal of Legal Studies 2(2), 323–349.

Cornes, R. (1993). Dyke maintenance and other stories: Some neglected types of public goods. The Quarterly Journal of Economics 108(1), 259–271.

Crisanto, J. C., J. U. Pelegrini, and J. Prenio (June 2023). Banks' cyber security – a second generation of regulatory approaches. FSI Insights on policy implementation No 50.

Crouzet, N., J. C. Eberly, A. L. Eisfeldt, and D. Papanikolaou (2022, August). The economics of intangible capital. Journal of Economic Perspectives 36(3), 29–52.

Dell'Ariccia, G., D. Kadyrzhanova, C. Minoiu, and L. Ratnovski (2021). Bank lending in the knowledge economy. The Review of Financial Studies 34(10), 5036–5076.

Deutsche Bundesbank (2022). Cybermapping mit ki – Deutsche Bundesbank und DFKI gründen gemeinsames Forschungslabor. https://www.bundesbank.de/de/presse/pressenotizen/cybermapping-mit-ki-deutsche-bundesbank-und-dfki-gruenden-gemeinsames-forschungslabor-889828.

Dixit, A. (1987). Strategic behavior in contests. American Economic Review 77(5), 891–898.

Doerr, S., L. Gambacorta, T. Leach, B. Legros, and D. Whyte (2022). Cyber risk in central banking. BIS Working Papers No 1039.

Duffie, D. and J. Younger (2019). Cyber runs. Hutchins Center Working Paper 51, Brookings Institution.

Dziubinski, M. and S. Goyal (2013). Network design and defence. Games and Economic Behavior 79, 30–43.

Eisenbach, T., A. Kovner, and M. J. Lee (2022). Cyber risk and the U.S. financial system: A pre-mortem analysis. Journal of Financial Economics, 145, 802–826.

Elestedt, L., U. Nilsson, and C.-J. Rosenvinge (2021). A cyber attack can affect financial stability. Economic Commentary No. 8, Sveriges Riksbank, Stockholm.

Englund, C. and C. Sosa (2022). An approach to quantifying operational resilience concepts. FEDS Notes, July 1, 2022.

European Systemic Risk Board (2022). Mitigating systemic cyber risk. Technical Report, Frankfurt.

Federal Financial Institutions Examination Council (2019). Business continuity management. Information Technology Examination Handbook.

Federal Reserve Board of Governors (2023). Cybersecurity and financial system resilience report. Report to Congress.

Fell, J., N. de Vette, S. Gardó, B. Klaus, and W. Wendelborn (2022). Towards a framework for assessing systemic cyber risk. Financial Stability Review, European Central Bank, Frankfurt.

Florackis, C., C. Louca, R. Michaely, and M. Weber (2020). Cybersecurity risk. NBER Working Paper 28196.

Frankel, D., S. Morris, and A. Pauzner (2003). Equilibrium selection in global games with strategic complementarities. Journal of Economic Theory 108(1), 1–44.

G7 (2018). G-7 fundamental elements for threat-led penetration testing. https://www.bundesbank.de/resource/blob/764690/792725ab3e779617a2fe28a03c303940/mL/2018-10-24-g-7-fundamental-elements-for-threat-led-penetration-testing-data.pdf.

Gatzert, N. and M. Schubert (2022). Cyber risk management in the us banking and insurance industry: A textual and empirical analysis of determinants and value. Journal of Risk and Insurance 89(3), 725–763.

Gilboa, I. and D. Schmeidler (1989). Maxmin expected utility with non-unique prior. Journal of Mathematical Economics 18(2), 141–153.

Goh, J., H. Kang, Z. X. Koh, J. W. Lim, C. W. Ng, G. Sher, and C. Yao (2020). Cyber risk surveillance: A case study of Singapore. IMF Working Paper 20/28.

Goldstein, I. and A. Pauzner (2005). Demand deposit contracts and the probability of bank runs. Journal of Finance 60(3), 1293–1327.

Gordon, L. and M. Loeb (2002). The economics of information security investment. ACM Transactions on Information and System Security 5(4), 438–457.

Goyal, S. and A. Vigner (2014). Attack, defence, and contagion in networks. The Review of Economic Studies 81, 1518–1542.

Grossklags, J., N. Christin, and J. Chuang (2008). Secure or insure? A game-theoretic analysis of information security games. In WWW '08: Proceedings of the 17th international conference on World Wide Web, pp. 209–218.

Hansen, L. and T. J. Sargent (2001, May). Robust control and model uncertainty. American Economic Review 91(2), 60–66.

Hirshleifer, J. (1983). From weakest-link to best-shot: The voluntary provision of public goods. Public Choice 41(3), 371–386.

HM Treasury (2022). Critical third parties to the finance sector: policy statement. https://www.gov.uk/government/publications/critical-third-parties-to

-the-finance-sector-policy-statement/critical-third-parties-to-the-finance-sector-policy-statement.

Jamilov, R., H. Rey, and A. Tahoun (2021). The anatomy of cyber risk. NBER Working Paper No. 28906.

Johnson, C., L. Badger, D. Waltermire, J. Snyder, and C. Skorupka (2016). Guide to cyber threat information sharing. Special publication 800-150, National Institute of Standards and Technology.

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

Kashyap, A. K. and A. Wetherilt (2019). Some principles for regulating cyber risk. AEA Papers and Proceedings 109, 482–87.

Kleijmeer, R., J. Prenio, and J. Yong (2019). Varying shades of red: How red team testing frameworks can enhance the cyber resilience of financial institutions. FSI Insights on policy implementation No 21.

Lev, B. and S. Radhakrishnan (2005). The valuation of organization capital. In Measuring capital in the new economy, pp. 73–110. University of Chicago Press.

Mauer, T. and A. Nelson (2020). International strategy to better protect the financial system against cyber threats. Technical report, Carnegie Endowment for International Peace.

European Central Bank (2024). Ecb to stress test banks' ability to recover from cyberattack. https://www.bankingsupervision.europa.eu/press/pr/date/2024/html/ssm.pr240103~a26e1930b0.en.html.

Mester, L. J. (2019). Cybersecurity and financial stability. Speech at the Federal Reserve Bank of Cleveland, Cleveland, Ohio. 21 November.

Microsoft (2012). memo from Bill Gates. https://news.microsoft.com/2012/01/11/memo-from-bill-gates/.

MIT Technology Review (2022). Wealthy cybercriminals are using zero-day hacks more than ever. https://www.technologyreview.com/2022/04/21/1050747/cybercrimi

nals-zero-day-hacks/.

Monetary Authority of Singapore (2021). Technology risk management guidelines. Guidelines for financial institutions.

Morris, S. and H. Shin (2003). Global games: Theory and applications. In M. Dewatripont, L. Hansen, and S. Turnovsky (Eds.), Advances in Economics and Econometrics (Proceedings of the 8th World Congress of the Econometric Society). Cambridge University Press.

Morris, S. and H. S. Shin (1998). Unique equilibrium in a model of self-fulfilling currency attacks. American Economic Review 88(3), 587–597.

Perlroth, N. (2021). This is how they tell me the world ends: The cyberweapons arms race. Bloomsbury Publishing.

Pretty, D. (2018). Reputation risk in the cyber age: The impact on shareholder value. Technical report, Aon and Pentland Analytics.

Rochet, J.-C. and X. Vives (2004). Coordination failures and the lender of last resort: was Bagehot right after all? Journal of the European Economic Association 2(6), 1116–47.

Rumsfeld, D. (2002). Defense Department Briefing – Secretary Donald Rumsfeld and General Richard Myers. https://www.c-span.org/video/?168646-1/defense-department-briefing.

Samuelson, P. (1954). The pure theory of public expenditure. The Review of Economics and Statistics 36(4), 387–389.

Shavell, S. (2009). Economic analysis of accident law. Harvard University Press.

Sophos (2023). The state of ransomware in financial services 2023. Sophos Whitepaper, July. Available: https://www.sophos.com/en-us/whitepaper/state-of-ransomware-in-financial-services.

S&P Global Market Intelligence (2019). S&P downgrades Malta-based Bank of Valletta. https://www.spglobal.com/marketintelligence/en/news-insights/trending/5mvfiykwlxliliri78qd-q2.

Strzalecki, T. (2011). Axiomatic foundations of multiplier preferences. Econometrica 79(1), 47–73.

Tarabay, J. (2021). How a dated cyber-attack brought a stock exchange to its knees. Bloomberg Businessweek.

The White House (2023). National cybersecurity strategy. https://www.whitehouse.gov/wp-content/uploads/2023/03/National-Cybersecurity-Strategy-2023.pdf.

Varian, H. (2004). System reliability and free riding. In L. J. Camp and S. Lewis (Eds.), Economics of information security, pp. 1–15. Springer.

Warren-Kachelein, D. (2022). JPMorgan Chase invests $12 billion in security updates. Information Security Media Group, January 19. Available: https://www.bankinfosecurity.com/jp-morgan-chase-invests-12-billion-in-security-updates-a-18329.

Woods, D. W., T. Moore, and A. C. Simpson (2021). The county fair cyber loss distribution: Drawing inferences from insurance prices. Digital Threats: Research and Practice 2(2), 1–21.