CYBERSECURITY AND FINANCIAL STABILITY

KARTIK ANAND, CHANELLE DULEY, AND PRASANNA GAI

ABSTRACT. Cyberattacks can trigger bank runs when vulnerabilities in IT systems are exploited. A trade-off between protection against cyberattacks and resilience in the event of attack shapes a bank's private incentives to invest in cybersecurity. Relative to the social optimum, there can be under- or over-investment in cybersecurity depending on the capability of the attacker and the nature of bank fragility. We clarify the circumstances where regulatory measures such as cyber-stress testing, red-teaming, and negligence rules are appropriate. Our results are robust to the introduction of liquidation costs, a lender of last resort, and a strictly positive recovery rate for investors.

Keywords: Cybersecurity, cyberattacks, bank runs, global games

JEL classifications: G01, G20, G21, G28.

We thank Toni Ahnert, Mei Dong, Thomas Eisenbach, Sayantan Ghosal, Tetsuya Hoshino, Charlie Kahn, Philipp J. König, Stephen Morris, Silvio Petriconi, Ryan Riordan, Lin Shen, Hyun Shin, Xavier Vives, Junyuan Zou 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, FIRS Conference, 2024, Berlin, Annual ECB Banking Supervision Research Conference, 2024, Frankfurt, and Banking in the Age of Challenges Conference, 2024, Paris 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.

"We have more engineers than Google or Amazon. Why? Because we have to. The fraudsters get smarter, savvier, quicker, more devious, more mischievous."

— Mary Erdoes, Chief Executive Officer of J.P. Morgan Asset & Wealth Management, 17 January 2024

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] And, according to the Bank of England's Systemic Risk Survey for 2023 H2, cyberattacks were perceived by financial market participants as the most significant source of risk to the financial system (Bank of England, 2023). Cyberattacks have the potential to trigger bank runs (Duffie and Younger, 2019) and disrupt wholesale funding markets (Eisenbach et al., 2022). The cyberattack on ICBC Financial Services in November 2023, which temporarily prevented it from settling transactions in the US Treasury market, is an important case study of the potential risk (Financial Times, 2024).[2] In light of these threats to financial stability, some policymakers advocate stress-tests to assess banks' preparedness 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, as well as how regulation addresses cyber risk, is limited. Our paper seeks to fill this gap.

[1]The European Union Agency for Cybersecurity (2023) reports a sharp increase of 22% in the volume of distributed-denial-of-service (DDoS) attacks on financial institutions during 2022. 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). The annual cost of cyberattacks — including theft, lost productivity and reputational harm — is estimated to reach $10.5tn globally by 2025 (Financial Times, 2024b). The IMF estimates annual financial firm losses from cyber attacks of up to the value of $2.2bn (IMF, 2024).

[2]The attacker was Lockbit, a deep-pocketed ransomware group, that gained access to ICBC's systems by exploiting a vulnerability in its remote desktop software, Citrix, that is commonly used by many banks. The attack impaired several operating systems, including those used by ICBC to clear US Treasury trades and repo transactions, and facilitated the theft of confidential data. Further disruptions were prevented after the parent company provided an emergency $9bn liquidity injection and paid an undisclosed ransom to Lockbit (The Banker, 2023).

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 generates future economic value by thwarting cyberattacks but 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 activities.[3] 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.

[3]JPMorgan Chase considers cybersecurity practices "run-the-bank priorities" and "change-the-bank" investments (Bank Info Security, 2022) to arm it against hacking attempts. The bank employs 62,000 technologists and allocates an annual budget of $15 billion, or one third of the value of its annual investments, to technological updates, focussing largely on cybersecurity protection (Financial Times, 2024a).

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 European Central Bank – being conducted during a period of heightened geo-political tensions with state-sponsored actors being accused of cybercrimes – is one such example.[4] The aim is to assess how supervised banks will respond to, 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 from risk-neutral wholesale depositors by issuing uninsured short-term 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; Tullock, 2008).

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 by professional fund managers to whom rollover decisions are delegated (Rochet and Vives, 2004), precipitating a bank run. And, even after the resumption of services, the bank may suffer permanent losses that further undermine its ability to service deposits. Bank failure may, thus, be illiquidity- or insolvency-based. We pin

[4]Business continuity plans and cyber hygiene notices also fall into this category.

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 improve. 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 attacker is highly sophisticated. The more enduring is the disruption of a successful cyberattack, 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 of credit to the real economy. 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. In the final section of the paper and in an online Appendix, we extend the model to allow for (a) asset illiquidity and lender-of-last-resort support by the authorities; (b) a positive recovery rate for investors; (c) observable cybersecurity investments; (d) ambiguity-averse investors facing Knightian uncertainty about "zero-day" vulnerabilities – security flaws previously not known to exist; (e) multiple banks and common IT solutions; and (f) 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.[5] Eisenbach et al. (2023) find evidence of heightened cyber risk during stressed financial conditions, suggesting an important interdependence between cyber risk and the financial cycle. In earlier work, Eisenbach et al. (2022) examine how cyberattacks impair a bank's

[5] Jamilov et al. (2021) is a related contribution to 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 Aldasoro et al. (2022), Kamiya et al. (2021), Woods et al. (2021) and Florackis et al. (2020).

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 are 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.

There are also points of contact with the economics of security.[6] 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.

[6]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 highlights the financial stability and cybersecurity aspects arising from the strategic interplay between an attacker and a defender.

Grossklags et al. (2008) examine how the option to invest in insurance against damages following a cyberattack influences decisions to contribute to cybersecurity. Ahnert et al. (2024) propose a principal-agent model of cybersecurity investment by firms that are delegated the task of protection by clients. They find that both the attacker's modality and firms' security investment decisions depend crucially on whether or not cybersecurity investment is observable. Our contribution to this literature shows how ex post coordination failures in the form of bank runs and the cyber-threat landscape shape banks' incentives to contribute to cybersecurity protection.

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 under perfect competition, a unit mass of investors, and an "attacker" intent on causing harm to the bank. All agents are risk-neutral. Investors are indifferent between consumption at $t = 1$ and $t = 2$, whereas 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 good.[7] 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 $t = 2$.[8] To finance this investment, the bank issues $D > 0$

[7] 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.

[8] 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 5.1.

of uninsurable short-term debt claims, such as certificates of deposit, to investors. We consider depositors to be wholesale investors rather than households, as in Rochet and Vives (2004).[9] The face value of debt, $F > 0$, 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*. Investors depositing with the bank do not observe the bank's cybersecurity investment.[10]

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. The bank is solely responsible for the security of these systems, i.e., finding vulnerabilities and mitigating (or patching) them. 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; Tullock, 2008).

As we discuss below, investing in cybersecurity improves the bank's chances of thwarting cyberattacks, which has economic benefits for the bank.[11] Moreover, these investments can readily be accounted for on the bank's balance 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. At $t = 0$, the attacker exerts a level of effort, $A \geq 0$, at marginal cost $c > 0$, directed towards discovering and exploiting vulnerabilities in the bank's IT systems. A successful attack provides the attacker with a prize $V > c$ at $t = 2$. The marginal cost of attack is inversely related to the attacker's sophistication — the higher is c, the less sophisticated the attacker.

[9] The inclusion of stable sources of funding, such as long-term debt or equity core, would not qualitatively alter the core trade-off generated by the model. While replacing some of short-term debt with stable funding would reduce the incidence of runs in the event of a cyberattack, the bank can still fail if the losses incurred are too large. Thus, the protection-resilience trade-off would persist but be modulated through the introduction of stable funding.

[10] For estimates of the degree of uncertainty that investors hold about banks' cyber risk exposure and the extent to which such uncertainty is priced in options markets, see Jamilov et al. (2021).

[11] 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.

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.

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.[12] 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 $t = 2$.

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

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

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.[13] Depositors are assumed to face a zero recovery rate upon bank failure. We relax this assumption in Section 5.2 and show that our results remain qualitatively unchanged.

[12] 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 (Bloomberg, 2021).

[13] 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.

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.[14] 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 1 summarises the bank's balance sheet at $t = 2$ following a successful attack and rollover of bank deposits. In Table 2 we list the key variables and parameters of interest used throughout the paper.

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

TABLE 1. 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

[14]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).

	Variable/Parameter	Interpretation
Endogenous	A	attacker effort
	D	short-term debt issued
	F	face value of debt
	I	investment by bank
	S	allocation towards protection
Exogenous	c	marginal cost of attack
	R	return on project
	r	risk-free technology
	V	attacker prize
	x_i	fund manager i's private signal
	α	impairment shock
	γ	conservatism ratio
	δ	deadweight loss
	ζ	LLR penalty rate
	κ	negligence rule penalty
	λ	social planner weight on bank failure
	ρ	impact of red teaming
	σ	cybersecurity subsidy
	ψ	fire-sale discount

TABLE 2. Exogenous and endogenous variables.

is $w - o > 0$.[15] 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

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

[15]Most wholesale deposits are held by collective investment schemes, whose managers tend to be remunerated on reputation (Chevalier and Ellison, 1997; Rajan, 2006).

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.

$t = 0$	$t = 1$	$t = 2$
1. Bank chooses the amount of debt, D, to issue, investment, I, and cybersecurity, S	1. Attacker succeeds with probability $1 - p(A, S)$ 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.

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

(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 $t = 2$.

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.

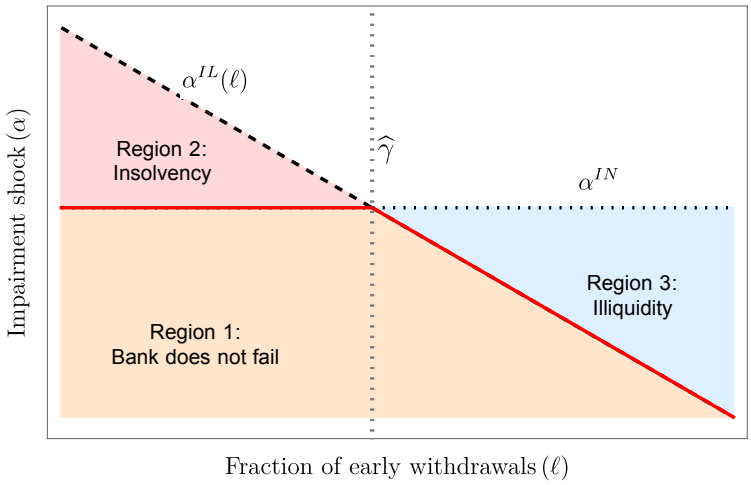


FIGURE 1. Failure conditions for a bank following a successful cyberattack.

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.[16]

Proposition 1. *Assume $\delta > 1 - \frac{r}{R}$ and suppose that $RI > FD$. 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 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 $\alpha > \alpha^{IL}(\gamma)$.

[16] 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).

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.

Proposition 1 relies on two conditions that ensure well defined limit dominance regions — a property required for uniqueness of the subgame equilibrium (Morris and Shin, 2003). The assumption $\delta > 1 - \frac{r}{R}$ ensures that the upper dominance bound is well defined. The upper dominance threshold implies that, irrespective of the bank's choices and 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 it cannot service its debts at $t = 2$ and fails due to insolvency. In the upper dominance region $(\alpha^{IN}, 1]$, therefore, fund managers have a dominant strategy to withdraw their funds, irrespective of the actions of other fund managers.[17] For the lower dominance bound to be well defined, we make a supposition involving ex ante parameters, $RI > FD$, which – as we show below in Lemma 4 – is always satisfied in equilibrium. In the lower dominance region, $[0, \alpha^{IL}(\gamma)]$, fund managers have a dominant strategy to roll over, irrespective of the actions of other fund managers.

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

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

[17] Note that when $\ell = 0$, $\alpha^{IL}(0) = 1$ and so whenever $\alpha^{IN} < 1$, the bank fails due to insolvency for all impairment shocks in the range $(\alpha^{IN}, 1]$.

Lemma 2. *The attacker expends effort*

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

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); Tullock (2008) and suppose that the bank pre-commits to its 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

$$\begin{aligned} \max_{\{D,I,S\}} \quad & \pi(D, I, S) \\ \text{subject to} \quad & D = I + S \\ & D \leq 1 \\ & \pi(D, I, S) \geq 0. \end{aligned}$$

The constraints reflect: (i) the bank's initial balance sheet, (ii) the resources available for the amount of debt that can be issued, and (iii) the participation constraint for the bank to engage in financial intermediation.

Lemma 3. *For any given I and S, the bank issues $D^* = 1$ of uninsured short-term 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.

Lemma 4. *It is never optimal for the bank to choose $S > \overline{S} \equiv 1 - \frac{F}{R}$ and so, in equilibrium, $RI \geq FD$.*

Lemma 4 follows by virtue of the bank's participation. The bank's profit in the event that no breach takes place is $RI - FD = R(1-S) - F$ by Lemma 3. An allocation to cybersecurity greater than $\overline{S}$ causes it to have negative equity value and default even in the event that no breach takes place since it has generated zero asset returns. And, whenever the attacker is successful, the bank always defaults. Thus, the bank's participation constraint is violated for $S > \overline{S}$. Lemma 4 thus ensures that $RI \geq FD$, which implies that in the $t = 1$ subgame, both dominance bounds are well-defined.

Proposition 2. *There exists a lower bound on returns, $\underline{R}$ such that 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*

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

such that the threshold $\widehat{\gamma} \equiv \widehat{\gamma}(S^*)$ *is unique and well defined. 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 do not directly observe bank investment in cybersecurity; instead, they respond to a schedule of protection allocations in setting the price of debt faced by the bank. 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 5. *The face value of debt, $F^*(S)$, is U-shaped in the bank's contribution to cybersecurity, S. The minimum of the U-shape is at $S = \tilde{S}$, which is larger when bank failure is illiquidity-driven.*

The ex ante behaviour of investors is also shaped by the protection-resilience trade-off. The value of a debt claim increases as the bank contributes more to cybersecurity when the benefit from protection is relatively large, i.e., if $S < \tilde{S}$. In these circumstances, investors prefer that the bank takes measures to forestall cyberattacks. So the face value of debt is decreasing in the bank's contribution to cybersecurity, $\frac{\partial F^*}{\partial S} < 0$. But for $S > \tilde{S}$, when the erosion of resilience is of greater concern, investors demand greater compensation per unit contributed to cybersecurity and so $\frac{\partial F^*}{\partial S} > 0$. The critical threshold, $\tilde{S}$, at which the switch occurs depends on the extent of rollover risk, γ. In particular, for $\gamma > \widehat{\gamma}$, when bank failure is driven by illiquidity risk, investors place

greater emphasis on protection over resilience and so the face value of debt is decreasing for a larger range of S.

Laissez-faire equilibrium. The joint equilibrium for the bank's contribution to cybersecurity and the face value of debt is summarised in Proposition 3 and depicted in Figures 2a and 2b.

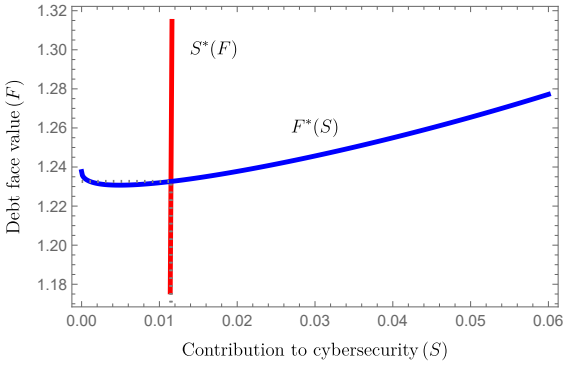
Proposition 3. *There exists a unique equilibrium, which is jointly characterised by* $\mathcal{V}(F^{**}, S^{**}) = r$, *where the bank's contribution to cybersecurity is given by* $S^{**} \equiv S^{*}(F^{**})$.

Figure 2a depicts the unique 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.

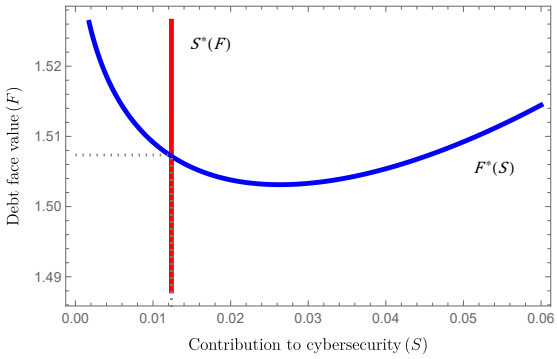
The equilibrium when bank failure is illiquidity-driven is shown in Figure 2b where the best-response correspondence for the face value of debt is downward sloping. When investors expect a higher contribution to cybersecurity by the bank, they require a lower face value of debt to lend. Expecting a low 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 4. *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* $c > \tilde{c}^{**}$.

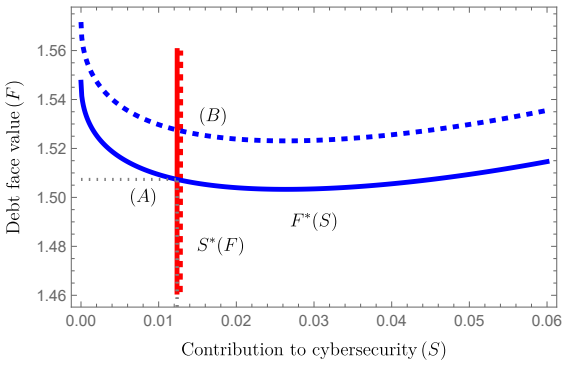


(a) $\gamma = 0.1$ and $\widehat{\gamma}(S^{**}, F^{**}) = 0.55$.

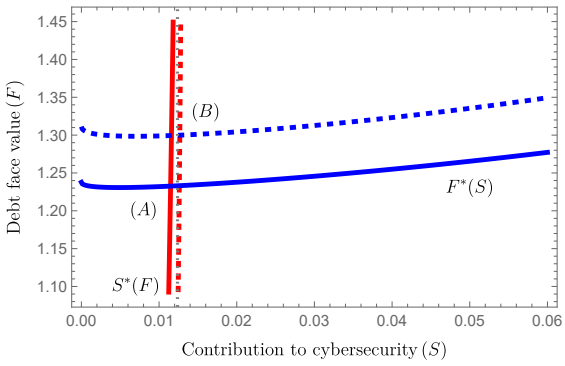


(b) $\gamma = 0.8$ and $\widehat{\gamma}(S^{**}, F^{**}) = 0.68$.

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



(a) Rollover risk increases from $\gamma = 0.8$ to $\gamma = 0.81$.



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

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

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 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 deadweight 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 $\lambda > 0$.

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

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

where $\Big(1 - p(S)\Big)\Big(1 - \alpha^*(S)\Big)$ is the ex ante probability that the bank fails due to a cyberattack. Although the cybersecurity contribution schedules of the planner and bank differ, the participation constraint of investors remains the same.

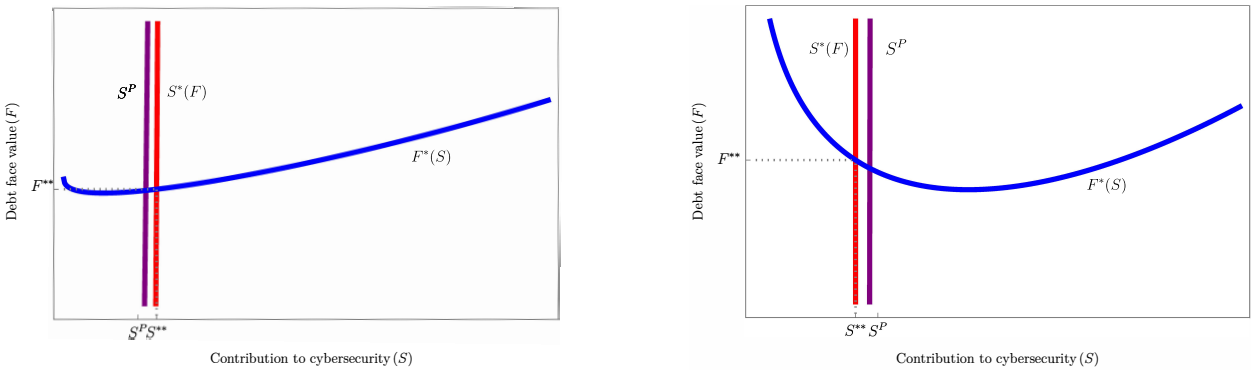
Proposition 5. *There exists a critical threshold for the attacker's sophistication, $\widehat{c}$, such that the bank over-invests in cybersecurity, $S^{**} > S^P$, if and only if $\gamma \leq \widehat{\gamma}$, i.e., its failure is driven by insolvency concerns, and $c < \widehat{c}$. Else, the bank 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. Up to 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 bank winning the contest is low, $p(S)$ is small, which scales up the social resilience effect.



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

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

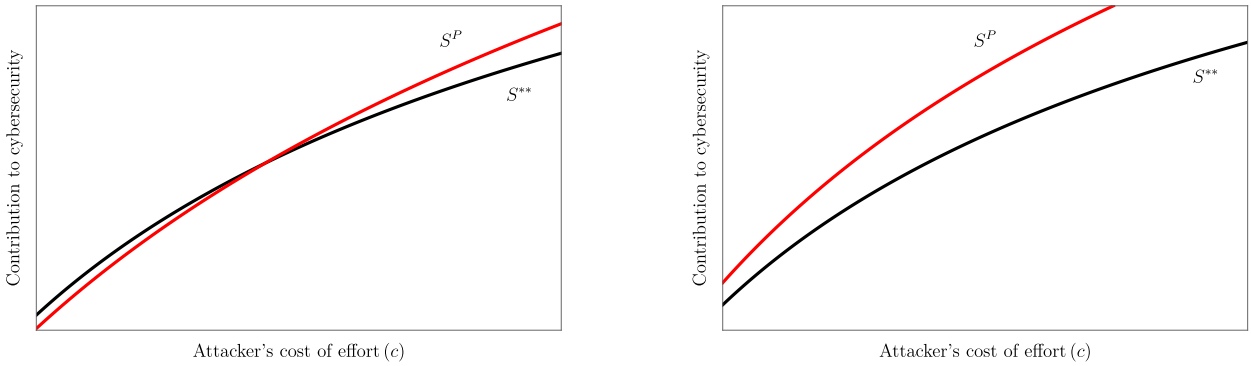
FIGURE 4. Joint equilibrium for the bank's contribution to cybersecurity as well as the social planner's allocation and the face value of debt. In producing the figure, we assume $R = 3.5$, $r = 1$, $\delta = 0.8$, $c = 0.1$, $V = 1$ and $\lambda = 1$.

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 resilience 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 6. *The difference, $|S^P - S^{**}|$ is increasing in c, i.e., as the attacker becomes less sophisticated, if either: (i) $\gamma > \widehat{\gamma}(S^{**}, F^{**})$, or (ii) $\gamma < \widehat{\gamma}(S^{**}, F^{**})$ and $c > \widehat{c}$. And when bank failure is illiquidity-driven, $|S^P - S^{**}|$ is increasing in rollover risk, γ.*

Proposition 5 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



(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 $R = 3.5$, $r = 1$, $\delta = 0.8$, $V = 1$ and$\lambda = 1$, and vary c from 0.5 to 2.5.

planner. Proposition 6 shows, moreover, that under the same conditions, 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 protection and allocates more to cybersecurity. Figure 5 illustrates. And when bank failure is illiquidity driven, the failure threshold is reduced following a marginal increase in rollover risk, which further strengthens the social protection effect.

4.1. **Operational resilience standards.** Our analysis highlights that the correct regulatory prescription crucially depends on the characteristics 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 7. *When the bank over-invests in protection, 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 5 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.[18]

Such constraints can be implemented in practice. Cyber stress-tests, cyber hygiene notices and business continuity planning are examples.[19] 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.[20] 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).[21]

4.2. **Red-team testing.** When bank failure from a cyberattack is driven by either liquidity risk or insolvency risk and the attacker sophistication is low, our analysis suggests that the planner's constrained efficient solution can be achieved by improving the efficacy of the bank's contribution to cybersecurity. By contributing S, the level of cybersecurity can be augmented 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.

[18] Brando et al. (2022) call for special attention to be paid to banks' cyber resilience beyond traditional mitigants, such as capital and liquidity, owing to the fact that "individual firms (including those in the supply chain) will underinvest in their own resilience, failing to internalize their effect on the system's stability and thus putting the system at more risk than is optimal".

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

[20] 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).

[21] 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).

Proposition 8. *When the bank under-invests in protection, 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.[22] These tests may be funded by the banks themselves, by authorities, or even crowd-sourced through initiatives that promote information sharing and vulnerability reporting.[23].

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 be the dollar value of the subsidy, and τ the lump sum tax used to finance the subsidy. Proposition 9 demonstrates that Pigouvian subsidies elicit higher investment in cybersecurity to bridge the gap between the level of protection offered by the bank and planner.

[22]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).

[23]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. See also Kleijmeer et al. (2019)

Proposition 9. *When the bank under-invests in protection, 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 10. *When the bank under-invests in protection, 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 4 summarises the use case for each tool as well as considerations for their interactions with other initiatives.

5. Robustness and extensions

Our results are robust to the inclusion of liquidation costs and a lender of last resort, as well as a strictly positive recovery rate for investors. We also describe below how (i) the protection-resilience trade-off is influenced by Knightian uncertainty about the cyber risk landscape; (ii) common IT systems shared by multiple banks introduce interdependencies in cybersecurity; (iii) technology vendors influence cyber risk management; and (iv) self-fulfilling outcomes arise when banks can learn about the extent of disruption caused by a cyber breach. A formal treatment of these extensions is provided in the online Appendix.

5.1. **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.

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 4. Proposed regulatory toolkit for addressing cyber threat landscape.

Our analysis of 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.

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

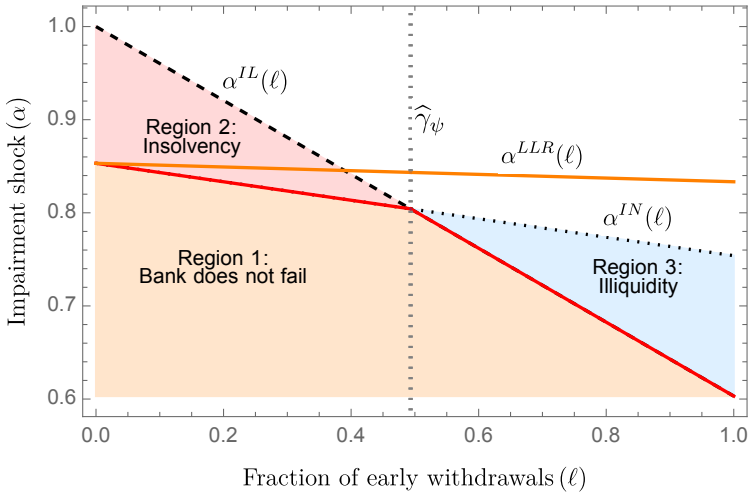


FIGURE 6. Failure conditions with asset illiquidity, $\psi < 1$, for the bank following a successful cyberattack. We also depict the failure threshold with the LLR.

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 improves 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 11. *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.

5.2. **Positive recovery rate.** Our analysis has so far assumed an extreme zero recovery rate for investors in the event that the bank fails. Under modest restrictions, our results follow in the absence of bankruptcy costs and when investors obtain pro rata shares of the bank's returns when the bank fails.

If the impairment shock is large enough to precipitate a run on the bank, $\alpha > \alpha^{IL}(\gamma)$, but not so large that it fails due to insolvency, $\alpha < \alpha^{IN}$, the bank fails due to illiquidity even though it could have paid depositors. So in the absence of bankruptcy costs, for $\alpha \in [\alpha^{IL}(\gamma), \alpha^{IN}]$, depositors continue to be repaid in full. For $\alpha \in [\alpha^{IN}, 1]$, a portion of the bank's recovered project returns can be shared across depositors. Accounting for this possibility, the value of the debt claim is thus

$$\mathcal{V}(F,S) = p(S)F + \Big(1 - p(S)\Big)\Big[\alpha^{IN}(S,F)\,F + \Omega(F,S)\Big], \tag{15}$$

where

$$\Omega(F,S) = \int_{\alpha^{IN}(F,S)}^{1} (1 - \alpha\delta)R(1-S)\,d\alpha\,, \tag{16}$$

is the expected return to depositors when they each receive pro rata shares of the bank's assets. Since $\frac{\partial \Omega}{\partial F} = \frac{F}{\delta R(1-S)} > 0$, it remains that the value of the debt claim is increasing in F. Moreover,

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

where $\Big(1 - \alpha^{IN}(F,S)\Big)F - \Omega(F,S) = \frac{((1-\delta)R(1-S)-F)^2}{2\delta R(1-S)} > 0$ and $\frac{\partial \alpha^{IN}}{\partial S}F + \frac{\partial \Omega}{\partial S} = \frac{(1-\delta)^2R^2(1-S)^2 - F^2}{2\delta R(1-S)} < 0$.

In the limit $S \to 0$, we continue to have that $\frac{\partial F^*}{\partial S} < 0$ implying that the protection effect is dominant for investors when the bank's contribution to cybersecurity is low. And so following a marginal increase in S, the face value of debt is decreasing. But when S is sufficiently large, investors value greater resilience over protection and so $\frac{\partial F^*}{\partial S} > 0$. Compared with our baseline analysis with a zero recovery rate, we note that the presence of $\Omega(F,S)$ reduces the weight placed

on the protection effect (term that multiplies into $\frac{\partial p}{\partial S}$) while it amplifies the resilience effect (term that multiplies into $1-p(S)$). Consequently the point at which the switch in the sign of $\frac{\partial F^*}{\partial S}$ occurs is shifted towards the left, i.e., for a lower level of S.[24] In sum, the core tension between protection and resilience remains intact when we relax the zero-recovery rate assumption for investors.

5.3. **Observable cybersecurity investments.** In deriving the schedule $S^*(F)$, we have assumed that the bank takes the face value of debt as given. In other words, the bank does not internalize how its investment choice impacts the face value of debt. In this section, we extent our analysis to consider what happens when the bank does internalise this impact. Formally, the bank's ex ante problem is given by

$$\max_S \pi(S) \equiv p(S)\Big(R(1-S)-F^*(S)\Big)+\Big(1-p(S)\Big)\int_0^{\alpha^*(S)}\Big[(1-\alpha\delta)R(1-S)-F^*(S)\Big]d\alpha\,, \quad (17)$$

where $F^*(S)$ solves $\Big[p(S)+(1-p(S))\alpha^*(S)\Big]F=r$. Focusing on the case where $\gamma<\widehat{\gamma}$, the total derivative of π with respect to S is given by

$$\begin{aligned}\frac{d\pi}{dS} &= \frac{\partial p}{\partial S}\left[R(1-S)-F^*(S)-\int_0^{\alpha^*(S)}\Big[(1-\alpha\delta)R(1-S)-F^*(S)\Big]d\alpha\right]\\ &- R\left[p(S)+\Big(1-p(S)\Big)\int_0^{\alpha^*(S)}(1-\alpha\delta)d\alpha\right]-\frac{r}{F^*(S)}\frac{dF^*}{dS}\,.\end{aligned}$$

Evaluating the above equation at the equilibrium S^{**} that was previously derived, we obtain

$$\left.\frac{d\pi}{dS}\right|_{S=S^{**}} = -\frac{r}{F^*(S^{**})}\left.\frac{dF^*}{dS}\right|_{S=S^{**}}\,.$$

When $\frac{dF^*}{dS^*}<0$ in equilibrium, the bank invests more in protection relative to the equilibrium S^{**} that was previously derived. Insofar that $\frac{dF^*}{dS}>0$, the protection-resilience trade-off is tilted towards increasing resilience and so the bank's contribution to cybersecurity is reduced.

[24] The level of cybersecurity investment at which $\frac{\partial \mathcal{V}}{\partial S}=0$ occurs at $\tilde{S}\equiv S: \frac{p(S)}{1-p(S)}-\frac{2S}{1-S}\left[\frac{2\delta R(1-S)\Omega(F,S)}{[(1-\delta)R(1-S)-F]^2}\right]=0$.

Turning to the normative implications, following a marginal increase in S, the wedge between the bank's and planner's choices is

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

where

$$\frac{d\alpha^*}{dS} = -\left(\frac{1}{\delta}-\alpha^*\right)\left(\frac{1}{1-S}+\frac{1}{F^*(S)}\frac{dF^*}{dS}\right) < 0.$$

Relative to our benchmark and following the lines of reasoning behind the result on Proposition 5, it follows that in the limit $c \to 0$, the result that there is over-investment in cybersecurity is strengthened. While in the limit $c \to V$, the incentives to under-invest in cybersecurity are weakened. On the other hand when $\gamma > \widehat{\gamma}$ such that $dF^*/dS < 0$, the the bank always under-invests in cybersecurity as in our benchmark analysis. Thus, our results are qualitatively robust to the assumption on whether or not the bank internalises the effects of its cybersecurity choice on the face value of debt.

5.4. **Extensions and further implications.** Our main analysis is centred around a simple benchmark to highlight the trade-off between protection and resilience, and how the shadow of coordination failure from a self-fulfilling bank run shapes bank investment incentives. But the core tension is perturbed in various ways when we allow for different informational structures, and types of banks and IT service providers, each generating empirical implications. Here, we summarise additional extensions and results contained in the Online Appendix.

Knightian uncertainty. For a given face value, F, the value of a debt claim, $\mathcal{V}(F,S)$, 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 (model) uncertainty. This is especially true for zero-day vulnerabilities, which are security flaws that were previously unknown to exist (Perlroth, 2021). In online Appendix 1, we reformulate the model in the spirit of Hansen and Sargent (2001) to allow for model misspecification and capture the genuine uncertainty associated with the possibility of the bank experiencing a breach. Investors hold a reference belief over the likelihood that the bank

is able to protect itself from a breach, and assign a measure to the value derived from contracting with the bank, accounting for aversion to model misspecification (Strzalecki, 2011). The trade-off between protection and ex-post resilience is thus shaped by a measure of the degree of model uncertainty.

Increases in the reference belief that the bank is successful at protecting itself against attack, and in the degree of trust in the reference model, cause a reduction in the face value of debt and the bank's investment in protection. As beliefs strengthen, investors become more optimistic about the bank's ability to stave off an attack. Similarly, as trust in the reference model improves, investors are less uncertainand place greater emphasis on their reference model.[25] Both effects increase the expected value of depositing with the bank, leading investors to demand a lower level of compensation. By Proposition 3, the bank is relatively less exposed to failure at any level of disruption in the event of a breach, through the indirect effect of a decrease in face value, leading it to allocate less towards cybersecurity protection as a result. These results generate testable implications on the effect of investor sentiment about cyber risk on the pricing of debt. They suggest that public dissemination of the ways in which the bank's efforts to protect itself map to a reduction in cyber risk may reduce its debt obligations and also lower its incentives to invest in protection ex ante.

Multiple banks. The cyber threat landscape is also shaped by interdependencies among banks. In addition to financial linkages, many banks use common IT systems provided by third-party vendors. We extend our baseline model in online Appendix 2 to include a number of heterogeneous banks operating on common IT software or hardware. In such a context, cybersecurity has the hallmarks of a *public good* (Mester, 2019). Variation in the types of shared risks faced by banks can be reflected in differences in the nature of this public good. For example, the public disclosure of software vulnerabilities to the benefit of other banks using similar systems is a best-shot public good — the provision of protection is proportional to the bank that contributes the most

[25]Using a Kullback-Leibler measure of divergence of the true probability of the bank's winning the contest from a reference belief, we provide limiting cases on the degree of trust in the reference model in Appendix B.1. When there is complete trust in the reference model, the value of a debt claim collapses into a standard expected utility criterion. On the other extreme where model uncertainty becomes paramount, the investor has no confidence in their reference model and values debt according to a maxmin criterion.

to uncovering bugs, and cybersecurity investments are substitutes across banks.[26] By contrast, weakest-link public goods refer to situations whereby aggregate protection is proportional to the bank that contributes the least, and contributions are perfect complements. Examples include improper configuration, updating or maintenance of systems by banks that allow attackers to perform virtual-machine or container breakout attacks, in which one client's access is used as a "backdoor" through which to gain privileged access to other clients' systems hosted on the same physical server.[27]

We parameterise the degree to which bank contributions are substitutes using a generalised public good aggregation function (Cornes, 1993) which has implications for the allocation towards cybersecurity protection by each bank. The public goods formulation of cybersecurity introduces externalities that aggravate or mitigate the underprovision we derive in Proposition 5. In particular, the extent of under- or over-provision by banks is systematically related to the degree of heterogeneity across banks and the type of public good aggregator. As cybersecurity investments become complements and a weaker- or weakest-link technology aggregates contributions, those banks with the lowest marginal rate of substitution between protection and resilience disproportionately affect the global provision of the public good, whereas a best-shot aggregator implies those with the highest marginal rate of substitution are instrumental in protection. Relative to the single bank case, free-riding incentives can lead banks to place a greater weight on private resilience vis-à-vis ex ante protection. Our analysis suggests that policies designed to address the externalities arising from underprovision of protection require tracing out the characteristics of common IT infrastructure across banks. Moreover, the characteristics of common IT infrastructure — that is, whether contributions are complements or substitutes, determine whether targeted efforts (for example, subsidies restricted to systemically important banks) should be directed at weaker or stronger links

[26] For example, bug bounty platform HackerOne has paid out a total of over $300 million in bug bounty prizes to white-hat hackers since its inception (HackerOne, 2023). It subsequently publishes vulnerabilities for public consumption.

[27] In February 2024, the CVE List, a vulnerability reporting site sponsored by the U.S. Department of Homeland Security and Cybersecurity and Infrastructure Agency, published a set of vulnerabilities that give attackers a way to break out of containers, a type of virtualisation of operating systems, and execute malicious actions on underlying host systems (CVE, 2024). Other examples include the 'Meltdown' and 'Spectre' hardware vulnerabilities present in every Intel processor since 1995, but only uncovered in 2018 (Lipp et al., 2018).

in the system. Efforts to uncover common IT infrastructure are underway in several jurisdictions (Brauchle et al., 2020; Deutsche Bundesbank, 2022).

IT vendors. We consider the role of IT vendors in the contest in online Appendix 3. Whereas the additional banks we model in online Appendix 2 are also driven by the risk of bank runs to invest in cybersecurity, the vendor simply cares about a loss of reputation in the event of an attack. Insofar as vulnerability reporting is a best-shot public good, if the vendor's concern for reputation is sufficiently high relative to the bank's marginal rate of substitution between protection and resilience, then the vendor contributes to vulnerability detection, and banks free ride on the vendor's allocation.

In practice, IT vendors and clients sign service level agreements centred around a shared-responsibility model (Harmon, 2020). But like the multiple bank case, the vendor may still fail to internalise how its contribution to cybersecurity influences the contributions of other banks and the social cost of bank failure. Owing to its role as service provider, the vendor has additional influence over the cyber landscape in its ability to promote sound cybersecurity practices and inform banks about their responsibility in maintaining and configuring their cloud infrastructure. The identification of critical vendors, and imposing minimum cybersecurity standards on them (HM Treasury, 2022), are first steps in addressing underprovision of cybersecurity arising from such interdependencies.

Learning about vulnerabilities. We also analyse how the nature of the contest, and the trade-off between protection and resilience, change when the attacker and banks can estimate the extent of disruption to operations induced by a successful breach. The unique equilibrium we derive in the contest between the bank and attacker gives way to multiplicity when the attacker knows the level of the impairment shock, α, following a successful attack, allowing the bank to estimate the scale of disruption. In online Appendix 4, we reformulate the attacker's problem so that the probability the bank is successful at detecting a vulnerability is a function of the disruption caused by the attacker in the event of a cyberattack.

A tripartite classification for the extent of the disruption, α, occurs when the attacker knows the degree of disruption its attack will induce. In particular, when *high-grade* attacks occur, i.e., the disruption is large that the bank fails, $\alpha > \widehat{\alpha}^{**}$, where $\widehat{\alpha}^{**}$ is the failure threshold that emerges in an environment with learning, the bank is incentivised to protect itself by allocating more resources towards cybersecurity. This, in turn, implies that if the attacker is successful at exploiting the vulnerabilities, the bank is less resilient, which validates its initial belief that attacks are high grade. By a similar logic, for $\alpha < \widehat{\alpha}^{*}$, where $\widehat{\alpha}^{*}$ is the corresponding bank failure threshold, only *low-grade* cyberattacks occur in equilibrium. By believing that it will remain resilient against a cyberattack, the bank invests more in assets, which in turn, bolsters its resilience. For $\alpha \in [\widehat{\alpha}^{*}, \widehat{\alpha}^{**}]$ there are multiple equilibria. Specifically, a vulnerability that is characterised by the disruption α can manifest as either a high-grade cyberattack or a low-grade one. The equilibrium that is selected depends on the bank's initially held belief about the severity of the disruption. Moreover, shocks to this belief can shift the equilibrium within the region of multiplicity.

The information ascertained by a bank about the disruption serves as a focal point to anchor its beliefs. Armed with the knowledge that a successful cyber attack would cause a certain level of disruption, the bank can better assess the value of greater protection. For $\alpha < \widehat{\alpha}^{*}$ and $\alpha > \widehat{\alpha}^{**}$, the bank's higher-order beliefs are also well anchored, leading to the unique equilibrium outcomes. While in the intermediate range, $\alpha \in [\widehat{\alpha}^{*}, \widehat{\alpha}^{**}]$, multiple equilibria are supported. Multiple equilibria may be eliminated if the bank's belief could be shaped. This can be achieved, for example, through institutions that both coordinate the cybersecurity activities of banks and also disseminate information on risks. For example, in the United Kingdom, the National Cyber Security Centre supported the creation of the Financial Sector Cyber Collaboration Centre (FSCCC) with the aim of identifying and coordinating responses to incidents that can impact the financial sector. The FSCCC works with both financial authorities and financial firms to combine and distribute information it gathers from across the sector. Other countries have similar institutional arrangements.[28]

[28] In the United States, this role is facilitated by the Cybersecurity and Information Security Agency. Likewise, both Australia and New Zealand also have national cybersecurity agencies.

Broader notion of resilience. We model a specific trade-off between a bank's ability to withstand withdrawals in the event of a successful breach and its capability in finding and patching vulnerabilities to ward off a cyber event entirely. But typical cyber risk management practices encompass all stages of an attack, and, in practice, resilience extends beyond the kind of balance sheet fragility we are concerned with. For example, the U.S. National Institute of Standards and Technology Cybersecurity Framework (NIST, 2024) is designed around six core functions that range from protection, through to detection and response and recovery activities. Industry initiatives such as Sheltered Harbor similarly promote resilience plans that constitute a wide range of activities beyond ensuring liquidity in the event of a breach. The BIS, IOSCO and U.S. supervisory agencies define operational resilience as an outcome rather than an activity, that features a combination of financial and operational resources to prepare for, adapt to, recover from, and withstand disruptions that include operational hazards extending beyond cyber risk (BIS and IOSCO, 2016; Federal Reserve Board of Governors, 2023).[29] Broader definitions of resilience combine financial and operational risks to include both first- and second-order effects (Englund and Sosa, 2022).[30]

By subsuming all anticipatory and recovery operations into 'protection' and (balance-sheet) 'resilience' respectively, we acknowledge that our treatment of resilience is limited. Future work should incorporate other dimensions including the interaction between cyber and other operational risks, commonalities and distinctions between protection and response, and the cascading of risk along financial and operational dimensions.

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

[29] For examples of standard setting to facilitate operational resilience, see the interagency report SR 20-24 (Federal Reserve Board of Governors, 2020).

[30] Englund and Sosa (2022) use a network-theoretic approach to quantify *system resilience*, measured by the percentage of nodes that remain active after a disruption, multiplied by the value of the network. Subsequent work has estimated the transmission of cyber risk through networks such as wholesale payment systems (Kosse and Lu, 2022).

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 liquidation costs, a lender of last resort, and a positive recovery rate for investors 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

$$\alpha > \alpha^{IN} \equiv \frac{1}{\delta}\left(1 - \frac{FD}{RI}\right), \tag{18}$$

and it fails due to illiquidity whenever

$$\alpha > \alpha_{IL}(\ell) \equiv 1 - \frac{\ell FD}{RI}. \tag{19}$$

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.,

$$\frac{1}{\delta}\left(1 - \frac{FD}{RI}\right) = 1 - \frac{\widehat{\gamma} FD}{RI}, \tag{20}$$

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 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 $\alpha > \alpha^{IN}$, which occurs with probability greater than zero whenever $\alpha^{IN} < 1$. 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. Since α^{IN} is increasing in D and decreasing in F and S, suppose that $D = 1$, $F = r$ and $S = 0$. A sufficient condition for $\alpha^{IN} < 1$ is, thus, $\delta > 1 - \frac{r}{R}$. This establishes a well-defined upper dominance region $(\alpha^{IN}, 1]$.

The supposition in Proposition 1 ensures that $\underline{\alpha} \equiv \alpha^{IL}(1) > 0$ is the largest shock that the bank can withstand even if all fund managers withdraw early.[31] Threshold $\alpha^{IL}(1) > 0$ if and only if $RI > FD$. When $\alpha \in [0, \underline{\alpha}]$, fund managers have a dominant strategy to roll over their claims, forming the lower dominance region. For $\alpha \in (\underline{\alpha}, \overline{\alpha})$, if the shock were common knowledge, the

[31] When all fund managers withdraw, $\alpha^{IL}(\gamma) < \alpha^{IN}$ since $\delta < 1$.

run dynamics of fund managers would be characterised by multiple, self-fulfilling equilibria, as shown in Figure 7.

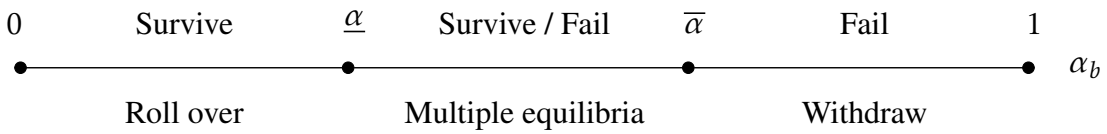


FIGURE 7. Tripartite classification of the accessibility shock.

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

$$s(x_k) = \begin{cases} \text{withdraw} & \text{if } x_k > x^*, \\ \text{roll over} & \text{if } x_k \leq x^*. \end{cases} \tag{21}$$

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

$$\ell(\alpha, x^*) = \Pr(x_k > x^* | \alpha) = 1 - H\Big(x^* - \alpha\Big), \tag{22}$$

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

$$\alpha^* = 1 - \ell^*(\alpha^*, x^*)\frac{F}{RI}. \tag{23}$$

For a given x^*, the left-hand side of Equation (23) 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

$$\gamma = \Pr(\alpha \leq \alpha^* | x_k = x^*). \tag{24}$$

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 Lemma 3.** 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. 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. 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 Lemma 4.** Suppose that the bank chooses $S = \overline{S}$. For all F, we have that $\pi(S = \overline{S}) \leq 0$ (with strict inequality for $\gamma > \widehat{\gamma}$) irrespective of whether a breach occurs, implying that this level of contribution to cybersecurity is not feasible. Moreover, at higher levels of cybersecurity contribution, $\pi(S > \overline{S}) < 0$. With a zero outside option, the bank will select $S < \overline{S}$ whenever it engages in financial intermediation.

A.5. **Proof of Proposition 2.** The Lagrange equation for the bank's problem is given by

$$\mathcal{L} = \pi(I,p) + \mu(1-I-S) + \phi\left(p - \sqrt{\frac{cS}{V}}\right), \tag{25}$$

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

$$\begin{aligned}\frac{\partial \pi}{\partial S} &= \frac{\partial p}{\partial S}\left[R(1-S) - F - \int_0^{\alpha^*(S)} EV(\alpha)d\alpha\right] \\ &- R\left[p(S) + \Big(1-p(S)\Big)\int_0^{\alpha^*(S)}(1-\delta\alpha)d\alpha\right] + \Big(1-p(S)\Big)EV(\alpha^*)\frac{\partial \alpha^*}{\partial S}\mathbf{1}_{\gamma > \widehat{\gamma}}. \qquad (26)\end{aligned}$$

Next, we need to argue that S^* is a maximum. Since $p(S)$ satisfies the Inada conditions, it follows that $\lim_{S\to 0}\frac{\partial \pi}{\partial S} = +\infty > 0$ and $\pi(S=0) > 0$. It follows from Lemma 4 that $\pi(S) \leq 0$ for all $S \geq \overline{S}$. 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

$$\begin{aligned}\frac{\partial^2 \pi}{\partial S^2} &= \frac{\partial^2 p}{\partial S^2}\left[R(1-S) - F - \int_0^{\alpha^*(S)} EV(\alpha)d\alpha\right] - 2\frac{\partial p}{\partial S}R\left[1 - \int_0^{\alpha^*(S)}(1-\delta\alpha)d\alpha\right] \qquad (27) \\ &- \Big(1-p(S)\Big)\frac{\partial \alpha^*}{\partial S}R\Big(1-\delta\alpha^*(S)\Big) \\ &- \frac{\partial \alpha^*}{\partial S}\left(EV(\alpha^*(S))\left[\frac{p(S)}{S} - \frac{1-p(S)}{1-S}\right] + \frac{F(1-\gamma\delta)}{1-S}\right) \times \mathbf{1}_{\gamma > \widehat{\gamma}}\end{aligned}$$

When failure is insolvency driven, the last term is zero. Combining the second and third terms, we get

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

We next introduce the following conjecture.

Conjecture 1. *In equilibrium, the bank's probability of winning the contest satisfies*

$$\frac{p(S)}{2S} - \frac{1-p(S)}{1-S} > 0.$$

Conjecture 1 immediately implies that $\frac{\partial^2 \pi}{\partial S^2} < 0$. We subsequently show in the proof of Proposition 3 that Conjecture 1 is always true in equilibrium.

When failure is illiquidity driven, combining the last three terms in Equation (27) yields

$$-\frac{p(S)}{S}\frac{R\delta}{2} - \frac{F^2\gamma}{R(1-S)^2}\left[p(S) + \left(\frac{p(S)}{S} - 1\right)(2-\gamma\delta)\right], \tag{28}$$

which is negative as long as $p(S) > \frac{(2-\gamma\delta)S}{2-\gamma\delta+S}$. Since the right-hand side of this condition is decreasing in $\gamma\delta$, a stricter condition is $p(S) > \frac{2S}{2+S}$, which is always satisfied under Conjecture 1.

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

$$\widehat{\gamma} = \frac{1}{\delta} - \left(\frac{1}{\delta} - 1\right)\frac{R\left(1 - S^*(\widehat{\gamma})\right)}{F}. \tag{29}$$

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 (29) is increasing in $\widehat{\gamma}$. At $\widehat{\gamma} = 0$, the left-hand side of Equation (29) 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.,

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

However, given the definition of $\overline{S}$, the above expression is exactly equal to zero, implying a contradiction. So the right-hand side of Equation (29) 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

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

Combining the terms together again, when failure is insolvency driven, we obtain

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

which is positive as long as

$$R>\underline{R}_F\equiv\frac{F}{(1-\delta)(1-S)}\frac{(1-p(S))}{p(S)}\left[\frac{p(S)}{1-p(S)}-\frac{2S}{1-S}\right].$$

Therefore, when $R>\underline{R}_F$, we have that $\frac{\partial^2\pi}{\partial S\partial F}>0$, and so $\frac{\partial S^*}{\partial F}>0$. When failure is illiquidity driven,

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

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

A.6. **Proof of Lemma 5.** Given F, the value of a debt claim is

$$\mathcal{V}(F) = \Big(p(S) + \Big(1 - p(S)\Big)\alpha^*(F,S)\Big)F\,. \tag{30}$$

Following a marginal increase in F, we have that

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

where $\alpha^*(S) + F\frac{\partial \alpha^*}{\partial F} = \frac{1}{\delta}\Big(1 - \frac{2F}{R(1-S)}\Big)$ for $\gamma < \widehat{\gamma}$, and $\alpha^*(S) + F\frac{\partial \alpha^*}{\partial F} = 1 - \frac{2\gamma F}{R(1-S)}$ when $\gamma > \widehat{\gamma}$. Since $\frac{\partial \mathcal{V}}{\partial F}$ is linear in F, and $\frac{\partial^2 \mathcal{V}}{\partial F^2} < 0$, it follows that $\mathcal{V}(F)$ is quadratic in F. In addition, for $F = r$, we have $\mathcal{V}(r) < r$, since $\alpha^* < 1$. Moreover, since $\frac{\partial \mathcal{V}}{\partial F}\Big|_{F=0} > 0$ and $\frac{\partial \mathcal{V}}{\partial F}\Big|_{F=R} < 0$, it implies that if a solution to $\mathcal{V}(F) = r$ exists, then there are always two solutions. In what follows, we focus on the smaller solution, which is justified under perfect competition. Denoting this F^*, we thus have that $\frac{\partial \mathcal{V}}{\partial F}\Big|_{F=F^*} > 0$.

Next, note that

$$\frac{1}{F}\frac{\partial \mathcal{V}}{\partial S} = \frac{\partial p}{\partial S}\Big(1 - \alpha^*(S)\Big) + \Big(1 - p(S)\Big)\frac{\partial \alpha^*}{\partial S}. \tag{31}$$

For $\gamma > \widehat{\gamma}$, we have that $\frac{\partial \alpha^*}{\partial S} = \frac{\alpha^*(S)-1}{1-S}$. This implies that $\frac{\partial \mathcal{V}}{\partial S} > 0$ if and only if $S < S^\dagger$, where $S^\dagger$ solves $\frac{p(S^\dagger)}{1-p(S^\dagger)} = \frac{2S^\dagger}{1-S^\dagger}$. By the implicit function theorem we obtain $\frac{\partial F^*}{\partial S} < 0$ if and only if $S < S^\dagger$.

For $\gamma < \widehat{\gamma}$, we have

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

For $S \geq S^\dagger$, the above expression in strictly negative. At the same time, in the limit $S \to 0$, we have $\frac{1}{F}\frac{\partial \mathcal{V}}{\partial S} > 0$. This suggests that there exists a critical threshold $S^{\dagger\dagger} < S^\dagger$ at which $\frac{1}{F}\frac{\partial \mathcal{V}}{\partial S}\Big|_{S=S^{\dagger\dagger}} = 0$. Therefore, by the implicit function theorem, we obtain that $\frac{\partial F^*}{\partial S} < 0$ if and only if $S < S^{\dagger\dagger}$. We thus define $\tilde{S} = S^\dagger$ if $\gamma > \widehat{\gamma}$ and $\tilde{S} = S^{\dagger\dagger}$ if $\gamma < \widehat{\gamma}$.

A.7. **Proof of Proposition 3.** In what follows, we argue that there is a single crossing between the curves $S^*(F)$ and $F^*(S)$. First, note that $S^*(F = r) < 1 < r < F^*(S = 0)$. And so, $S^*(F = r) < F^*(S = S^*(F = r))$. Thus, the curve $F^*(S)$ starts off "above" $S^*(F)$.

Suppose $\gamma > \widehat{\gamma}$. Denote $F^\dagger$ as the face value of debt that induces the bank to choose $S^*(F^\dagger) = S^\dagger$. We can explicitly solve for $F^\dagger$ and obtain

$$F^\dagger = \frac{R}{\gamma(1-\delta)}\left\{\delta\left[\frac{3V}{c} - \frac{3\sqrt{V(V-c)}}{c} - 2\right] + (2-\delta)\frac{V}{c}\sqrt{\frac{2V - c - 2\sqrt{V(V-c)}}{V}}\right\},$$

where $\frac{\partial F^\dagger}{\partial c} > 0$ and $\lim_{c\to 0} F^\dagger = \frac{R}{\gamma} > R$, implying that $S^\dagger$ is not feasible in any equilibrium. This, in turn, proves Conjecture 1 that $\frac{p(S^*)}{2S^*} - \frac{1-p(S^*)}{1-S^*} > 0$, or equivalently, $S^*(F) < S^\dagger$.

And so $F^\dagger > R > F^*(S = S^\dagger)$. The curve $S^*(F)$ thus ends up "above" $F^*(S)$. Since $F^*(S)$ is decreasing over $[0, S^\dagger]$, while $S^*(F)$ is increasing, the intersection is unique. As we have previously argued, at this equilibrium, the threshold $\widehat{\gamma}(F^{**}, S^{**})$ is well defined.

Suppose $\gamma \leq \widehat{\gamma}$. Denote by $F^{\dagger\dagger}$ the face value of debt such that $S^*(F^{\dagger\dagger}) = S^{\dagger\dagger}$. There are two cases to consider. First, suppose $F^*(S^{\dagger\dagger}) < F^{\dagger\dagger}$. In this case $S^*(F)$ intersects $F^*(S)$ in its downward sloping section. Moreover, this intersection is unique – and the threshold $\widehat{\gamma}(F^{**}, S^{**})$ is well defined – since $\tilde{F} > R$, where $\tilde{F}$ is the face value of debt such that $S^*(\tilde{F}) = S^\dagger$, where $S^\dagger > S^{\dagger\dagger}$. We can explicitly solve for $\tilde{F}$ and obtain

$$\tilde{F} = \frac{R}{1-\delta}\left\{(2\delta-1)\left[\frac{3V}{c} - \frac{3\sqrt{V(V-c)}}{c} - 2\right] + \frac{V}{c}\sqrt{\frac{2V - c - 2\sqrt{V(V-c)}}{V}}\right\},$$

which satisfies $\frac{\partial \tilde{F}}{\partial c} > 0$ and $\lim_{c\to 0} \tilde{F} = 2R > R$.

Next, suppose that $F^*(S^{\dagger\dagger}) > F^{\dagger\dagger}$. In this case, the intersection occurs in the upward sloping part of $F^*(S)$. Moreover, $\frac{\partial^2 F^*}{\partial S^2} \propto -\left[\frac{\partial \mathcal{V}}{\partial F}\frac{\partial^2 \mathcal{V}}{\partial S^2} - \frac{\partial^2 \mathcal{V}}{\partial F \partial S}\frac{\partial \mathcal{V}}{\partial S}\right] > 0$ while $\frac{\partial^2 S^*}{\partial F^2} \propto -\left[\frac{\partial^2 \pi}{\partial S^2}\frac{\partial^3 \pi}{\partial S \partial F^2} - \frac{\partial^3 \pi}{\partial S^2 \partial F}\frac{\partial^2 \pi}{\partial S \partial F}\right] < 0$, implying that the intersection is unique and the threshold $\widehat{\gamma}(F^{**}, S^{**})$ is well defined.

A.8. **Proof of Proposition 4.** 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

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

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

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

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

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

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.9. **Proof of Proposition 5.** In what follows, let us refer to the planner's objective by $W(S)$, which is concave and well defined, and define $\Delta(S) \equiv -\lambda\Big(1-p(S)\Big)\Big(1-\alpha^*(S)\Big)$. We have that $\frac{\partial W}{\partial S} = \frac{\partial \pi}{\partial S} + \frac{\partial \Delta}{\partial S}$. Evaluating the planner's first-order condition at the bank's private choice, S^*, such that $\frac{\partial \pi}{\partial S}\Big|_{S=S^*} = 0$, we obtain that $\frac{\partial W}{\partial S}\Big|_{S=S^*} = \frac{\partial \Delta}{\partial S}\Big|_{S=S^*}$. Consequently, if $\frac{\partial \Delta}{\partial S}\Big|_{S=S^*} < 0$, we have that the bank over-invests in cybersecurity, while if $\frac{\partial \Delta}{\partial S}\Big|_{S=S^*} > 0$, it under-invests. In what follows, 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

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

Next, in the limit $c \to 0$, we have that $\lim_{c\to 0} S^* = 0$. And so we obtain

$$\lim_{c\to 0}\frac{1}{\lambda}\frac{\partial \Delta}{\partial S}\bigg|_{S=S^*} = \Big(1-\alpha^*(0)\Big)\left(\lim_{c\to 0}\frac{p(S^*)}{2S^*} - 1\right) - \left(\frac{1}{\delta}-1\right). \tag{32}$$

For there to be over-investment in cybersecurity, we require that

$$\lim_{c\to 0}\frac{p(S^*)}{2S^*} < 1 + \frac{\frac{1}{\delta}-1}{1-\alpha^*(0)}. \tag{33}$$

The expression on the left-hand side does not converge to zero but instead to some positive value. To see this, note that, following L'Hôpital's rule, $\lim_{c\to 0}\frac{c}{S^*} = \frac{1}{\frac{\partial S^*}{\partial c}\big|_{c=0}} > 0$. The term $\frac{\partial S^*}{\partial c}\big|_{c=0} > 0$ reflects the marginal change in the bank's contribution to cybersecurity following an increase in c when $c = 0$. Since there is a small chance that the bank will still win the contest against the attacker, the bank responds by increasing its contribution to cybersecurity. Finally, since the left-hand side of Equation (33) is decreasing in R, while the right-hand side is increasing in R, there exists $\underline{R}_c$, which solves $\frac{1}{2}\sqrt{\frac{1}{\frac{\partial S^*}{\partial c}\big|_{c=0,R=\underline{R}_c}}} = 1 + \frac{F(1-\delta)}{F-R(1-\delta)}$, such that for $R > \underline{R}_c$, the inequality in Equation (33) is satisfied. In this case, we obtain that $S^P(F) < S^*(F)$, i.e., the bank contributes excessively to cybersecurity.

Next, in the limit $c \to V$, we have $S^* = \overline{S}$ and $\alpha^*(\overline{S}) = 0$. Moreover, as the attacker becomes less sophisticated, the probability that the bank wins the contest increases. Accounting for both the direct and indirect effect of c on $p(S^*)$, we obtain that $\lim_{c\to V} p(S^*) = 1$. In sum, we obtain $\lim_{c\to V}\frac{1}{\lambda}\frac{\partial \Delta}{\partial S}\big|_{S=S^*} > 0$ implying that $S^P(F) > S^*(F)$ and there is under-investment in cybersecurity by the bank.

Finally, to determine the critical threshold in terms of c where the S^* and S^P curves cross, note that

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

and $\frac{\partial\pi}{\partial S}\Big|_{S=S^*} = 0$ by definition for all c. Since we have already shown that $\lim_{c\to 0}\frac{\partial\Delta}{\partial S}\Big|_{S=S^*} < 0$ and $\lim_{c\to\infty}\frac{\partial\Delta}{\partial S}\Big|_{S=S^*} > 0$, it follows from the Intermediate Value Theorem that there exists some $\widehat{c}$ such that $\frac{\partial\Delta}{\partial S}\Big|_{S=S^*,\ c=\widehat{c}} = 0$, in which we obtain $S^* = S^P$. 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

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

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

A.10. **Proof of Proposition 6.** To show that the difference between the planner's solution and the bank's private solution is increasing in c, we need to establish that

$$\frac{\partial}{\partial c}\frac{1}{\lambda}\frac{\partial\Delta}{\partial S}\bigg|_{S=S^*} > 0. \tag{34}$$

That is, accounting for the indirect effect, $\frac{\partial S^*}{\partial c}$, we need that the cross-partial of the externality term is positive.

We first consider the case where bank failure is driven by illiquidity and then turn our focus to the case where it is driven by insolvency.

Illiquidity. For $\gamma > \widehat{\gamma}$, we have that

$$\begin{aligned}
\frac{\partial}{\partial c}\frac{1}{\lambda}\frac{\partial \Delta}{\partial S}\bigg|_{S=S^*} &= \frac{1}{\lambda}\frac{\partial^2 \Delta}{\partial S \partial c}\bigg|_{S=S^*} + \frac{1}{\lambda}\frac{\partial^2 \Delta}{\partial S^2}\bigg|_{S=S^*} \times \frac{\partial S^*}{\partial c} \\
&= \Big(1-\alpha^*(S^*)\Big)\Big(\frac{1}{2S^*}+\frac{1}{1-S^*}\Big)\times \frac{\partial p}{\partial c}\bigg|_{S=S^*} \\
&+ \left\{ -\frac{\partial \alpha^*}{\partial S}\bigg|_{S=S^*}\left(\frac{p(S^*)}{2S^*}-\frac{1-p(S^*)}{1-S^*}\right)\right. \\
&+ \left.\Big(1-\alpha^*(S^*)\Big)\left[-\frac{p(S^*)}{4(S^*)^2}-\frac{2S^*-p(S^*)(1+S^*)}{2S^*(1-S^*)^2}\right]\right\}\times \frac{\partial S^*}{\partial c}.
\end{aligned}$$

Since $S^* < \tilde{S}$, it follows that $p(S^*) > \frac{2S^*}{1+S^*}$. Thus, insofar that $\frac{\partial S^*}{\partial c} > 0$, it follows that a sufficient condition for the inequality in Equation (34) to be satisfied is

$$p(\tilde{S}) > \frac{4(\tilde{S})^2}{(\tilde{S}-2)^2+3}. \tag{35}$$

If the above condition is satisfied at $\tilde{S}$, then it will be satisfied at S^*. The condition above involves only two variables, c and V and it is true for all $c < V$. Thus, $|S^P - S^*|$ is increasing in c.

We can extend the logic from above to show $|S^P - S^*|$ is increasing in γ. In this case, the direct effect is $-\frac{\partial \alpha^*}{\partial S}\big|_{S=S^*}\left(\frac{p(S^*)}{2S^*}-\frac{1-p(S^*)}{1-S^*}\right) > 0$. And for the indirect effect, since $\frac{\partial S^*}{\partial \gamma} > 0$, we have our result whenever the condition in Equation (35) is satisfied.

Insolvency. For $\gamma < \widehat{\gamma}$, the expression $\frac{1}{\lambda}\frac{\partial \Delta}{\partial S}\big|_{S=S^*}$ differs with that in the case of illiquidity by a factor

$$\Xi \equiv -\left(\frac{1}{\delta}-1\right)\frac{1-p(S^*)}{1-S^*}.$$

So in the case $c > \hat{c}$, we have that

$$\begin{aligned}
\frac{d\Xi}{dc}\bigg|_{S=S^*} &= \frac{\partial \Xi}{\partial S}\bigg|_{S=S^*} \times \frac{\partial S^*}{\partial c} \\
&= -\left(\frac{1}{\delta}-1\right)\frac{2S^*-p(S^*)(1+S^*)}{(1-S^*)^2} \\
&= \left(\frac{1}{\delta}-1\right)\frac{p(S^*)(1+S^*)-2S^*}{(1-S^*)^2} > 0,
\end{aligned}$$

since $S^* < \tilde{S}$. And so the additional effect via Ξ reinforces the other indirect effects identified in the illiquidity case, implying that for $c > \hat{c}$, the difference $|S^P - S^*|$ is increasing in c.

A.11. **Proof of Proposition 7.** The planner's benchmark is derived in Proposition 5, which shows that the bank over-invests in cybersecurity when $\gamma < \widehat{\gamma}$. Proposition 2 suggests that $\frac{\partial \pi}{\partial S} > 0$ at $S^P < S^*$. Therefore, the introduction of a constraint $S \leq S^P$ through the enforcement of minimum operational resilience standards is binding and the bank allocates exactly $S = S^P$ towards cybersecurity.

A.12. **Proof of Proposition 8.** By Proposition 5, 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

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

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.13. **Proof of Proposition 9.** 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 5, 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.14. **Proof of Proposition 10.** With a negligence rule in place, expected profits include a penalty, $\kappa(S)$, that is implemented conditional on a successful cyberattack

$$\begin{aligned} \pi(S) &= p(S)(RI-F) && (40)\\ &+ (1-p(S))\int_0^{\alpha^*(I)}\Big[RI(1-\alpha\,\delta)-F-\kappa(S)\Big]d\alpha\,. && (41)\end{aligned}$$

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 5 and 6, 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.15. **Proof of Proposition 11.** For small but positive noise, the bank's expected profits are given by

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

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

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

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.

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, P. Gai, and J. Chapman (2019). Asset encumbrance, bank funding and fragility. Review of Financial Studies 32(6), 2422–2455.

Ahnert, T., M. Brolley, D. A. Cimon, and R. Riordan (2024). Cyber risk and security investment. Mimeo, European Central Bank.

Aldasoro, I., L. Gambacorta, P. Giudici, and T. Leach (2022). The drivers of cyber risk. Journal of Financial Stability 60, 100989.

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

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

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.

BIS and IOSCO (2016, June). Guidance on cyber resilience for financial market infrastructures. CPMI Papers, No. 146.

Bloomberg (2021, February 5). How a dated cyber-attack brought a stock exchange to its knees. Bloomberg Businessweek.

Brando, D., A. Kotidis, A. Kovner, M. Lee, and S. L. Schreft (2022). Implications of cyber risk for financial stability. FEDS Notes, May 12.

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.

Chevalier, J. and G. Ellison (1997). Risk taking by mutual funds as a response to incentives. Journal of Political Economy 105(6), 1167–1200.

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.

CVE (2024, February). Cve-2024-21626. https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2024-21626.

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.

Eisenbach, T., A. Kovner, and M. J. Lee (2023). When it rains it pours: Cyber vulnerabilities and financial conditions. Federal Reserve Bank of New York Staff Report, no. 1022.

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.

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 (2020). Sr 20-24: Interagency paper on sound practices to strengthen operational resilience. https://www.federalreserve.gov/supervisionreg/srletters/SR2024a1.pdf.

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.

Financial Times (2024a, January 17). JPMorgan suffers wave of cyber attacks as fraudsters get 'more devious'. Financial Times.

Financial Times (2024b, November 15). The alarming cyber hack at ICBC. Financial Times. https://www.ft.com/content/d3c7259c-0ea6-414b-9013-ac615b1a8177.

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.

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.

HackerOne (2023, October 25). 7th annual hacker-powered security report. https://www.hackerone.com/reports/7th-annual-hacker-powered-security-report.

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

Harmon, R. L. (2020). Cloud concentration risk II: what has changed in the past two years? Technical report, Cloudera.

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.

IMF (2024). Global financial stability report: The last mile: Financial vulnerabilities and risks.

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

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.

Kosse, A. and Z. Lu (2022). Transmission of cyber risk through the canadian wholesale payment system. Journal of Financial Market Infrastructures 10(4).

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.

Lipp, M., M. Schwarz, D. Gruss, T. Prescher, W. Haas, A. Fogh, J. Horn, S. Mangard, P. Kocher, D. Genkin, Y. Yarom, and M. Hamburg (2018). Meltdown: Reading kernel memory from user space. In 27th USENIX Security Symposium (USENIX Security 18), Baltimore, MD, pp. 973–990. USENIX Association. August.

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

Bank of England (2023). Systemic Risk Survey Results – 2023 H2. https://www.bankofengland.co.uk/systemic-risk-survey/2023/2023-h2.

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.

European Union Agency for Cybersecurity (2023). Enisa threat landscape 2023. The European Union Agency for Cybersecurity Report, October.

Financial Times (2024, January 15). Cyber attacks reveal fragility of financial markets. https://www.ft.com/content/a8b8de58-8691-4ece-ade3-5b7be63dbef2.

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

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

NIST (2024, February 26). The nist cybersecurity framework (csf) 2.0. https://doi.org/10.6028/NIST.CSWP.29.

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.

Rajan, R. G. (2006). Has finance made the world riskier? European Financial Management 12(4), 499–533.

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.

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

Sophos (2023). The state of ransomware in financial services 2023. Sophos Whitepaper, July. 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.

The Banker (2023, November 22). The significance of the ICBC FS hack on the US Treasury market. The Banker. https://www.thebanker.com/The-significance-of-the-ICBC-FS-hack-on-the-US-Treasury-market-1700578028.

Tullock, G. (2008). Efficient Rent Seeking, pp. 105–120. Berlin, Heidelberg: Springer Berlin Heidelberg.

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

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.