CYBERSECURITY AND FINANCIAL STABILITY

KARTIK ANAND, CHANELLE DULEY, PRASANNA GAI

ABSTRACT. Cyber risk exposes banks to operational disruptions that can trigger runs. A bank chooses its cybersecurity by trading off protection against attacks with remaining resilient if an attack succeeds. Cybersecurity functions as a risk-management decision: it reduces the bank's exposure to adverse outcomes but entails lower balance-sheet returns. Equilibrium cybersecurity depends on whether failure is driven by insolvency or illiquidity. When failure is insolvency-driven, bank and creditor actions reinforce one another: greater cybersecurity leads to a higher debt burden, which strengthens incentives for protection. When failure is illiquidity-driven, additional cybersecurity lowers the debt burden, eliminating the bank's private risk–return trade-off. Socially optimal cybersecurity differs from the private choice, and corrective instruments must target either the protection or resilience margins. We extend the model to a system-wide environment in which cybersecurity is a public good, highlighting free-riding and the need for targeted regulation.

Keywords: Cybersecurity, cyberattacks, bank runs, global games

JEL classifications: G01, G20, G21, G28.

We thank the managing editor, Marcus Opp, for suggestions that materially improved the paper. We also 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.

1. INTRODUCTION

From mobile phone-based solutions for customers to virtual data rooms automating due diligence processes, modern banks bear little resemblance to the brick-and-mortar institutions of yesteryear. But as the digital transformation in banking has gathered pace, so too have cyber risks.[1] Cyberattacks are perceived by financial market participants as the most significant source of risk to the financial system (Bank of England, 2023). Academics and policymakers have also highlighted how cyberattacks have the potential to trigger bank runs and disrupt wholesale funding markets (Duffie and Younger, 2019; 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, illustrates the potential risk to financial stability. An attacker ("Lockbit") gained access to ICBC's systems by exploiting a vulnerability in its remote desktop software ("Citrix") commonly used by many banks. The attack impaired operating systems, including those used by ICBC to clear US Treasury trades and repo transactions, and facilitated the theft of confidential data. ICBC's parent company averted widespread financial disruption by providing an emergency liquidity injection of $9bn and paying an undisclosed ransom (The Banker, 2023). Although the financial consequences of the attack on ICBC financial services remain unclear, Huang and Wang (2021) show that such data breaches significantly increase the cost of borrowing for affected firms.

In this paper, we study how cyberattacks influence banks' security choices and the risk of bank runs in a theoretical model that takes financial market reactions into account. At the heart of our analysis is the observation that a bank's cybersecurity investment is a *risk-management decision*. Increasing the protection of its IT infrastructure reduces the likelihood of a successful cyberattack—tilting probability mass away from adverse outcomes—but doing so diverts resources

[1]The European Union Agency for Cybersecurity (2023) reported an increase of 22% in the volume of DDOS (distributed-denial-of service) attacks on financial institutions in 2022. And industry surveys suggest that the rate of ransomware attacks on financial services rose from 21% in 2021 to 64% in 2023 (Sophos, 2023). The annual cost of cyberattacks – including theft, lost productivity and reputational harm – is estimated at $10.5tn globally (Financial Times, 2024a).

from profitable investment.[2] This reduces the bank's equity value and leaves its balance sheet less able to absorb losses in the event that an attack does occur. The bank therefore faces a fundamental trade-off: greater protection lowers downside risk, but at the cost of lower returns and weaker balance-sheet resilience.[3]

This interaction between protection and balance-sheet resilience is central to how cybersecurity affects run risk: weaker resilience raises the likelihood that a successful attack triggers depositor withdrawals. Thus, cybersecurity spending influences not only the likelihood of loss but also the bank's susceptibility to runs: lower resilience raises the chance that a successful attack triggers withdrawals. While we study cyberattacks, our underlying analytical framework captures a wider class of operational failures in which the likelihood and severity of losses depend on the bank's preventive investment. Cyber risk is thus a prominent, policy-relevant instance of a more general operational risk management problem.

Main results. In the unique equilibrium of our model, the bank's decision to allocate resources to cybersecurity depends crucially on the nature of bank fragility. When bank failure is driven by insolvency considerations (i.e., the run is "fundamentals" based), the actions of the bank and its creditors are strategic complements. Anticipating that the bank chooses more cybersecurity to thwart cyberattacks, creditors reason that, in the event of an attack, the loss of balance sheet resilience increases the ex-ante likelihood of bank failure. So they demand greater compensation. Faced with a higher face value of debt and a lower opportunity cost to improve its protection, the bank is incentivised to increase cybersecurity.

But when bank failure is driven by illiquidity (i.e. the run is self-fulfilling or "belief-based" in nature), the actions of the bank and creditors are strategic substitutes. The increased fragility implies that the bank is more likely to fail in the event of a cyberattack due to a bank run. As a consequence, by investing more in cybersecurity, creditors perceive a higher likelihood of repayment

[2]The resources dedicated by banks towards IT security are considerable. For example, JP Morgan Chase allocates an annual budget of $15 billion, or one-third of the value of its annual investments, towards cybersecurity. The budget allows for 62,000 technologists to arm the bank against hacking attempts (Financial Times, 2024b).

[3]We use the term "resilience" to describe the strength of a bank's balance sheet and its ability to service withdrawals in the event of a cyberattack. This is distinct from notions of "resilience" in cybersecurity circles, where it refers to anticipating, withstanding, and recovering from cyberattacks, (NIST, 2020).

and require a lower face value of debt in compensation. Although this increases the opportunity cost of protection, the bank is nevertheless incentivised to increase its cybersecurity allocation because of the increased fragility and likelihood of failure. Across both regimes, the equilibrium level of cybersecurity is determined by the same underlying logic: the bank increases protection up to the point where the marginal benefit of shifting probability mass toward the no-attack outcome equals the marginal cost of reduced investment and resilience. What differs between insolvency- and illiquidity-driven failure is how investor beliefs and the failure threshold shape this marginal benefit.

In our model, cyber risk reshapes the bank's risk–return trade-off. When failure is illiquidity-driven, a marginal increase in cybersecurity investment delivers two reinforcing effects. First, by reducing the likelihood of a successful attack, it lowers run risk and therefore raises the probability that investors are repaid. Investors respond by requiring a lower face value of debt, reducing the bank's debt burden and lifting expected equity. Second, greater protection directly reduces the bank's chance of suffering losses. Together, these forces imply that—in contrast to the usual risk–return tension— returns and fragility move in the same direction. More protection both raises expected equity value and makes the bank less susceptible to runs.

When failure is insolvency-driven, additional protection crowds out resilience, and the familiar risk–return trade-off re-emerges. Beyond the cybersecurity application, the analysis provides a framework for bank risk-management under incomplete information, showing how strategic complementarities, debt pricing, and run thresholds jointly shape the link between operational risk mitigation, returns, and balance-sheet resilience. The global-games structure that generates the illiquidity region is therefore central not only to our cyber setting but also to understanding operational risk practices more broadly.

Our analysis generates a rich set of comparative static results. The deep parameters of our model collectively define a *cybersecurity landscape*—capturing attacker sophistication, rollover risk, the severity of cyber–induced losses, and the availability of emergency liquidity. Each element of this landscape reshapes the balance between protection and resilience. A less sophisticated attacker

increases the bank's ability to detect and mitigate vulnerabilities. This strengthens the return to protection: the bank invests more in cybersecurity, while creditors perceive lower repayment risk and therefore demand a lower face value of debt. Greater protection reduces downside risk but diverts resources from resilience, so the net effect on resilience is ambiguous in equilibrium.

Higher rollover risk makes cyberattacks more salient by raising the likelihood that even modest impairments trigger inefficient, coordination-driven runs. A cyberattack therefore becomes a more potent source of fragility. This amplifies the value of protection relative to resilience, leading to more cybersecurity investment. From the investor side, higher rollover risk reduces the likelihood of repayment in the event of an attack; creditors respond by seeking a higher face value of debt, which further reinforces the bank's incentive to allocate more to cybersecurity.

When cyberattacks become more costly, the gap between in returns between the no-attack and attack states widens. Protection becomes more valuable because preventing an attack avoids a larger fall in equity value. For creditors, there is a reduced likelihood of repayment if bank failure is insolvency driven, leading them to demand a higher face value of debt. Both forces push the bank toward stronger protection.

Finally, the robustness of emergency liquidity assistance also shapes the cybersecurity landscape. Improved emergency payment infrastructure reduces the consequences of a successful attack, lowering the bank's incentive to invest in cybersecurity: with reliable backup mechanisms, the payoff to protection falls. At the same time, creditors feel more assured of repayment when disruptions can be bridged by such facilities, leading them to require a lower face value of debt. Consequently, stronger emergency liquidity reduces protection but improves balance-sheet resilience.

In our model, the bank does not take into account the social consequences of its cybersecurity choices, such as disruptions to the macroeconomy in the event of a cyberattack. A social planner, on the other hand, takes these into consideration when choosing cybersecurity. The misalignment between the bank and the planner produces a wedge between the private and socially optimal contribution to cybersecurity, and the direction of the inefficiency depends critically on the nature of bank fragility, i.e. whether cyberattacks trigger illiquidity or insolvency.

When failure is illiquidity driven, the bank underinvests in cybersecurity relative to the planner. This is because when creditors are increasingly skittish, the marginal benefit of higher protection to the bank is greater. But the planner prefers an even higher allocation, given the social costs of bank failure. By contrast, when failure is triggered by insolvency, the bank overinvests in security relative to the planner. Concerned with its own survival, the bank is incentivised to allocate more resources towards cybersecurity. But the diversion of these resources means that the bank is less resilient in the event of a successful attack. The planner, who seeks to reduce the social costs of a bank failure prefers, instead, that the bank focuses on balance sheet resilience and allocates less to cybersecurity.

Our analysis sheds light on the role of subsidies in dealing with cyber risks. For example, revenue-neutral Pigovian policies targeted at either shoring up the bank's protection or balance sheet resilience, depending on whether bank failure is driven by illiquidity or insolvency, achieve the planner's constrained-efficient outcome. These subsidies alter the bank's incentives to invest in cybersecurity without changing the behaviour of the bank's creditors. Examples include the Dutch Central Bank's *subsidie cyberweerbarheid* program and the Monetary Authority of Singapore's *Cybersecurity Capability Grant* program. Direct technical assistance, e.g., so-called *red-team testing* is another type of subsidy offered in many jurisdictions. Such technical assistance is more than just an IT exercise to improve protection—it boosts investor confidence and leads to the favourable valuation of debt. Since the bank's chances of repelling an attack are improved, technical assistance incentivises it to allocate more towards cybersecurity. But, unlike Pigovian policies, technical assistance is a second-best policy that cannot achieve the constrained-efficient outcome.

A key feature of the cyberattack on ICBC was the reliance on a software platform (Citrix) widely used by other market participants. Motivated by this, we extend our core analysis to allow for multiple and heterogeneous banks and situations where cybersecurity has public good characteristics. While banks exhibit incentives to free-ride on the cybersecurity investments of others, we show that our result on overinvestment in cybersecurity when bank failure is driven by insolvency remains robust. But policies aimed at achieving the constrained-efficient solution must be tailored to the type of public good. When cybersecurity is a best-shot public good, policies should target the

bank with the strongest incentive to invest in cybersecurity. And when cybersecurity is a weakest-link public good, policies should apply to all banks, but be formulated based on the bank with the lowest incentives to invest in cybersecurity.

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. Existing literature in the area is largely empirical in nature.[4] Duffie and Younger (2019) describe cyber-runs and conduct a stress test to understand the resilience of systemically important banks to wholesale depositor runs following a cyber-attack. Relatedly, Jamilov et al. (2021) study cyber risk and contagion using a text-based measure of cyber risk. They also explore whether cyber risk exposure influences asset pricing and how the effects of this exposure propagate among firms.[5] And Eisenbach et al. (2025) highlight the interdependence between cyber risk and the financial cycle, finding evidence of heightened cyber risk during stressed financial conditions.

Our work complements Eisenbach et al. (2022). They examine how cyberattacks impair a bank's ability to repay withdrawing creditors and discuss how this influences creditors' incentives to run. In their analysis, 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. By contrast, in our model, we distinguish between cyber-attack induced crises driven by solvency and liquidity considerations. Whenever rollover risk is low, bank failure following a cyberattack is driven 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 the bank is transient in nature and inefficient self-fulfilling runs are a source of bank failure.

[4] A growing policy literature has also examined the relationship between cyberattacks and financial stability (Kashyap and Wetherilt, 2019; Adelmann et al., 2020; Elestedt et al., 2021; Fell et al., 2022). These papers suggest that more should be done to bolster banks' resilience in the face of an attack but do not provide any formal modelling in support of their policy prescriptions.

[5] See also Aldasoro et al. (2021) Kamiya et al. (2021), Woods et al. (2021) and Florackis et al. (2023).

We also contribute to the literature on the economics of cybersecurity. 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 firms' 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) extends this analysis to the case of multiple firms with network externalities, and where cybersecurity has the properties of a public good. He shows that under-provision of cybersecurity at the system level can be rectified with negligence rules. Ahnert et al. (2024) propose a principal-agent model of cybersecurity investment by firms that are delegated the task of protection by clients.[6] They find that both the attacker's mobility and firms' security investment decisions depend crucially on whether cybersecurity investment is observable. Our analysis shows how ex post coordination failure in the form of bank runs, the sophistication of the attacker, and the nature of bank fragility shape a bank's incentives to provide cybersecurity.

Our analysis complements two strands of the banking literature. The first is the bank risk-taking literature, in which higher leverage creates risk-shifting incentives: shareholders prefer mean-preserving spreads because they capture upside gains while debtholders bear downside losses (e.g., Jensen and Meckling, 1976). This mechanism underlies much of the modern theory of capital regulation (e.g., Biais and Casamatta, 1999; Repullo, 2004; Martinez-Miera and Repullo, 2017). Our model departs from this paradigm by showing that when operational loss risk is endogenous—as with cyberattacks—greater leverage can discipline the bank. The prospect of attack-induced losses strengthens the incentive to invest in protection, reversing the canonical risk-shifting logic. Moreover, in the presence of rollover risk, the risk–return trade-off can locally vanish, generating a virtuous cycle in which greater protection simultaneously improves expected equity value and lowers fragility. This positive feedback echoes the charter-value logic of Keeley (1990), in contrast with the negative feedback loops highlighted in models of financial fragility such as He and Xiong (2012) and Moreno and Takalo (2016).

[6]See also Ramírez (2025) who considers an attacker-defender framework to study cybersecurity and argues that the size of the defender is a key determinant shaping equilibrium behaviour.

Second, our framework connects to the literature on risk-management. In models such as Froot et al. (1993), firms use costly hedging instruments to reallocate payoffs away from adverse states, reducing the dispersion of outcomes at the expense of lower average returns. However, risk-management competes with productive investment given balance sheet constraints (Rampini and Viswanathan, 2010). We model cybersecurity investment as a decision that affects tail risk and expected returns (e.g., Biais et al., 2010; Hoffmann et al., 2022): it reallocates probability mass away from adverse states (protection), but reduces productive investment and hence the bank's ability to absorb losses if an attack succeeds (resilience).[7] This dual role generates an interplay between protection, resilience, and funding conditions.

Finally, we add to the large literature on bank runs and global games (Morris and Shin, 2003; Goldstein and Pauzner, 2005; Ahnert et al., 2019; Kashyap et al., 2024). We specifically build on Rochet and Vives (2004), where unsecured debtholders delegate their rollover decisions to professional managers so that rollover decisions are global strategic complements. Our contribution shows how cyber risk interacts with run risk in such a setting.

2. Model

We develop a model of a bank facing adversarial or preventable operational risks. Although our exposition focuses on cyberattacks, where an external attacker invests effort to breach the bank's defences, the analytical framework also applies to more general operational-risk environments in which loss events are endogenously affected by the bank's own preventive investment, including sophisticated fraud schemes, insider manipulation, or avoidable IT disruptions. We consider cybersecurity to be a salient and policy-relevant instance of this broader issue.

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

[7] This mechanism also bears resemblance to the classic self-protection and self-insurance trade-off highlighted by Ehrlich and Becker (1972).

endowed with a unit of the consumption good and has access to a risk-free storage technology that yields $r > 1$ per unit of investment. We assume that the bank is subject to limited liability and that the attacker is deep pocketed.[8] Without loss of generality, the bank's endowment is normalised to zero.

Bank balance sheet. The bank has access to a safe and liquid asset at $t = 0$ with a return $R > r$ per unit of investment at $t = 2$. To finance this investment, the bank issues $D > 0$ of uninsurable short-term debt claims to investors.[9] Let $F > 0$ be the face value of debt and suppose that it is independent of the withdrawal date. The bank invests $I \leq D$ in the asset and allocates the rest, $S \equiv D - I$, to its *cybersecurity*. Investors who deposit with the bank do not observe its cybersecurity allocation.[10]

IT systems and cybersecurity. Banks use 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, e.g., finding vulnerabilities and mitigating them. But if the attacker discovers these vulnerabilities, it can exploit them for private gain and disrupt the bank. Allocating more to cybersecurity improves the bank's chances of thwarting cyberattacks, with economic benefits for the bank.[11] Such investments are accounted for on the bank's balance sheet. But since cybersecurity is ingrained in, and inalienable from, the institution, the bank cannot sell or trade cybersecurity (Crouzet et al., 2022).

Cyberattacks. We treat the interaction between the bank and attacker as a strategic contest (Dixit, 1987; Tullock, 2008). At $t = 0$, the attacker exerts a level of effort, $A \geq 0$, at marginal cost, $c > 0$, directed at discovering and exploiting vulnerabilities in the bank's IT systems.[12] A successful

[8] Recent analysis suggests that cyber criminals have amassed considerable wealth in recent years to support their efforts (MIT Technology Review, 2022). State sponsored hackers can also be assumed to have deep pockets.

[9] Including stable sources of funding, such as long-term debt or equity, moderates but does not qualitatively alter the core trade-off in our model. Replacing some short-term debt with stable funding reduces the incidence of runs in the event of a cyberattack, but the bank can still fail if the losses incurred are too large.

[10] We relax this assumption and consider observable and contractible cybersecurity investments in Appendix B and show that our results remain qualitatively unchanged.

[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). They suggest that banks actively managing cyber risks are valued almost 11% higher than those without formal cybersecurity practices.

[12] For parsimony we keep the attacker's cost and prize invariant to bank returns. Allowing them to vary with bank returns mainly rescales the contest, but would not alter the qualitative mechanisms we emphasise.

attack provides 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.

At $t = 1$, the outcome of the contest between the attacker and the bank is determined. With probability

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

the bank is successful in identifying and resolving the vulnerability. We interpret $p(A, S)$ as a measure of the bank's ability to protect itself from cyberattacks. By allocating more to cybersecurity, i.e., S increases, the bank improves its protection against cyberattacks and its chances of winning the contest. With probability $1 - p(A, S)$, the attacker wins the contest and the cyberattack breaches the bank's IT systems. With greater effort, i.e., increasing A, the attacker improves its chances of winning the contest.

If the bank wins the contest, the attacker's payoff is 0, while the bank obtains the equity value, $RI - FD$, at $t = 2$ after investors are paid. But if the attacker wins the contest and discovers the vulnerability, it deploys malicious code that impairs the bank's IT systems and temporarily affects its recourse to liquidity.[13] Specifically, the bank experiences an impairment shock, $\alpha \in [0, 1]$, which is a uniformly distributed random variable drawn at $t = 1$. When $\alpha = 0$, the bank's recourse to liquidity is not affected, while $\alpha = 1$ implies a complete denial of liquidity. The impairment occurs at $t = 1$ and is resolved by $t = 2$. By impairing the bank's recourse to liquidity, a cyberattack can precipitate bank runs (Duffie and Younger, 2019).

Bank failure. 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. Clearly, greater investment, I, at $t = 0$ can prevent the bank from failing. The balance sheet is *resilient* whenever, following a cyberattack, the value of unimpaired assets exceeds the amount needed to service

[13] The cyberattack on the New Zealand stock exchange in December 2020, which prevented the posting of market announcements and led to a trading suspension for several days, is one example (Bloomberg, 2021).

depositor withdrawals. The bank fails and its equity value is wiped out whenever $\alpha > \alpha^{IL}(\ell) \equiv 1 - \frac{\ell FD}{RI}$.[14] Depositors face a zero recovery rate upon bank failure.[15]

Cyberattacks can also have lasting repercussions. For example, there may be loss of secret information pivotal to the bank's role as a financial intermediary, losses from paying ransoms, and even physical damage to IT systems.[16] We capture this possibility by assuming that the bank is subject to a deadweight loss proportional to the shock, α. After the bank regains access to its IT systems at $t = 2$, its investments yield $(1 - \delta\alpha)RI$, where $\delta < 1$ reflects the deadweight loss incurred due to the cyberattack. The larger the deadweight loss, the greater 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 return from assets is insufficient to repay total debt claims. The bank fails whenever $\alpha > \alpha^{IN} \equiv \frac{1}{\delta}\left(1 - \frac{FD}{RD}\right)$, where the threshold is independent of the fraction of withdrawals and increases in the bank's investment, I. If the bank does not fail due to illiquidity or insolvency, it receives $E(\alpha) = (1 - \delta\alpha)RI - FD$ as its equity value. Table 1 illustrates the bank's balance sheet at $t = 2$ following a cyberattack and the rollover of bank debt.

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

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

[14] In many bank run models (Rochet and Vives, 2004; Ahnert et al., 2019) the cost of runs at $t = 1$ is due to firesales or costly asset liquidation. We abstract from these issues to highlight the core tension between ex ante allocations to cybersecurity and ex post resilience to cyberattacks. In our model, the cost of a run at $t = 1$ is the trigger of bankruptcy that completely erodes the value of the bank's assets.

[15] We relax this assumption in Appendix B by allowing for a positive recovery rate and show that our results remain qualitatively unchanged.

[16] The credit downgrading of the Maltese bank, Valetta PLC, following a cyberattack and concerns over its operational risk management, illustrates how cyberattacks can threaten bank solvency (S&P Global Market Intelligence, 2019).

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

Banks are often reluctant to broadcast details of a cyber attack for fear of further attacks and reputational concerns (Biener et al., 2015; Pretty, 2018). So fund managers operate under incomplete information about the level of the bank's impairment following a cyberattack. Specifically, we suppose that fund managers, indexed $k \in [0,1]$, each receive a noisy signal about the impairment shock when making their rollover decision, namely

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

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

Table 2 summarises the timeline 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 2. Timeline of events.

3. Analysis

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

(a) 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^*;

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

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

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

We construct the equilibrium backward, solving first the optimal rollover decision of fund managers before establishing the optimal choice of the attacker and the bank, along with the face value of debt. All proofs are contained in Appendix A.

Rollover risk. For a given mass of early withdrawals, ℓ, the bank does not fail, provided that the impairment shock is sufficiently small. The largest shock that the bank can withstand depends on whether the 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 can service them after the cyberattack provided that the impairment shock is not too large, $\alpha < \alpha^{IN}$. But for larger shocks, as in Region 2,

the losses borne by the bank are so high that it has too few resources to repay claims that are rolled over. So, although the bank can meet interim liquidity needs, the cyberattack renders it insolvent at $t = 2$.

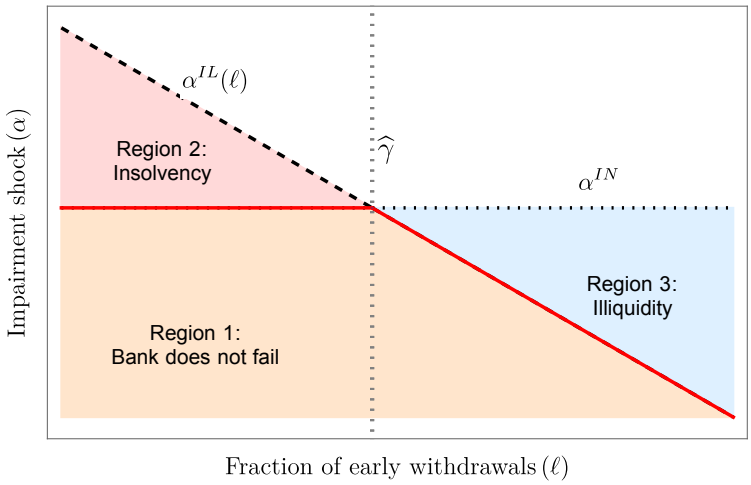

FIGURE 1. Failure conditions for a bank following a successful cyberattack.

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, $\alpha < \alpha^{IN}$. In this case, the bank fails due to illiquidity even though it is solvent.

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.[17]

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 $\alpha > \alpha^*$. The failure threshold increases with greater investment, $\frac{\partial \alpha^*}{\partial I} > 0$.

[17] Our result follows as long as ϵ is sufficiently small. For tractability, we take the limit of vanishing private noise, which is conventional in the global games literature (Morris and Shin, 1998).

Proposition 1 relies on two conditions that ensure well-defined limit dominance regions – a property required for the 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, in the worst case of $\alpha = 1$, if all debt claims are rolled over, $\ell = 0$, then the losses 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, irrespective of the actions of other fund managers. For the lower dominance bound to be well-defined, we make a supposition involving ex ante parameters, namely $RI > FD$. In Lemma 3 below, we show that this inequality is always satisfied in equilibrium. In the lower dominance region, $[0, \alpha^{IL}(1))$, fund managers have a dominance strategy to rollover, irrespective of the actions of other fund managers.

Bank failure is therefore driven by illiquidity when rollover risk is sufficiently large, $\gamma \geq \widehat{\gamma}$. Following an impairment shock, the bank is only able to service debt claims if a few fund managers withdraw. But if sufficiently many 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 that others do too. There is a unique impairment shock, $\alpha^{IL}(\gamma)$, such that the bank fails due to illiquidity whenever $\alpha > \alpha^{IL}(\gamma)$.

When rollover risk is low, $\gamma < \widehat{\gamma}$, bank failure is driven by insolvency. Concerns about the deadweight losses borne by the bank incentivise 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 the *balance sheet resilience* of the bank, i.e., its ability to continue servicing its debts in the event of a cyberattack.

Protection-resilience trade-off. The bank's optimal choice of cybersecurity trades off protection from cyber attacks against balance sheet resilience in the event of a cyberattack. By investing more in cybersecurity, the bank improves its chances winning the contest against the attacker. But this comes at the cost of reducing profitable investment, which has two implications. First, irrespective of the outcome of the contest, the bank has a lower equity value. And second, in the event of an

attack, the bank has fewer resources at its disposal to service withdrawals, and, therefore, less able to weather a run or buffer itself against the deadweight losses.

At $t = 0$, the attacker chooses how much effort to expend to uncover vulnerabilities to maximise its expected prize, taking as given the bank's level of cybersecurity. At the time of choosing the effort, the attacker does not know what vulnerabilities it will find and how disruptive their exploitation might be to the bank. The attacker's problem is

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

Lemma 2. *The attacker expends effort*

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

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

Lemma 2 is intuitive. The attacker expends a strictly positive effort as long as the prize is sufficiently valuable. The bigger the prize, the larger is equilibrium effort. And, as the marginal cost of effort increases, the potential gain from effort for a less sophisticated attacker is lower and there is less effort.

Consistent with the notion that banks commit to their IT budgets at the start of the financial year, we treat the bank as a first-mover in the contest with the attacker and suppose that it pre-commits to its cybersecurity choice in the spirit of Dixit (1987) and Tullock (2008). If $A^*(S) > 0$, the probability that the bank successfully protects itself from cyberattacks 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 equity value is given by

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

which combines outcomes with, and without, a successful cyberattack. When the bank successfully protects itself (probability $p(S)$), investors are repaid in full and the bank retains the residual value $RI - FD$. When the attack succeeds (probability $1 - p(S)$), equity value depends on the realised impairment shock, α, and whether the losses trigger bank failure, summarised by the critical threshold $\alpha^* = \alpha^*(D,I)$.

Equation (10) highlights the risk-management role that cybersecurity investment, S, plays in shifting probability mass toward the safe, no-attack state. When S increases, the probability of the bank avoiding a cyberattack, and retaining $RI - FD$, rises, while the probability of being subject to a cyberattack and suffering a loss to its equity value falls. This re-weighting of payoffs across states directly parallels the hedging motive in Froot et al. (1993): by investing in protection, the bank reduces the dispersion of outcomes (a benefit) but at the cost of lowering expected returns, since higher S diverts resources from productive investment. If the failure threshold α^* were exogenous, so that balance-sheet resilience did not vary with S, the bank's problem would exhibit the classical mean–variance logic: cybersecurity spending reduces the returns spread with and without a cyberattack (variance-like risk) while lowering the overall mean return by diverting resources from investment.[18]

In our setting, however, the failure threshold is endogenous: allocating more to protection reduces resilience by lowering the resources available to withstand shocks, i.e., α^* decreases. This additional channel alters the classic risk-management logic. The bank now internalises not only how S tilts probability mass between attack and no-attack states (the *protection* margin) but also how it reshapes losses within the attack state through its effect on α^* (the *resilience* margin). As

[18]Formally, with an exogenous α^* that is small enough for the bank not to default, expected equity can be written as $\pi(S) = R(1-S) - FD - \frac{1}{2}\Big(1 - p(S)\Big)\delta(\alpha^*)^2 R(1-S)$. The first term, $R(1-S) - FD$, corresponds to the expected return in the event of no cyberattack, and it declines with S (the cost of protection). The second term represents the difference in the equity values with and without a cyberattack: it scales with the probability of a cyberattack, $1 - p(S)$, and the expected loss, $\delta(\alpha^*)^2 R(1-S)/2$, making it variance-like. Increasing S lowers this difference term (benefit) but reduces the expected return (cost).

a result, cybersecurity investment acts both as a hedge and a determinant of the bank's capacity to absorb shocks.

The optimisation problem for the bank is thus

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

where 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.

The net marginal benefit to the bank from issuing debt is strictly positive, $\frac{\partial \pi}{\partial D} > 0$, even after accounting for the potential losses in the event of a cyberattack. This is the bank's intermediation margin, i.e., the difference between what it earns from investing the debt raised and the additional repayment to investors. If $R - F > 0$, the intermediation margin is strictly positive and the bank is encouraged to issue as much debt as possible.

In what follows, we exploit that $D^* = 1$ to simplify our analysis. An immediate consequence is that, due to limited liability, the participation constraint is satisfied for any I and S. The result in Lemma 3 immediately follows.

Lemma 3. *It is never optimal for the bank to invest more than* $\overline{S} \equiv 1 - \frac{F}{R}$ *in cybersecurity, thus implying that, in equilibrium,* $RI \geq F$.

The bank's profit in the event there is no cyberattack is $RI - F = R(1 - S) - F$. Allocations to cybersecurity greater than $\overline{S}$ cause the bank to have a negative equity value and default, even in the absence of a cyberattack. And in the event of a cyberattack, the bank also always defaults. Thus,

the bank's participation constraint is violated for any $S > \overline{S}$. Lemma 3 ensures that $RI > F$, 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*

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

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., $\frac{\partial S^*}{\partial F} > 0$.

Equation (11) formalises the *protection-resilience* trade-off for the bank. The left-hand side is the ratio of the marginal impact on expected profit from an increase in protection (numerator), and the marginal impact from an increase in investment enabling the balance sheet to remain resilient (denominator). The right-hand side is the marginal rate of transformation between the bank's allocation 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.

This first-order condition highlights that the optimal S^* equates the marginal benefit from improved protection—the expected reduction in losses and failure risk—with the marginal cost of reduced investment and resilience. By increasing S, the bank reallocates probability mass towards the no-attack outcome and away from the states in which it suffers an attack and looses equity value. The cost of doing so, however, is a reduction in resilience: diverting resources from investment lowers average returns and leaves the bank more susceptible to failure in the event of an attack.

The protection-resilience trade-off depends on the nature of bank failure. When rollover risk is low, $\gamma < \widehat{\gamma}$, bank failure is driven by insolvency concerns, i.e., $\alpha^* = \alpha^{IN}$. Since fund managers are not subject to coordination failure, their rollover decisions are efficient and, at the critical failure threshold, the bank's equity value is wiped out, i.e., $E(\alpha^*) = 0$. But when rollover risk is high,

$\gamma \geq \widehat{\gamma}$, illiquidity concerns are paramount and fund managers' rollover decisions are subject to coordination failure. Fund managers may withdraw because they believe others will withdraw and so the bank fails due to a lack of liquidity despite being solvent. The run is inefficient, which is reflected by the failure threshold $\alpha^* = \alpha^{IL}(\gamma)$ where the bank has strictly positive equity value, $E(\alpha^*) > 0$. The term $E(\alpha^*) > 0$ captures the amount of equity value that is lost when the bank is subject to coordination failure induced runs. Avoiding these additional losses creates an additional effect that increases the marginal impact of investing more in cybersecurity so as to reduce the likelihood that the bank is subject to cyberattack induced runs. Therefore, the bank optimally invests more in cybersecurity when its failure is driven by illiquidity than when it is due to insolvency.

The effect of a higher face value of debt on cybersecurity investment can be understood by decomposing the cross-derivative $\frac{\partial^2 \pi}{\partial S \partial F}$. In the case where bank failure is driven by insolvency, we obtain

$$-\frac{\partial p}{\partial S} + \alpha^* \frac{\partial p}{\partial S} - (1 - p(S))R(1 - \delta\alpha^*)\frac{\partial \alpha^*}{\partial F}.$$

The first term reflects how the marginal benefit of protection changes as the face value of debt increases in the event that the bank successfully protects itself. As the face value of debt increases, the equity value that accrues to the bank decreases. This, in turn, weakens the bank's incentive to invest in its protection—the standard *risk-shifting effect*.

The second term captures how a change in F alters the marginal benefit of protection in the event of a cyberattack. Conditional on survival ($\alpha < \alpha^*$), a higher F reduces the bank's equity value and increases the losses borne by the bank. This strengthens the incentive to avoid the attack altogether by increasing protection. Finally, the third term reflects how a higher F affects the marginal value of protection through its impact on the failure threshold. An increase in F lowers α^*, making the bank less resilient and more likely to fail for smaller shocks. This amplifies the benefit of protection, since avoiding failure now preserves more equity value.

In sum, the incentives to improve protection dominate dominate the classical risk-shifting effect, implying that $\frac{\partial^2 \pi}{\partial S \partial F} > 0$ and, by the implicit-function theorem, investment in cybersecurity is increasing in the face value of debt, $\frac{\partial S^*}{\partial F} > 0$. This reversal of the standard risk-shifting result

(e.g., Hellmann et al., 2000; Repullo, 2004; Dell'Ariccia et al., 2014) also holds when failure is illiquidity-driven, $\gamma \geq \widehat{\gamma}$, since the effect via the failure threshold is even stronger, thus amplifying the incentive to invest in protection.

Unique equilibrium. Since investors do not observe the bank's allocation towards cybersecurity, they set the face value of debt based on their expectation of the bank's contribution. Investors are repaid in full when the bank does not fail. This occurs if the bank is successful in protecting itself or remaining resilient through a successful attack. 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^*, is such that investors are indifferent between lending to the bank and investing in the storage technology, $\mathcal{V}(F^*) = r$.

Lemma 4. *The face value of debt, $F^*(S)$, is U-shaped in the bank's cybersecurity allocation, reaching the minimum at $S = \tilde{S}$. The minimum $\tilde{S}$ is weakly increasing in rollover risk, γ.*

When investors anticipate low cybersecurity, $S < \tilde{S}$, improved protection is viewed as the prudent course of action that increases the probability of repayment. So the face value of debt decreases in cybersecurity, $\frac{\partial F^*}{\partial S} < 0$. But when cybersecurity is considered high, $S \geq \tilde{S}$, any further increase in protection is considered counterproductive to their likelihood of being repaid and they require greater compensation for depositing with the bank, $\frac{\partial F^*}{\partial S} \geq 0$. The turning-point $\tilde{S}$ is also influenced by rollover risk. When failure is driven by illiquidity ($\gamma > \widehat{\gamma}$), fund managers have more incentive to withdraw in the event of a cyberattack. Under such circumstances, investors regard greater protection against cyberattacks to be more prudent, and so the $\tilde{S}$ increases.

Proposition 3. *There exists a unique equilibrium, jointly characterised by $\mathcal{V}(F^{**}, S^{**}) = r$, where the bank's cybersecurity choice is $S^{**} \equiv S^*(F^{**})$. The rollover risk threshold $\widehat{\gamma}^{**} \equiv \widehat{\gamma}(F^{**}, S^{**})$ that distinguishes whether bank failure is due to illiquidity or insolvency exists and is unique.*

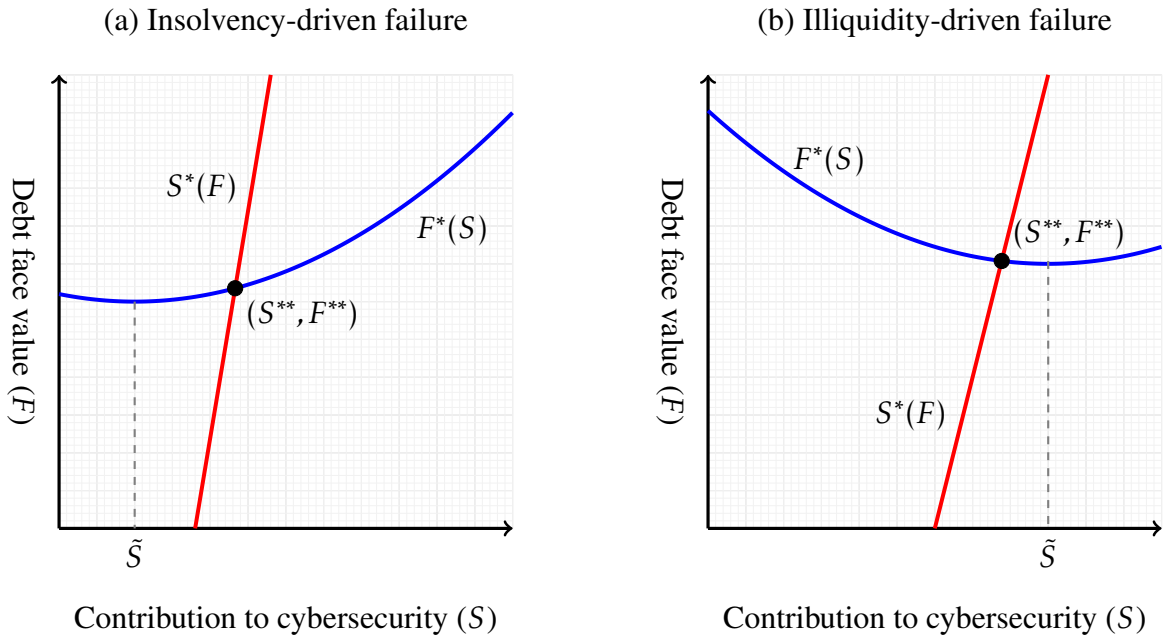

FIGURE 2. Unique equilibrium and the nature of bank fragility.

Figure 2 (panels a and b) depicts the unique equilibrium when bank failure is driven by insolvency and illiquidity, respectively. In panel (a), in the vicinity of the equilibrium, both $F^*(S)$ and $S^*(F)$ are increasing functions. Anticipating that the bank chooses more cybersecurity to guard against cyberattacks, investors expect that, in the event of an attack, the loss of balance sheet resilience will increase the ex ante likelihood of bank failure. So they demand greater compensation. In turn, the bank faces a higher face value of debt and, thereby, a lower opportunity cost to improve its protection. So it increases cybersecurity even further. Through this iterative process, a unique equilibrium emerges in which the behaviour between investors and the bank is characterised by strategic complementarities.

In panel (b), bank failure is driven by illiquidity and the $F^*(S)$ schedule is downward sloping in the vicinity of the unique equilibrium. In this case, the iterative process and unique fixed point arises from strategic substitutability in the behaviour of investors and the bank. When investors perceive that the bank will increase its cybersecurity, they consider the net effect to increase their likelihood of repayment. They accordingly require a lower face value of debt as compensation. But since the bank is less likely to fail in the event of a cyberattack, this increases the opportunity cost of strengthening protection. So the bank is incentivised to reduce its contribution to cybersecurity.

It follows from Proposition 2 that both S^{**} and F^{**} are higher when bank failure is driven by illiquidity than when failure is driven by insolvency.

An interesting implication for the illiquidity regime is that the equilibrium exists in a region where the usual risk-return trade-off is absent. Along the locus for the face value of debt, we have that

$$\left.\frac{d\pi}{dS}\right|_{(S^{**},F^{**})} = -\frac{r}{F^{**}} \left.\frac{dF^*(S)}{dS}\right|_{S^{**}} > 0,$$

so a downward-sloping $F^*(S)$ schedule at the equilibrium implies that a marginal increase in the bank's cybersecurity investment raises expected equity value once debt is priced. The intuition is that when debt is priced competitively, higher cybersecurity investment does not merely increase protection—it also reduces the bank's debt burden. By reducing the likelihood of bank failure in the event of a cyberattack, investors regard lending to the bank to be safer, and, hence seek a lower face value of debt. This reduces the bank's debt burden and raises its equity value. In equilibrium, more cybersecurity simultaneously reduces failure risk and increases expected returns.

4. Cyber threat landscape

The cyber threat landscape refers to the threats and risks that banks face in the digital environment. It encompasses, for example, the potential attackers and their experience in launching attacks, as well as the economic costs of cyberattacks. In our model, the cyber threat landscape is summarised by attacker sophistication (c), the deadweight loss term (δ), and the nature of bank fragility (γ). We consider how changes to these parameters shape the protection-resilience trade-off. Proposition 4 and Figure 3 summarise these comparative static results.

The cyber-threat landscape is also shaped by policy initiatives to deal with cyber risks. In what follows, we consider the impact of a recent proposal to have so-called *emergency payment nodes*, which are narrow utility banks that become active during periods of operational stress to offer liquidity assistance and facilitate payments (Duffie and Younger, 2019). Although not entirely immune from cyberattacks, they are likely to be less susceptible owing to their narrow function.

Proposition 4. *The bank's investment in cybersecurity, S^{**}, is decreasing in the sophistication of the attacker, and increasing in rollover risk γ, and the deadweight loss of attack, δ.*

Attacker sophistication. If the attacker is less sophisticated, i.e., c is higher, it is more costly for the attacker to invest effort. Ceteris paribus, reduced attacker effort increases the likelihood of the bank winning the contest. Since each unit of protection now translates more effectively into a higher probability of avoiding an attack, the bank gains more from tilting probability mass toward the no-attack outcome. As a result, the marginal return to protection rises and the $S^*(F)$ schedule shifts outward. At the same time, from the investors' perspective, there is a higher likelihood of being repaid, implying that lending to the bank is marginally safer. So they require a lower face value of debt, shifting the $F^*(S)$ schedule downward. In general, the direct effect via the contest strictly dominates the indirect effect, $\frac{\partial S^*}{\partial c} + \frac{\partial S^*}{\partial F}\frac{\partial F^*}{\partial c} > 0$. As a consequence, in equilibrium the bank's protection against cyberattack increases, while the effect on balance sheet resilience is ambiguous.

Rollover risk. When rollover risk is low, $\gamma < \widehat{\gamma}^{**}$, bank failure is insolvency driven. In this case, the bank's incentives to protect itself is unresponsive to increases in rollover risk since the failure threshold, α^{IN}, is not driven by fund manager withdrawals.

But if rollover risk exceeds the critical threshold, $\gamma > \widehat{\gamma}^{**}$, the bank must contend with the possibility of a self-fulfilling run. As rollover risk increases, α^* declines, implying that smaller shocks are sufficient to trigger failure. A cyberattack, therefore, becomes a more potent catalyst for inefficient, coordination-driven runs that destroy equity value even when the bank remains fundamentally solvent. Thus, a higher rollover risk increases the benefit from greater cybersecurity investment because stronger protection mitigates both direct losses and lowers the probability of attack-induced runs. In sum, the $S^*(F)$ schedule shifts outward.

From an investor's perspective, a marginal increase in rollover risk reduces balance sheet resilience and the likelihood of repayment in the event of a cyberattack. To be compensated for this higher risk, they demand greater compensation and seek a higher face value of debt. The $F^*(S)$ schedule also shifts outwards, reinforcing the bank's incentives to spend more on cybersecurity.

Deadweight loss. An increase in the deadweight loss, δ, can be similarly decomposed. A rise in δ means that the losses borne by the bank in the event of a cyberattack are greater, which reduces its equity value and increases the wedge between the outcomes with and without a breach.[19] An increase in cybersecurity investment reduces the wedge and, therefore, strengthens the bank's motive to invest in protection. For investors, an increase in the deadweight loss reduces the likelihood of repayment when bank failure is insolvency driven. They require greater compensation and the face value of debt increases, further reinforcing the bank's incentives to increase cybersecurity. As Figure 3(c) illustrates, both $S^*(F)$ and $F^(S)$ shift outwards.

Emergency payment node (EPN). Our model sheds light on how an EPN can improve a bank's resilience in the event of a cyberattack, and its implications for the protection-resilience tradeoff. We consider the liquidity assistance offered by the EPN as a bridge loan at $t = 1$ to cover withdrawals, ℓFD. Normalising the fee charged to zero, the EPN can perfectly eliminate panic runs by fund managers and so the bank only fails at $t = 2$ due to insolvency.

But suppose that, in the event of a cyberattack, the EPN can also become impaired and unable to facilitate payments. We capture such considerations by assuming that the EPN is active at $t = 1$ with probability β. The parameter β can be interpreted as a measure of the robustness of the EPN in the event of a cyberattack.

Proposition 5. *Suppose $\gamma > \widehat{\gamma}^{**}$. A marginal increase in β decreases the bank's contribution to cybersecurity, S^{**}, and the face value of debt, F^{**}. The balance sheet resilience of the bank improves as a result.*

In the event of a cyberattack, borrowing from the EPN eliminates bank runs induced by coordination failure. The bank only fails when it is fundamentally insolvent. But if the EPN is itself inoperable, the bank is subject to a self-fulfilling run. As the robustness of the EPN increases (β increases), the bank places greater emphasis on its ability to facilitate payments in the event of a cyberattack. This weakens the bank's incentives to invest in its protection since a more robust EPN

[19] The failure threshold is strictly decreasing in δ when bank failure is driven by insolvency, i.e., $\gamma < \widehat{\gamma}^{**}$.

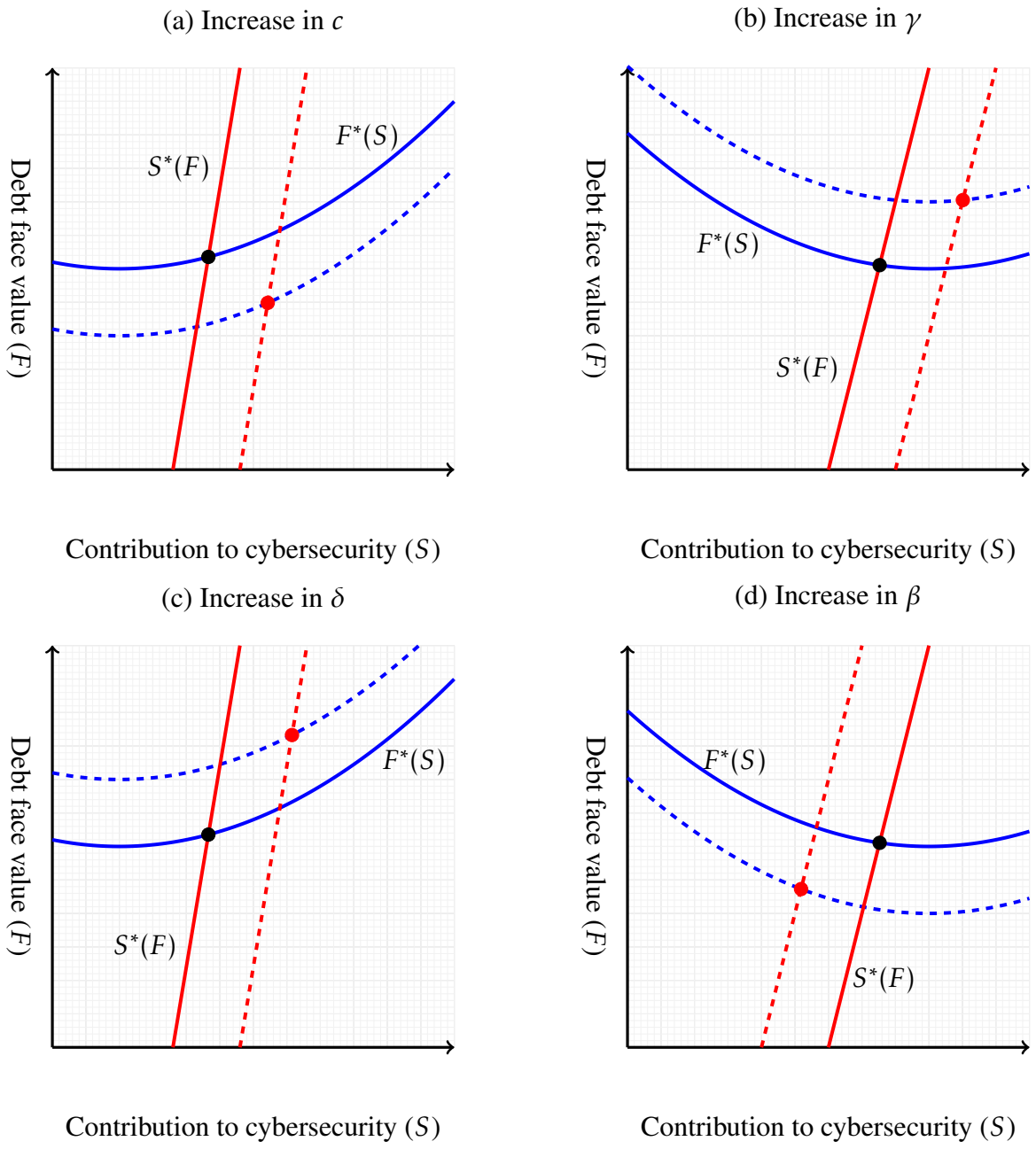

FIGURE 3. Implications of a changing cyber threat landscape for cybersecurity provision.

cushions the downside risk of a successful cyberattack. As the marginal benefit of great protection is reduced, the $S^*(F)$ schedule shifts inwards as Figure 3(d) shows. As a consequence of lower cybersecurity and cheaper debt, balance sheet resilience improves unambiguously.

To serve as an effective backstop and to avoid being subject to the same vulnerabilities, the EPN must use robust, and distinct, infrastructure from the banks that it services. While this may be challenging due to the specialised and concentrated nature of much of the IT service industry at present, an advantage of the narrow function and highly restricted access points of the EPN is that it can be made resistant to a breach relatively more easily than an ordinary bank (Duffie and Younger, 2019). But an implication of the bank's and investors' optimal responses to the EPN is that protection is reduced and attacks are more likely to succeed. The bank will be subject to more frequent disruption as a result, and our analysis shows that the mere existence of an EPN makes it more likely to be deployed.

5. Financial stability implications

Bank failures are socially costly. There can be a loss of payment services when a bank is in financial distress, and a sharp credit contraction can create significant macroeconomic costs. In what follows, we treat the social costs of bank failure as exogenous and represent them with the parameter $\lambda > 0$. Although the bank does not take the social cost of its failure into account when making decisions, a social planner takes these into consideration.

Social planner. The planner takes as given the information structure of the game, i.e., fund managers' incomplete information about the shock and the bank's unobservable investment choice) and the the planner chooses a contribution to cybersecurity for a given face value of debt, i.e.,

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

The term that multiplies into λ is the ex ante probability that the bank fails due to a cyberattack. The face value of debt continues to be set via the investors' participation constraint. With a slight abuse of notation, denote S^P as the cybersecurity choice of the planner in equilibrium.

Proposition 6. *The private equilibrium is constrained inefficient. When bank failure is driven by insolvency, $\gamma < \widehat{\gamma}^{**}$, the bank overinvests in cybersecurity, $S^{**} > S^P$, and the face value of debt*

is too high, $F^{**} > F^P$. When failure is driven by illiquidity, $\gamma > \widehat{\gamma}^{**}$, the bank underinvests in cybersecurity, $S^{**} < S^P$, and the face value of debt is too high, $F^{**} > F^P$.

From the planner's perspective, a marginal increase in cybersecurity investment can be decomposed into the marginal benefit of increasing protection to stop cyberattacks, $\lambda \frac{\partial p}{\partial S}(1 - \alpha^*(S))$, and the marginal cost, which reflects how loss of balance sheet resilience increases the likelihood of bank failure in the event of a cyberattack, $\lambda(1 - p(S))\frac{\partial \alpha^*}{\partial S}$. As Proposition 6 suggests, the relative weight of these effects depends on the extent of rollover risk or, equivalently, whether bank runs are fundamentals-based or driven by self-fulfilling investor beliefs.

Ceterus paribus, the bank is more likely to fail when rollover risk is high, i.e., $\alpha^{IL}(\gamma) < \alpha^{IN}$, whenever $\gamma > \widehat{\gamma}$. Intuitively, a higher proclivity for runs exacerbates fragility. This implies that the marginal benefit of increasing protection is higher, and also the marginal cost is lower when bank failure is driven by illiquidity. In other words, faced with the prospect of runs in the event of a cyberattack, the planner is incentivised to reduce the likelihood that the bank is subject to them by enhancing its protection against cyberattacks. For $\gamma > \widehat{\gamma}^{**}$, the planner allocates more to cybersecurity than the bank $S^P > S^{**}$, leading to a lower face value of debt, $F^P < F^{**}$.

But when failure is driven by insolvency $\gamma < \widehat{\gamma}^{**}$. the probability of bank failure is lower, $\alpha^{IN} > \alpha^{IL}(\gamma)$, and the planner's incentive to increase protection against cyberattacks is reduced. At the same time, the planner faces a higher marginal cost for providing protection. Intuitively, when the risk of bank failure is relatively low, it is imprudent to divert resources away from maintaining a healthy and resilient balance sheet. So the planner chooses $S^P < S^{**}$, which is associated with a lower face value of debt, $F^P < F^{**}$ (see Figure 4). Crucially, the planner's solution, S^P is always closer to the minimum, $S = \tilde{S}$, of the investors' pricing schedule, $F^*(S)$, than the bank's private choice.

The idea that the bank can overinvest in cybersecurity reflects the underlying reality that cyberattacks are a matter of "when" not "if". Faced with potential failure, the bank, concerned primarily with its own survival, is incentivised to allocate resources to find and mitigate vulnerabilities at the expense of balance sheet resilience. Resources are allocated towards measures that maximise

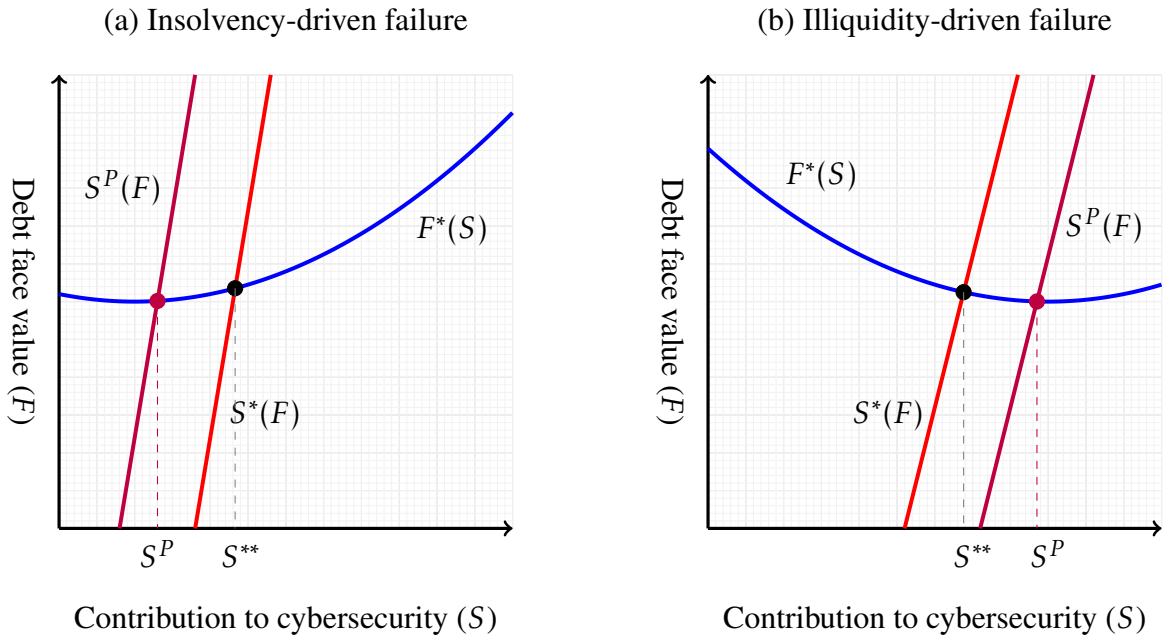

FIGURE 4. Comparison between the planner's equilibrium and the bank's equilibrium.

expected equity value rather than remaining resilient in the event of a cyberattack. From the social planner's perspective, this creates inefficiency because resources diverted to cybersecurity could otherwise strengthen the bank's ability to maintain critical financial services in the event of an attack, thereby reducing the social costs associated with bank failure.

Cyber subsidies. Our analysis highlights the need for a nuanced approach to regulating cybersecurity. Under some circumstances, the bank should be incentivised to increase its investment in cybersecurity so as to increase its protection against cyberattacks. While in other circumstances, emphasis should be placed on improving balance sheet resilience so that the bank is better able to cope in the event of a cyberattack. Proposition 7 considers the role of Pigovian policies aimed at either subsidising cybersecurity investment or balance sheet resilience in implementing the planner's constrained efficient solution.

Proposition 7. *When failure is insolvency driven, $\gamma < \widehat{\gamma}^{**}$, the planner can achieve the constrained efficient solution, (S^P, F^P) via a contingent linear tax on cybersecurity investment at $t = 2$, where*

the optimal tax rate is

$$\tau^* = \frac{\lambda\left[\frac{\partial p}{\partial S}\Big|_{S=S^P}(1-\alpha^*(S^P,F^P)) + (1-p(S^P))\frac{\partial \alpha^*}{\partial S}\Big|_{S=S^P,F=F^P}\right]}{p(S^P)+(1-p(S^P))\alpha^*(S^P,F^P)}, \tag{13}$$

combined with a lump-sum rebate of the generated revenue, $\Sigma \equiv \tau^* S$. *While, when failure is illiquidity driven,* $\gamma \geq \widehat{\gamma}^{**}$, *the constrained efficient solution is achieved via a subsidy,*

$$\sigma^* = \frac{\lambda\left[\frac{\partial p}{\partial S}\Big|_{S=S^P}(1-\alpha^*(S^P,F^P)) + (1-p(S^P))\frac{\partial \alpha^*}{\partial S}\Big|_{S=S^P,F=F^P}\right]}{p(S^P)+(1-p(S^P))\alpha^*(S^P,F^P)}, \tag{14}$$

on cybersecurity investment at $t = 2$, *which is funded via lump-sum taxation,* $T \equiv \sigma^* S$.

Figure 5 illustrates how taxes and subsidies achieve the constrained efficient outcome. In the case of insolvency-driven bank failure, the policy is given by a linear tax on cybersecurity investment at $t = 2$ combined with a lump-sum rebate. Since the bank overinvests in cybersecurity, the tax makes it more costly for the bank to forgo maintaining a healthy and resilient balance sheet in favour of increasing its protection. Moreover, as the tax is revenue neutral and rebated back to the bank, it does not directly impact the failure threshold and hence the face value of debt. When failure is illiquidity driven, a subsidy may be introduced to correct for the underinvestment in cybersecurity. The marginal subsidy rate is set to equal the marginal social value of increased protection, encouraging the bank to align its allocation with that of the planner. To fund the policy, a lump-sum tax is levied on the bank that is equal to the subsidy provided. Examples include the Dutch Central Bank's *subsidie cyber-weerbarheid* program and the Monetary Authority of Singapore's *Cybersecurity Capability Grant* program.

In many jurisdictions, regulators offer banks more direct subsidies in the form of cybersecurity assistance. The European Central Bank has established the Threat Intelligence-based Ethical Red Teaming for the European Union (TIBER-EU), the Monetary Authority of Singapore conducts the Adversarial Attack Simulation Exercise (AASE), and the Australian authorities draw on the Cyber Operational Resilience Intelligence led Exercise (CORIE).[20] Assistance takes the form of ethical

[20]These frameworks, in turn, draw inspiration from the Bank of England's Cybersecurity Benchmarking and Threat Intelligence (CBEST) toolkit.

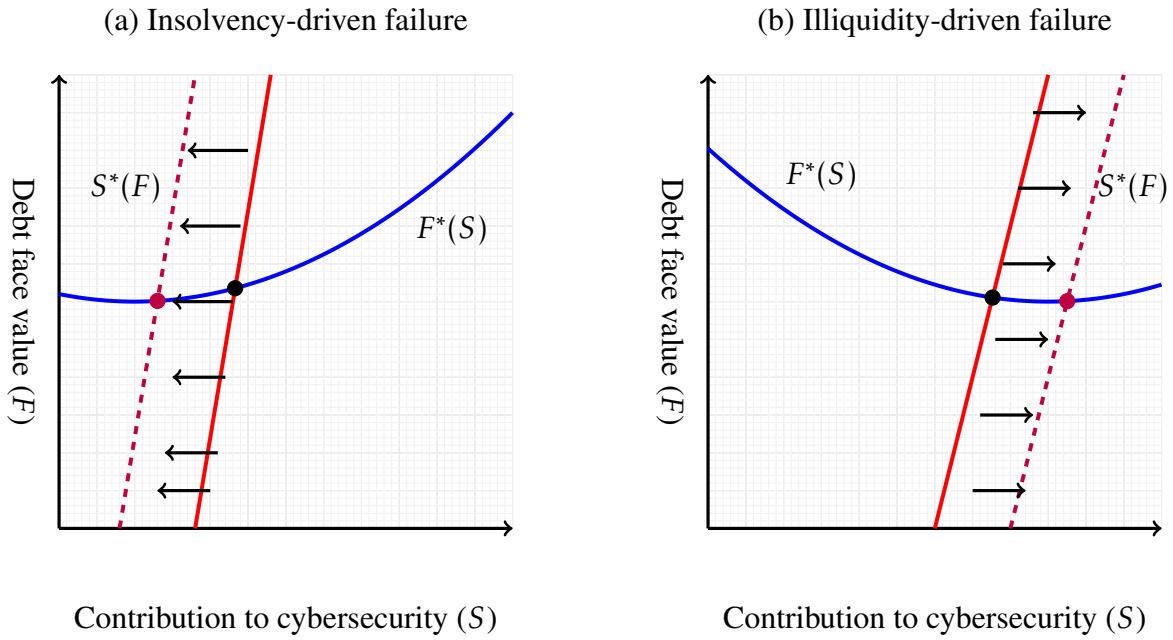

FIGURE 5. Introducing revenue-neutral Pigovian policies at $t = 2$ can achieve constrained efficiency. When failure is insolvency-driven a tax with a rebate is the optimal policy, while when failure is illiquidity driven, the optimal policy is a subsidy that is taxed lump-sum.

hackers (the 'red team') mimicking real-world attackers by breaching banks' IT systems, while the banks (the 'blue team') try to thwart them. The idea is to improve banks' understanding of the cyber threat landscape and their ability to guard against attack.

We model red team testing as a transfer of knowledge from ethical hackers to banks — technical assistance improves the quality of human capital (Prenio et al., 2019). Technical assistance is a constant returns to scale technology that increases the bank's cybersecurity by factor ρ, which captures the extent of knowledge transfer. The probability that the bank wins the contest against the attacker and is not subject to a cyberattack at $t = 1$ is

$$p(S) = \sqrt{\frac{c\,S(1+\rho)}{V}},$$

which is increasing in knowledge transfer, ρ, from red-team testing.

Proposition 8. *A marginal increase in ρ leads to an increase in the bank's investment in cybersecurity, S^{**}, and a decrease in the face value of debt, F^{**}. Thus, as a policy instrument, ρ cannot be used to obtain the planner's constrained-efficient solution, (S^P, F^P).*

Proposition 8 shows that red teaming incentivises the bank to allocate more resources to cybersecurity. By improving the bank's cybersecurity, a marginal increase in ρ increases the likelihood of winning the contest and so encourages the bank to invest more in cybersecurity. The increase in ρ also improves the likelihood of investors being repaid in full, inducing a reduction in the face value of debt sought in compensation for lending to the bank, which in turn, discourages cybersecurity investment by the bank. In equilibrium, the direct effect dominates the indirect effect, implying an increase in cybersecurity investment. Crucially, while red-team testing incentivises greater investment in cybersecurity, the indirect effect via the face value of debt implies that it cannot be used to implement the planner's constrained efficient solution.

The indirect effect of red-team testing on lowering the face value of debt implies that, in the event of a cyberattack, the bank is better able to service withdrawals. Thus, the bank's balance sheet resilience improves as a consequence. Red teaming is, thus, more than just a technical IT exercise to improve protection. By boosting investor confidence, it reduces the face value of debt and improves the resilience of the bank.

6. Extension - public good aspects of cybersecurity

Our analysis focuses on the protection-resilience tradeoff for a single bank. At the system level, cybersecurity has the hallmarks of a public good (Mester, 2019).[21] In practice, many banks use common IT systems provided by third-party vendors to avoid the costs of owning and maintaining inhouse services. As mentioned in the introduction, ICBC used a common software platform "Citrix". And Oracle Corporation sells a suite of software solutions (FLEXCUBE) to banks such

[21]The (state-sponsored) cyberattack on SolarWinds, a third-party vendor to many U.S. banks, highlights the public good nature of shared IT solutions. Banks that installed a SolarWinds software update became compromised, opening a back door for attackers to gain privileged access to the banks' IT systems for some time (FBI, 2021).

as HSBC, Citigroup and Wells Fargo. The use of common IT systems creates interdependencies among banks and generates free-riding problems that aggravate underinvestment in cybersecurity.

We extend our setting to allow for $N \geq 2$ banks sharing a common IT system. Each bank runs the same software suite or uses a common hardware component provided by a third party vendor. We use $X \equiv X(\vec{S})$ to denote the level of cybersecurity in the banking system, given the contributions of individual banks, $\vec{S} = (S_1, \ldots, S_N)$.

In the game between the attacker and the banks, and given the attacker's effort, A, banks are successful in uncovering vulnerabilities and mitigating the attack with probability

$$p(A, X(\vec{S})) = \frac{X(\vec{S})}{A + X(\vec{S})}. \tag{15}$$

But with probability $1 - p(A, X(\vec{S}))$, the attack is successful and a cyberattack is launched on *all* banks, causing a common level of disruption $\alpha \in [0, 1]$ that is uniformly distributed. The prize earned by the attacker remains $V > 0$.

Assuming no overlap in sets of fund managers across the different banks and that, in the event of attack, fund managers receive noisy signals about the disruption to their bank, we can extend the result of Proposition 2 to define the failure threshold α_b^* for each bank, $b = 1, \ldots, N$. If a fraction of fund managers affiliated with bank $b = 1, \ldots, N$ withdraw, then the bank fails due to illiquidity at $t = 1$ whenever $\alpha > \alpha_b^{IL}(\ell_b)$. And, if bank b is able to service its debts at $t = 1$, it may fail due to insolvency at $t = 2$ whenever $\alpha > \alpha_b^{IN}$. Since the face value of debt is determined by both the level of protection at the system-level, and the bank's (private) balance sheet resilience, there is a schedule, $F_b^*(S_b)$, specific to each bank.

We capture heterogeneity in expected equity value at the bank level in the measure of deadweight losses, δ_b, incurred by banks to reflect their differential exposure to lasting impact (physical or reputational) from a cyberattack. Let banks be placed in descending order of measure δ_b so that $\delta_1 > \ldots > \delta_N$. We assume that $\delta_b > 1 - \frac{1}{R}$ for all $b = 1, \ldots, N$ so that a unique failure threshold arises as in Proposition 1. Lemma 1 establishes that $\widehat{\gamma}$ is proportional to the deadweight loss from

a cyberattack. It follows that there is a regime threshold, $\widehat{\gamma}_b(\delta_b)$, for each bank such that the bank fails due to insolvency whenever $\gamma < \widehat{\gamma}_b$ and due to illiquidity otherwise, with $\widehat{\gamma}_1 > \ldots > \widehat{\gamma}_N$.

Best shot public good. When contributions to protection by banks are perfect substitutes, cybersecurity is a *best-shot* public good (Hirshleifer, 1983; Bliss and Nalebuff, 1984), and so aggregate cybersecurity takes the form $X(\vec{S}) = \max\{S_1, \ldots, S_N\}$. Public disclosure of software vulnerabilities to the benefit of other banks using similar systems is one example of how one bank can become a best-shot provider of the public good of cybersecurity.[22] We can express bank b's expected equity value as a function of investment, I_b, and the level of protection, p, so that the bank's $t = 0$ optimisation problem is

$$\begin{aligned} \max_{\{I_b, X, S_b\}} \quad & \pi_b(I_b, p) \\ \text{subject to} \quad & 1 = I_b + S_b, \\ & p = \sqrt{\frac{cX}{V}}, \\ \text{and} \quad & X = \max\{S_1, \ldots, S_N\}. \end{aligned}$$

In the private equilibrium, accounting for the incentives of all banks and how cybersecurity investment affects the face value of debt, bank b contributes $S_b^*(F_b^*)$ analogously to S^* derived in equation (11) in the single-bank case. But with cybersecurity assuming a best-shot formulation, the marginal contribution by bank b to aggregate cybersecurity, $\partial X/\partial S_b$, is equal to either 1 or zero, depending on whether b is the largest contributor. By our ordering, bank 1 has the largest marginal rate of substitution of resilience for protection and so, in equilibrium, while accounting for the endogenous face value of debt, $F_1^*(S_1)$, bank 1's contribution is given by $S_1^{**} \equiv S_1^*(F_1^{**}) > 0$ while all other banks contribute $S_2^{**} = \ldots = S_N^{**} = 0$.

The social planner maximises the sum of banks' expected equity values minus the social cost of bank failure, which we assume is linearly separable in individual bank failure. Unlike banks, which maximise their own equity value taking others' cybersecurity contributions as given, the

[22] 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.

planner considers the social benefits and costs of each bank's allocation, the social cost of bank failure, as well as the bank's own balance sheet conditions, taking into account how cybersecurity contributions influence the face value of debt. Formally,

$$\begin{aligned}
\max_{\{\{I_b\},X,\{S_b\}\}} \quad & \sum_{b=1}^{N} \pi_b(I_b,p) - \lambda\Big(1-p\Big)\sum_{b=1}^{N}\Big(1-\alpha_b^*(I_b)\Big) \\
\text{subject to} \quad & 1 = I_b + S_b \quad \forall\, b = 1,...,N\,, \\
& p = \sqrt{\frac{cX}{V}}\,, \\
\text{and} \quad & X(\vec{S}) = \max\{S_1,\ldots,S_N\}\,.
\end{aligned}$$

Accounting for the endogenous face value of debt, in the planner's equilibrium, we have that $S_2^P = \ldots = S_N^P = 0$, while for bank 1, $S_1^P > 0$ so that

$$\frac{\frac{\partial \pi_1}{\partial X} + \lambda\frac{p}{2X}(1-\alpha_1^*) + \sum_{j=2}^{N}\frac{\partial \pi_j}{\partial X} + \lambda\frac{p}{2X}(1-\alpha_j^*)}{\frac{\partial \pi_1}{\partial I_1} + \lambda(1-p)\frac{\partial \alpha_1^*}{\partial I_1}} = 1\,. \tag{16}$$

Equation (16) is a version of the 'Samuelson Rule' (Samuelson, 1954). The left-hand side reflects the sum of marginal rates of substitution at the bank level, and since these rates of substitution are equal to zero at the planner's optimum for banks $2,\ldots,N$, the condition degenerates into the marginal *social* rate of substitution for bank 1. The right-hand side is the marginal rate of transformation of investment into protection by bank 1.

The numerator on the left-hand side indicates that, relative to the single bank case, the planner now considers how bank 1's allocation to protection offers both additional marginal profits to other banks (the third term), as well as extra social benefits from preventing the failure of other banks (the fourth term). Thus, the shadow of socially costly bank failure ex post shapes incentives to protect the system ex ante. This social protection must be balanced against the private balance-sheet resilience of bank 1 (the denominator on the left-hand side).

Proposition 9. *When cybersecurity is characterised by a best-shot public good, aggregate protection is shaped by the bank with the highest marginal rate of substitution (bank 1), and only this*

bank contributes positively to protection. There is underinvestment in protection whenever bank failure is due to illiquidity, $\gamma \geq \widehat{\gamma}_1$. When failure is due to insolvency, $\gamma < \widehat{\gamma}_1$, there exists a threshold $\overline{\delta}_1$ such that for $\delta_1 < \overline{\delta}_1$, there is overinvestment in protection.

Proposition 9 shows that when illiquidity is a central concern, the failure of bank 1 to account for both the social costs of its failure, λ, and the social benefit to other banks from its contributions to cybersecurity, $\sum_{j=2}^{N} \partial \pi_j / \partial X \geq 0$, leads to *system-wide underinvestment* in protection. Two points are worth highlighting. Firstly, compared with the single-bank case, the underinvestment that occurs when the bank fails due to illiquidity is aggravated due to the positive spillovers from one bank's allocation to protection on other banks sharing IT infrastructure. Secondly, in addition to the direct free-riding benefits enjoyed by banks $2,\ldots,N$ from the best-shot contributions (reflected in the third term of the numerator on the left-hand side of equation (16)), these banks enjoy an added indirect free-riding benefit from the pricing of debt. Since the planner assigns all of these banks' contributions to balance sheet resilience instead of protection, the pricing of debt is at its minimum in equilibrium, further bolstering the banks' expected equity values.

By contrast, when failure is insolvency driven, the *system-wide overinvestment* derived in Proposition 6 (the single-bank case) is tempered in Proposition 9 by social benefits from protection, and is reversed when the strongest link's deadweight losses are sufficiently high. The higher is δ_1, the more bank 1 contributes to protection in the private allocation. But at the same time, the marginal social benefit from heightened protection is also increasing in δ_1 at a larger rate. Therefore, when $\delta_1 > \overline{\delta}_1$, the planner prefers to increase bank 1's contribution. Our result highlights the countervailing effects imposed by the marginal cost to the planner of increasing bank 1's protection, $\lambda(1-p(S_1))\frac{\partial \alpha_1^*}{\partial S_1}$, and the marginal social benefits to protection, $\sum_{j=2}^{N} \partial \pi_j / \partial X + \lambda \frac{p}{2X}(1-\alpha_j^*) \geq 0$.

Weakest link public good. When cybersecurity investments become perfect complements, $X(\vec{S})$ converges to a *weakest-link* public good, namely

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

Examples of cybersecurity as a weakest-link public good include social engineering attacks, where people with privileged IT access are manipulated by attackers to divulge passwords or grant access to attackers, often disguised as employees, to protected systems.[23] Another example is the improper updating or maintenance of systems that allows the attacker "backdoor" access to other banks' systems hosted on the same server.[24]

It is a well-established result that in public goods games with a weakest link aggregation, the provision of the public good is shaped by the marginal rate of substitution of the *weakest-link* (Varian, 2004), with all other agents matching this contribution. Each bank then finds it optimal to set its allocation to protection to match that of bank N. Unlike the best-shot allocation, this environment permits a range of Nash equilibria, with the largest (and Pareto-dominant) equilibrium corresponding to the allocation denoted $(S_1^{**},\ldots,S_N^{**})$, where all banks allocate S_N^{**} to protection. Therefore, the allocation is shaped solely by the marginal rate of substitution of the *weakest link*: bank N.

In the Pareto-dominant equilibrium corresponding to the planner's allocation, accounting for how the face value of debt responds to cybersecurity allocation for each bank, we have an allocation for each bank equal to $S_N^P > 0$. The Samuelson condition becomes

$$\sum_{b}^{N} \frac{\frac{\partial \pi_b}{\partial X} + \lambda \frac{p}{2X}(1-\alpha_b^*)}{\partial \pi_b / \partial I_b + \lambda(1-p)\frac{\partial \alpha_b^*}{\partial I_b}} = 1\,. \tag{18}$$

Unlike in condition (16), the planner's allocation in the weakest link features positive allocations for all banks $b = 1,\ldots,N$, so that the marginal rate of transformation is equated to the sum of marginal rates of substitution for all participating banks. Importantly, the level S_b^P is driven entirely by bank N, since this bank's marginal rate of substitution anchors all other allocations.

[23] Companies including Visa and New York Life Insurance Co. have been the recent targets of social engineering attacks by a hacker group known as "Scattered Spider". The attackers trick staff into changing passwords or resetting authentication processes (Financial Times, 2025).

[24] 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).

Proposition 10. *When cybersecurity is a weakest-link public good, system-wide protection is proportional to the bank with the lowest marginal rate of substitution and all banks match this allocation. If bank N fails due to illiquidity, $\gamma \geq \widehat{\gamma}_N$ it underinvests in protection in all private equilibria. When failure due to insolvency occurs, $\gamma < \widehat{\gamma}_N$, there is overinvestment if and only if $\delta_N < \overline{\delta}_N$.*

The systemic implications of Proposition 9 largely extend to Proposition 10, but there are three key differences. Firstly, whereas in best-shot cybersecurity environments all banks but one allocate nothing to protection, in the case of a weakest link public good, there are positive contributions by all participating banks. Secondly, in the best shot setting, the indirect free-riding benefits that banks $2,\ldots,N$ enjoy over and above their own lack of contributions, via lower debt pricing, is absent in a weakest link environment because banks are incentivised to match the contribution of the bank with the lowest marginal rate of substitution. And whereas the marginal benefit of higher allocation to cybersecurity by non-contributing banks, b, equals zero in the best shot case, the effect is positive for all banks in the weakest link setting. So although there may be positive marginal social benefits to increasing protection by a given bank beyond what the bank would choose, doing so would render $\partial X/\partial S_b = 0$, so there is underinvestment only when $S_b^* < S_N^P$.

Regulatory implications. From a financial stability standpoint, the socially optimal balance between protection and balance-sheet resilience hinges critically on the nature of bank failure. When a bank's efforts to shore up its defences against unwarranted intrusion offer only private benefits, Proposition 6 shows that whether there is over- or underinvestment in protection, relative to a planner's choice, is crucially dependent on the market conditions (i.e., γ) that make illiquidity or insolvency a central concern for the bank. But when effective protection requires the collective effort of banks sharing common IT infrastructure, balance-sheet resilience can become a secondary concern to shoring up system-wide protection.

Propositions 9 and 10 together highlight the financial stability concerns associated with banks' collective misallocation, and that the direction of correction depends critically on the characteristics of the banks and the mechanism of failure. Table 4 summarises how targeted efforts shift to banks with different balance-sheet characteristics depending on market conditions.

	Cybersecurity is a best shot public good	**Cybersecurity is a weakest link public good**
Banks fail due to insolvency	Reduce protection of strongest links if $\delta_1 < \overline{\delta}_1$	Reduce protection of weakest links if $\delta_N < \overline{\delta}_N$
Banks fail due to illiquidity	Increase protection of strongest links	Increase protection of weakest links

TABLE 4. Optimal policy targeting under different cyber risk environments

Cyber risks resembling best-shot public goods need intervention only on the "*strongest links*", i.e., those banks with the highest marginal rates of substitution – that is, those banks for whom balance-sheet resilience is a lesser concern. While in the case of a weakest-link environment, it is the "*weakest links*", i.e., banks for whom balance-sheet resilience is paramount that should be encouraged to increase their allocation to protection.

7. CONCLUSION

This paper develops a framework for analysing how cyberattacks interact with bank fragility and, more broadly, how financial institutions manage operational risks that threaten both cash flows and stability. The model highlights that cybersecurity is a risk-management choice: increasing protection reduces the bank's exposure to adverse outcomes but also diverts resources from investment, reducing resilience if an attack succeeds. Whether the bank privately over- or underinvests in security depends on whether failure is insolvency- or illiquidity-driven—a distinction with important regulatory implications. Regulators must be alert to this, along with key elements of the cybersecurity landscape and the public good characteristics of cybersecurity, when designing policy.

To date, cybersecurity regulation has largely focused on preventive measures. But ex-ante protection and ex-post resilience are interdependent. Excessive investment in cybersecurity can, in some cases, weaken a bank's balance sheet resilience, and exacerbate financial fragility. Regulation may benefit from a more targeted approach, recognising that optimal policy may involve

reallocating resources from improving protection against cyberattacks to improving balance sheet resilience in the event of an attack, rather than seeking to uniformly increase expenditures.

The core insights of our analysis carry over into more general settings. The introduction of a positive recovery rate for investors and contractible cybersecurity investments 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

$$\alpha > \alpha^{IN} \equiv \frac{1}{\delta}\left(1 - \frac{FD}{RI}\right), \tag{19}$$

and it fails due to illiquidity whenever

$$\alpha > \alpha^{IL}(\ell) \equiv 1 - \frac{\ell FD}{RI}. \tag{20}$$

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 $\alpha^{IL}(\widehat{\gamma}) = \alpha^{IN}$, i.e.,

$$\frac{1}{\delta}\left(1 - \frac{FD}{RI}\right) = 1 - \frac{\widehat{\gamma} FD}{RI}, \tag{21}$$

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 decreasing in S and F, suppose that $F = r$ and $S = 0$. By the supposition, we have that the bank finds it optimal to take on as much debt as possible so that $D = 1$, which we show to be the case later on. 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 also 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.[25] 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,

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

forming the lower dominance region. For $\alpha \in (\underline{\alpha}, \overline{\alpha})$, if the shock were common knowledge, the run dynamics of fund managers would be characterised by multiple, self-fulfilling equilibria, as shown in Figure 6.

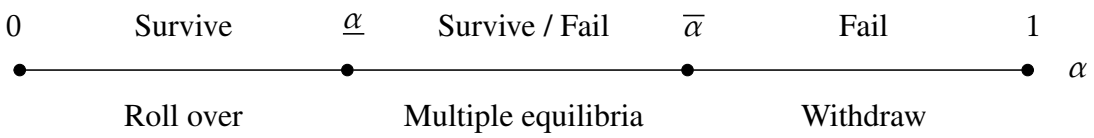

FIGURE 6. Tripartite classification of the accessibility shock.

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

$$s(x_k) = \begin{cases} \text{withdraw} & \text{if } x_k > x^*, \\ \text{roll over} & \text{if } x_k \leq x^*. \end{cases} \tag{22}$$

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

$$\ell(\alpha, x^*) = \Pr(x_k > x^* | \alpha) = 1 - H\Big(x^* - \alpha\Big), \tag{23}$$

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

$$\alpha^* = 1 - \ell^*(\alpha^*, x^*)\frac{F}{RI}. \tag{24}$$

For a given x^*, the left-hand side of Equation (24) 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

$$\gamma = \Pr(\alpha \leq \alpha^* | x_k = x^*). \tag{25}$$

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 2.** The marginal net benefit of expending effort in an attack is

$$-V\frac{\partial p(A,S)}{\partial A} - c. \tag{26}$$

Using the definition for $p(A,S)$ in (1), this can be written as

$$\frac{SV}{(A+S)^2} - c. \tag{27}$$

For all $S \geq \frac{V}{c}$, attacker payoffs are weakly decreasing in effort and the attacker's optimal level of effort is $A^* = 0$. While for all $S < \frac{V}{c}$, the optimal level of effort solves

$$\frac{SV}{(A^*+S)^2} = c, \tag{28}$$

which, on rearranging, yields the expression in the first piece of (8).

A.4. **Proof of Lemma 3.**

Intermediation margin. The marginal benefit from issuing debt is given by

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

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 $(1-p)\mathbb{1}_{\gamma>\widehat{\gamma}}$ 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., $D^* = 1$.

Participation constraint. 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

$$\mathcal{L} = \pi(I, p) + \mu(1 - I - S) + \phi\left(p - \sqrt{\frac{cS}{V}}\right), \tag{29}$$

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

$$\begin{aligned}\frac{\partial \pi}{\partial S} &= \frac{\partial p}{\partial S}\left[R(1-S)-F-\int_0^{\alpha^*(S)} E(\alpha)d\alpha\right] \\ &- R\left[p(S)+\Big(1-p(S)\Big)\int_0^{\alpha^*(S)}(1-\delta\alpha)d\alpha\right]+\Big(1-p(S)\Big)E(\alpha^*)\frac{\partial \alpha^*}{\partial S}\,\mathbb{1}_{\gamma>\widehat{\gamma}}=0. \qquad (30)\end{aligned}$$

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 3 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

$$\begin{aligned}\frac{\partial^2 \pi}{\partial S^2} &= \frac{\partial^2 p}{\partial S^2}\left[R(1-S)-F-\int_0^{\alpha^*(S)} E(\alpha)d\alpha\right]-2\frac{\partial p}{\partial S}R\left[1-\int_0^{\alpha^*(S)}(1-\delta\alpha)d\alpha\right] \qquad (31)\\ &- \Big(1-p(S)\Big)\frac{\partial \alpha^*}{\partial S}R\Big(1-\delta\alpha^*(S)\Big) \\ &- \frac{\partial \alpha^*}{\partial S}\left(E(\alpha^*(S))\left[\frac{p(S)}{S}-\frac{1-p(S)}{1-S}\right]+\frac{F(1-\gamma\delta)}{1-S}\right)\times \mathbb{1}_{\gamma>\widehat{\gamma}}.\end{aligned}$$

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 (31) yields

$$-\frac{p(S)}{S}\frac{R\delta}{2}-\frac{F^2\gamma}{R(1-S)^2}\left[p(S)+\left(\frac{p(S)}{S}-1\right)(2-\gamma\delta)\right], \qquad (32)$$

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

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

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

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

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

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

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

The terms on the second line may be expressed as

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

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 $\frac{\partial^2\pi}{\partial S\partial F}\big|_{\gamma<\widehat{\gamma}}>0$, and so $\frac{\partial S^*}{\partial F}>0$. When failure is illiquidity driven,

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

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

A.6. **Proof of Lemma 4.** Given F, the value of a debt claim is

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

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 $\gamma\geq\widehat{\gamma}$. 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

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

For $\gamma \geq \widehat{\gamma}$, 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

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

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 $\gamma \geq \widehat{\gamma}$ 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 $\gamma \geq \widehat{\gamma}$. 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 in Proposition 2, at this equilibrium, the threshold $\widehat{\gamma}(F^{**}, S^{**})$ is well defined.

Suppose $\gamma < \widehat{\gamma}$. 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. 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 $S^*(\tilde{F}) = S^{\dagger}$, and since $S^{\dagger} > S^{\dagger\dagger}$. 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

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

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

$$\left.\frac{\partial^2 \pi}{\partial S \partial c}\right|_{S=S^*} = \frac{\partial p/\partial c}{p(S^*)}\left\{R\int_0^{\alpha^*(S^*)}(1-\delta\alpha)d\alpha - E(\alpha^*)\frac{\partial \alpha^*}{\partial S}\mathbb{1}_{\gamma>\widehat{\gamma}}\right\} > 0. \tag{36}$$

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 outweighs the indirect effects via the pricing of debt.

This is tantamount to requiring that

$$\frac{R\left(\int_0^{\alpha^*}(1-\delta\alpha)d\alpha\right) - E(\alpha^*)\frac{\partial \alpha^*}{\partial S}\mathbb{1}_{\gamma>\widehat{\gamma}}}{F(1-\alpha^*)} > \frac{p(S)\frac{\partial^2 \pi}{\partial S \partial F}}{p(S)+(1-p(S))\left[\alpha^* + F\frac{\partial \alpha^*}{\partial F}\right]}. \tag{37}$$

In the limit $c \to 0$, such that the bank is almost surely to loose the contest, $p(S) \to 0$, the right-hand side of Equation (37) is zero, and so the condition is trivially satisfied. Thus, insofar that the right-hand side on Equation (37) is decreasing in c, it implies that the condition holds for all levels of attacker sophistication. To see this, note that the derivative of the denominator term on the right-hand side of Equation (37) with respect to c is (up to a factor $dp(S)/dc$) proportional to $\left(2(1-\alpha^*) + \left(\frac{1}{\delta}-1\right)\mathbb{1}_{\gamma\leq\widehat{\gamma}}\right) > 0$.

In the case where bank failure is driven by insolvency, the derivative of the numerator is weakly decreasing in c as long as

$$p(S) \geq \frac{FS}{F(1+S)-(1-\delta)R(1-S)^2}.$$

Since this inequality holds in the limit $S \to 0$ and the right-hand side is bounded from above by $\frac{rS}{r(1+S)-(1-\delta)R(1-S)^2}$, which is decreasing in S, the condition holds for all S. And when bank failure is illiquidity driven, the derivative of the numerator term on the right-hand side of Equation (37) is weakly decreasing in c as long as

$$p(S) \geq \frac{S}{1+S-\frac{R(1-\delta)(1-S)^2}{(2-\gamma\delta)F}}.$$

As in the case where bank failure is driven by insolvency, this condition holds in the limit $S \to 0$ and since the right-hand side is decreasing in S, the condition holds for all S.

Thus, irrespective of whether bank failure is driven by illiquidity or insolvency, the entire right-hand side term in Equation (37) is decreasing in c. And so, for all values of c, the direct effect dominates the indirect effect, implying that $\frac{dS^{**}}{dc} > 0$.

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

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

where

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

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 $\frac{\partial F^*}{\partial \delta} > 0$. The upward shift in the $F^*(S)$ schedule reinforces the effect of a marginal increase in δ on the $S^*(F)$ schedule. For $\gamma \geq \widehat{\gamma}$, note that $\mathcal{V}$ does not depend on δ, and so the only effect stems from that on the $S^*(F)$ schedule. Thus, in either case, we have that $\frac{dS^{**}}{d\delta} > 0$.

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

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

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

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

Thus, the cross derivative is strictly positive and so by the implicit function theorem, $\frac{\partial S^*}{\partial \gamma} > 0$ when $\gamma \geq \widehat{\gamma}$. And since α^{IN} does not depend on γ, we have that $\frac{\partial S^*}{\partial \gamma} = 0$ when $\gamma < \widehat{\gamma}$.

Turning to the investors' schedule for the face value of debt, for $\gamma < \widehat{\gamma}$, this is similarly independent of γ. While for $\gamma \geq \widehat{\gamma}$, we obtain $\frac{\partial \mathcal{V}}{\partial \gamma} = F\Big(1 - p(S)\Big)\frac{\partial \alpha^*}{\partial \gamma} < 0$, 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 $\frac{dS^{**}}{d\gamma} \geq 0$.

A.9. **Proof of Proposition 5.** It follows from Propositions 2 - 4, that the allocation to cybersecurity is greater when its failure is driven by illiquidity when compared to the case where bank failure is driven by insolvency. The introduction of an EPN implies that the bank's expected profit in the event of a successful attack is a probability-weighted average of its profits when failure is insolvency driven and when it is illiquidity driven, given by

$$\pi(S) = p(S)(R(1-S) - F) + (1 - p(S))\left[\beta \int_0^{\alpha^{IN}} E(\alpha)d\alpha + (1-\beta)\int_0^{\alpha^{IL}(\gamma)} E(\alpha)d\alpha\right]. \tag{38}$$

As β increases, the objective approaches the expected equity value weighted by the insolvency failure condition. We thus obtain that $\frac{\partial^2 \pi}{\partial S \partial \beta} < 0$, implying that $\frac{\partial S^*}{\partial \beta} < 0$.

From the investors' perspective, the value of the debt claim is

$$\mathcal{V}(F) = \Big(p(S) + \Big(1 - p(S)\Big)\Big(\beta\,\alpha^{IN} + (1-\beta)\,\alpha^{IL}(\gamma)\Big)F\,, \tag{39}$$

implying that $\frac{\partial \mathcal{V}}{\partial \beta} > 0$ and so $\frac{\partial F^*}{\partial \beta} < 0$. The total effect in equilibrium is thus an unambiguous decrease in the bank's investment in cybersecurity, $\frac{dS^{**}}{d\beta} < 0$.

A.10. **Proof of Proposition 6.** 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 overinvests in cybersecurity, while if $\frac{\partial \Delta}{\partial S}\Big|_{S=S^*} > 0$, it underinvests. 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, $\gamma \geq \widehat{\gamma}$.

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

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

Substituting $\alpha^*(S) = \frac{1}{\delta}\left(1-\frac{F}{R(1-S)}\right)$ into the above expression, we have

$$\frac{1}{\lambda}\frac{\partial \Delta}{\partial S}\bigg|_{S=S^*} = \frac{F}{\delta R(1-S^*)}\left(\frac{p(S^*)}{2S^*} - \frac{1-p(S^*)}{1-S^*}\right) - \frac{p(S^*)}{2S^*}\left(\frac{1}{\delta}-1\right) < 0, \tag{40}$$

where the condition holds by the assumption that $R > \underline{R}$ in Proposition 2. Therefore $S^{**} > S^P$ and the bank overinvests in cybersecurity.

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.11. **Proof of Proposition 7.**

Insolvency. With a tax scheme in place, the bank's expected profits include a tax that is linear in investment in cybersecurity, τS, and a lump-sum rebate, Σ,

$$\pi(S) = p(S)[R(1-S)-F-\tau S+\Sigma] + (1-p(S))\left[\int_0^{\alpha^*(S)} (R(1-S))(1-\alpha\delta) - F - \tau S + \Sigma) d\alpha\right]. \tag{41}$$

The bank's first-order condition is:

$$\begin{aligned}\frac{\partial \pi}{\partial S} = \frac{\partial p}{\partial S}\left[R(1-S) - F - \tau S + \Sigma - \int_0^{\alpha^*(S)} (R(1-\alpha\delta)(1-S) - F - \tau S + \Sigma)\, d\alpha\right] \\ - p(S)R - (1-p(S))\int_0^{\alpha^*(S)} (1-\alpha\delta) d\alpha - \tau[p(S) + (1-p(S))\alpha^*(S)].\end{aligned} \tag{42}$$

For the tax to be revenue neutral, we require $\Sigma = \tau S$, which simplifies the first-order condition to

$$\frac{\partial \pi}{\partial S} = \frac{\partial \pi^0}{\partial S} - \tau[p(S) + (1 - p(S))\alpha^*(S)], \tag{43}$$

where π^0 is the bank's profit function in absence of the policy. To recover the planner's solution, we require

$$\frac{\partial \pi}{\partial S} = \frac{\partial W}{\partial S} = \frac{\partial \pi^0}{\partial S} + \lambda \left[\frac{\partial p}{\partial S}(1 - \alpha^*(S)) + (1 - p(S))\frac{\partial \alpha^*}{\partial S} \right], \tag{44}$$

which optimally sets the tax to

$$\tau^* = \frac{\lambda \left(\frac{\partial p}{\partial S}\Big|_{S=S^P} (1 - \alpha^*(S^P, F^P) + (1 - p(S^P)\frac{\partial \alpha^*}{\partial S}\Big|_{S=S^P, F=F^P} \right)}{p(S^P) + (1 - p(S^P))\alpha^*(S^P, F^P)}. \tag{45}$$

Illiquidity. It is straightforward to repeat the exercise to elicit the optimal subsidy in equation (14).

A.12. **Proof of Proposition 8.** First consider the $S^*(F)$ schedule. The cross derivative of the bank's expected profits with respect to the level of cybersecurity assistance, evaluated at S^*, is

$$\frac{\partial^2 \pi}{\partial S \partial \rho}\Big|_{S=S^*} = \frac{\partial p / \partial \rho}{p(S^*)} \left\{ R \int_0^{\alpha^*(S)} (1 - \delta\alpha) d\alpha - E(\alpha^*) \frac{\partial \alpha^*}{\partial S} \mathbb{1}_{\gamma > \widehat{\gamma}} \right\} > 0. \tag{46}$$

The $S^*(F)$ schedule thus shifts rightwards. The effect on the face value of debt is given by

$$\frac{\partial F^*}{\partial \rho} = \frac{-F}{\frac{\partial \mathcal{V}}{\partial F}} \left\{ \frac{p(S)}{2(1 + \rho)} (1 - \alpha^*(S)) \right\} < 0, \tag{47}$$

since $\frac{\partial \mathcal{V}}{\partial F} > 0$ by Lemma 4. The direct effect dominates and S^{**} is increasing in ρ whenever

$$\frac{\partial S^*}{\partial \rho} + \frac{\partial S^*}{\partial F} \frac{\partial F^*}{\partial \rho} > 0. \tag{48}$$

Condition (48) holds by the same properties that ensure $\frac{\partial S^{**}}{\partial c} > 0$, which we have shown to be true in the proof of Proposition 4. To see this, if we compare the comparative static in (36) with (46), and $\frac{\partial F^*}{\partial c}$ with condition (47), the effects are clearly qualitatively identical since c and $(1 + \rho)$ enter into $p(S)$ in the same way. Since S^{**} is increasing, while F^{**} is decreasing in ρ, the planner's solution at (S^P, F^P) cannot be obtained simultaneously.

A.13. **Proof of Proposition 9.** In the private equilibrium among banks and investors, the level of protection obtained by all banks is proportional to the allocation chosen by the bank with the highest marginal rate of substitution, $b = 1$.[26] The contribution by this bank, $S_1^{**} > 0$, satisfies

$$\frac{\frac{\partial \pi_1}{\partial X} + \lambda \frac{p}{2S_1^{**}}(1-\alpha_1^*)}{\frac{\partial \pi_1}{\partial I_1} + \lambda(1-p)\frac{\partial \alpha_1^*}{\partial I_1}} = 1\,. \tag{49}$$

Illiquidity. For there to be underinvestment in protection at S_1^{**}, relative to allocation S_1^P defined in equation (16), we require

$$\left.\frac{\partial \pi_1/\partial X}{\partial \pi_1/\partial I_1}\right|_{S_1=S_1^{**}} < \left.\frac{\frac{\partial \pi_1}{\partial X} + \lambda \frac{p}{2X}(1-\alpha_1^*) + \sum_{j\neq 1}^N \frac{\partial \pi_j}{\partial X} + \lambda \frac{p}{2X}(1-\alpha_j^*)}{\frac{\partial \pi_1}{\partial I_1} + \lambda(1-p)\frac{\partial \alpha_1^*}{\partial I_1}}\right|_{S_1=S_1^{**}}. \tag{50}$$

Using the fact that $\frac{\partial \pi_1/\partial X}{\partial \pi_1/\partial I_1} = 1$ at $S_1 = S_1^{**}$, condition (50) can be written as follows

$$(1-p)\left.\frac{\partial \alpha_1^*}{\partial I_1}\right|_{S_1=S_1^{**}} < \frac{p}{2S_1^{**}}(1-\alpha_1^*) + \frac{1}{\lambda}\left(\sum_{j\neq 1}^N \frac{\partial \pi_j}{\partial X} + \lambda \frac{p(S_1)}{2S_1}(1-\alpha_j^*)\right), \tag{51}$$

Noting that $\left.\frac{\partial \alpha_1^*}{\partial I_1}\right|_{\gamma \geq \widehat{\gamma}_1} = (1-S_1^{**})(1-\alpha_1^*)$, and defining $\Delta_b(S_b) \equiv -\lambda(1-p(X))(1-\alpha_b^*(S_b))$, it follows from Proposition 6 that $\left.\frac{\partial \Delta_1}{\partial S_1}\right|_{S_1=S_1^{**}} > 0$ whenever $\gamma \geq \widehat{\gamma}_1$, and underinvestment is aggravated since

$$\sum_{j\neq 1}^N \frac{\partial \pi_j}{\partial X} + \lambda \frac{p(S_1)}{2S_1}(1-\alpha_j^*) \geq 0\,.$$

Insolvency. In the case where $\gamma < \widehat{\gamma}_1$, the equilibrium allocation S_1^{**} is greater than in the equilibrium corresponding to the planner's allocation, S_1^P, whenever

$$\frac{\partial \Delta_1}{\partial S_1} + \frac{1}{\lambda}\left(\sum_{j\neq 1}^N \frac{\partial \pi_j}{\partial X} + \lambda \frac{p(S_1)}{2S_1}(1-\alpha_j^*)\right) < 0\,. \tag{52}$$

[26] To see this, note that for any allocation by other banks, bank 1 finds it optimal to allocate S_1^{**} to protect against attack. Under this allocation, the marginal benefit to any other bank $j \neq 1$ from allocating a unit to cybersecurity, $\frac{\partial \pi_j}{\partial X}\frac{\partial X}{\partial S_j} = 0$, and so each other bank free rides on bank $b = 1$ and allocates all resources towards shoring up resilience.

Substituting $\alpha_1^* = \frac{1}{\delta_1}\left(1 - \frac{F_1}{RI_1}\right)$ into condition (52) and rearranging, we have overinvestment by bank 1 whenever

$$\delta_1 < \overline{\delta}_1 \equiv \frac{\frac{(1-p)}{(1-S_1^{**})}\left(\frac{F_1^{**}}{RI_1}\right) + \frac{p}{2S_1^{**}}\left(1 - \frac{F_1^{**}}{RI_1}\right)}{\frac{p}{2S_1^{**}} + \frac{1}{\lambda}\left(\sum_{j\neq b}^{N}\frac{\partial \pi_j}{\partial X} + \lambda\frac{p(S_b)}{2S_b}(1-\alpha_j^*)\right)}, \tag{53}$$

where $\lim_{S_1^{**}\to 0}\overline{\delta}_1 = 0$, $\lim_{S_1^{**}\to \overline{S}_1}\overline{\delta}_1 > 0$ and $\frac{\partial \overline{\delta}_1}{\partial S_1} > 0$.

A.14. **Proof of Proposition 10.** The private equilibrium among banks and investors features an identical level of investment by each bank, S_N^{**}, where S_N^{**} satisfies

$$\frac{\partial \pi_N/\partial X}{\partial \pi_N/\partial I_N} = 1\,. \tag{54}$$

In what follows, it suffices to focus on bank N since over or underinvestment for any bank b is identical to the extent to which bank N over- or underinvests.[27]

Illiquidity. There is underinvestment at S_N^{**} relative to S_N^P as defined in (18) if

$$\frac{\partial \pi_N/\partial X}{\partial \pi_N/\partial I_N} < \frac{\frac{\partial \pi_N}{\partial X} + \lambda\frac{p}{2X}(1-\alpha_N^*) + \sum_{j\neq N}\frac{\partial \pi_j}{\partial X} + \lambda\frac{p}{2X}}{\frac{\partial \pi_N}{\partial I_N} + \lambda(1-p)\frac{\partial \alpha_N^*}{\partial I_N}}, \tag{55}$$

which can be reformulated, as in the best-shot case, using the fact that $\frac{\partial \pi_N/\partial X}{\partial \pi_N/\partial I_N} = 1$ at $S_N = S_N^{**}$, as

$$(1-p)\frac{\partial \alpha_N^*}{\partial I_N}\bigg|_{S_N = S_N^{**}} < \frac{p}{2S_N^{**}}(1-\alpha_N^*) + \frac{1}{\lambda}\left[\sum_{j\neq N}\frac{\partial \pi_j}{\partial X} + \lambda\frac{p}{2X}(1-\alpha_j^*)\right]. \tag{56}$$

By Proposition 6, $\frac{\partial \Delta_N}{\partial S}\Big|_{S_N = S_N^{**}} > 0$ whenever $\gamma \geq \widehat{\gamma}_N$, and since $\sum_{j\neq N}\frac{\partial \pi_j}{\partial X} + \lambda\frac{p}{2X}(1-\alpha_j^*) \geq 0$, there is underinvestment at S_N^{**}.

[27] For each bank other than the weakest link, $j \neq N$, the marginal product of cybersecurity investment given allocation S_N^{**} by bank N is zero, $\frac{\partial X}{\partial S_j}\Big|_{S_N = S_N^{**}} = 0$, in both the private and planner's allocations.

Insolvency. When $\gamma < \widehat{\gamma}_N$, substituting $\alpha_N^* = \frac{1}{\delta_N}\left(1 - \frac{F}{RI_N}\right)$ into condition (56), and using the fact that $\frac{\partial \pi_N/\partial X}{\partial \pi_N/\partial I_N} = 1$ at $S_N = S_N^{**}$, there is underinvestment if and only if

$$\delta_N > \overline{\delta}_N \equiv \frac{\frac{(1-p)}{(1-S_N^{**})}\frac{F_N}{RI_N} + \frac{p}{2S_N^{**}}\left(1 - \frac{F_N}{RI_N}\right)}{\frac{p}{2S_N^{**}} + \frac{1}{\lambda}\left(\sum_{j \neq N} \frac{\partial \pi_j}{\partial X} + \lambda \frac{p}{2X}(1 - \alpha_j^*)\right)}. \tag{57}$$

Appendix B. Robustness exercises

Our results are robust to the inclusion of a strictly positive recovery rate for investors, and allowing the bank's allocation to cybersecurity to be observable by investors. We consider each in turn.

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

$$\mathcal{V}(F,S) = p(S)F + \Big(1 - p(S)\Big)\Big[\alpha^{IN}(S,F)\,F + \Omega(F,S)\Big], \tag{58}$$

where

$$\Omega(F,S) = \int_{\alpha^{IN}(F,S)}^{1} (1 - \alpha\delta)R(1-S)\,d\alpha\,, \tag{59}$$

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 allocation 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.[28] In sum, the core tension between protection and resilience remains intact when we relax the zero-recovery rate assumption for investors.

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 internalise how its investment choice impacts the face value of debt. We extend our analysis to consider what happens when the bank does internalise this impact. Formally, the bank's ex ante problem is given by

$$\max_S \pi(S) \equiv p(S)\Big(R(1-S)-F^*(S)\Big) + \Big(1-p(S)\Big)\int_0^{\alpha^*(S)}\Big[(1-\alpha\delta)R(1-S)-F^*(S)\Big]d\alpha, \quad (60)$$

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

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

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

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 allocation to cybersecurity is reduced.

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 6, it follows that there is overinvestment in cybersecurity, $S^{**} > S^P$. On the other hand, when $\gamma > \widehat{\gamma}$ such that $dF^*/dS < 0$, the the bank always underinvests 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.

REFERENCES

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., J. Frost, L. Gambacorta, and D. Whyte (2021). Covid-19 and cyber risk in the financial sector. BIS Bulletin No. 37.

Bank of England (2023). Thematic findings from the 2022 cyber stress test. https://bankofengland.co.uk/-/media/boe/files/prudential-regulation/letter/2023/thematic-findings-2022-cyber-stress-test.pdf.

Biais, B. and C. Casamatta (1999). Optimal leverage and aggregate investment. The Journal of Finance 54(4), 1291–1323.

Biais, B., T. Mariotti, J.-C. Rochet, and S. Villeneuve (2010). Large risks, limited liability, and dynamic moral hazard. Econometrica 78(1), 73–118.

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.

Bliss, C. and B. Nalebuff (1984). Dragon-slaying and ballroom dancing: The private supply of a public good. Journal of Public Economics 25(1-2), 1–12.

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

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., L. Laeven, and R. Marquez (2014, January). Real interest rates, leverage, and bank risk-taking. Journal of Economic Theory 149, 65 – 99.

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

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

Ehrlich, I. and G. S. Becker (1972). Market insurance, self-insurance, and self-protection. Journal of Political Economy 80(4), 623–648.

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 (2025). When it rains it pours: Cyber vulnerabilities and financial conditions. Federal Reserve Bank of New York Economic Policy Review 31, 1–24.

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

European Union Agency for Cybersecurity (2023). ENISA Threat Landscape 2023. The European Union Agency for Cybersecurity Report, October.

FBI (2021, January 5). Joint Statement by the Federal Bureau of Investigation (FBI), the Cybersecurity and Infrastructure Security Agency (CISA), the Office of the Director of National Intelligence (ODNI), and the National Security Agency (NSA). FBI press release.

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

Financial Times (2024b, January 17). JPMorgan suffers wave of cyber attacks as fraudsters get 'more devious'. https://on.ft.com/48BMDgE.

Financial Times (2025, May 26). In cyber attacks, humans can be the weakest link. https://on.ft.com/4kyjvw6.

Florackis, C., C. Louca, R. Michaely, and M. Weber (2023). Cybersecurity risk. The Review of Financial Studies 36(1), 351–407.

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

Froot, K. A., D. S. Scharfstein, and J. C. Stein (1993). Risk management: Coordinating corporate investment and financing policies. The Journal of Finance 48(5), 1629–1658.

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.

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.

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

He, Z. and W. Xiong (2012). Rollover risk and credit spreads. The Journal of Finance 67(2), 391–430.

Hellmann, T. F., K. C. Murdock, and J. E. Stiglitz (2000). Liberalization, moral hazard in banking, and prudential regulation: Are capital requirements enough? American Economic Review 91(1), 147–165.

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

Hoffmann, F., R. Inderst, and M. M. Opp (2022). The economics of deferral and clawback requirements. Journal of Finance 77(4), 2423–2470.

Huang, H. H. and C. Wang (2021). Do banks price firms' data breaches? The Accounting Review 96(3), 261–286.

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

Jensen, M. C. and W. H. Meckling (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. Journal of Financial Economics 3(4), 305–360.

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., D. P. Tsomocos, and A. P. Vardoulakis (2024). Optimal bank regulation in the presence of credit and run risk. Journal of Political Economy 132(3), 772–823.

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

Keeley, M. C. (1990). Deposit insurance, risk, and market power in banking. The American Economic Review 80(5), 1183–1200.

Martinez-Miera, D. and R. Repullo (2017). Search for yield. Econometrica 85, 351–378.

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/.

Moreno, D. and T. Takalo (2016). Optimal bank transparency. Journal of Money, Credit and Banking 48(1), 203–231.

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 (2020). Protecting controlled unclassified information in nonfederal systems and organizations. https://doi.org/10.6028/NIST.SP.800-171r2.

Prenio, J., J. Yong, and R. Kleijmeer (2019). Varying shades of red: How red team testing frameworks can enhance the cyber resilience of financial institutions. FSI Insights No. 21. https://www.bis.org/fsi/publ/insights21.htm.

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

Rampini, A. A. and S. Viswanathan (2010). Collateral, risk management, and the distribution of debt capacity. The Journal of Finance 65(6), 2293–2322.

Ramírez, C. A. (2025, May). On equilibrium cyber risk. Economics Letters 251.

Repullo, R. (2004, April). Capital requirements, market power, and risk-taking in banking. Journal of Financial Intermediation 13(2), 156–182.

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.

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

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.

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.