NBER WORKING PAPER SERIES

BANK HETEROGENEITY AND FINANCIAL STABILITY

Itay Goldstein
Alexandr Kopytov
Lin Shen
Haotian Xiang

Working Paper 27376
http://www.nber.org/papers/w27376

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2020

Itay Goldstein has been retained by a law firm as an expert on a case involving a liquidity crisis and government intervention in a European bank. This case has not influenced the research in this paper, and the research in this paper has not influenced the case. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Bank Heterogeneity and Financial Stability
Itay Goldstein, Alexandr Kopytov, Lin Shen, and Haotian Xiang
NBER Working Paper No. 27376
June 2020
JEL No. G01,G21,G23

ABSTRACT

We study how heterogeneity in banks' asset holdings affects fragility. In the model, banks face a risk of bank runs and have to liquidate long-term assets in a common market to repay runners. Liquidation prices are depressed when many banks sell their assets at the same time. When banks are homogeneous, their selling behaviors are synchronized, and bank runs are exacerbated. We show that differentiating banks to some extent enhances the stability of all banks, even those whose asset performance ends up being weaker. Our analyses provide new insights about the regulation of banking sector's architecture and the design of government support during crises.

Itay Goldstein
The Wharton School
University of Pennsylvania
3620 Locust Walk
Philadelphia, PA 19104
and NBER
itayg@wharton.upenn.edu

Alexandr Kopytov
Princeton University
alexandrkopytov91@gmail.com

Lin Shen
Boulevard de Constance
INSEAD
77300, Fontainebleau
France
lin.shen@insead.edu

Haotian Xiang
Peking University
xiang@gsm.pku.edu.cn

1 Introduction

One of the key questions coming out of the financial crisis of 2008 surrounds the way by which shocks can spread across financial institutions (for short, banks) and bring down the system as a whole. A key challenge going forward is understanding how the financial system should be designed to make it more resilient and less prone to fragility. Of particular concern is the issue of homogeneity. On the one hand, various regulations push towards greater homogeneity across banks, e.g., by encouraging banks to diversify their assets and be less exposed to local shocks. On the other hand, there is a growing recognition that homogeneity might synchronize fragility across institutions making its overall impact on the system greater (e.g. Haldane, 2009, Haldane and May, 2011, Yellen, 2013).

In this paper, we provide theoretical analysis on this particular issue: how bank asset heterogeneity affects financial stability. We model a banking sector in which assets of an individual bank are subject to both aggregate and idiosyncratic shocks. Banks finance themselves with short-term demandable debt, and so are exposed to the risk of a run a la Diamond and Dybvig (1983). Banks are indirectly interconnected due to the fact that, in the face of a run, they liquidate their assets in a common asset market, which leads them to impose negative fire-sale externalities on one another.

This structure leads to two layers of coordination problems. First, typical to a bank-run model, there is a *within-bank strategic complementarity*. Investors withdrawing money from the bank impose a negative externality on those who stay, because withdrawals lead to costly liquidations and reduce the amount available to those who do not withdraw. Hence, investors' inclination to withdraw increases when they expect that more of the other investors in their bank will do so. Second, there is a *cross-bank strategic complementarity*. Because of the liquidation fire-sale externality among different banks in the asset market, an individual bank needing to pay its own investors has to bear a higher liquidation cost when more banks in the economy liquidate their long-term assets to repay their investors. This implies that the inclination of investors to withdraw increases also when they expect that more investors in other banks will do so.

Crucially for our results, it turns out that these two types of strategic complementarities amplify each other. Investors are more affected by the expected behavior of investors in other banks when they think that more investors in their own bank are likely to run. This is intuitive: the fire-sale externality matters only when the bank needs to liquidate

significant amounts in the asset market, which is the case only when the bank is under a run. We think that this framework, where investors are concerned about the behavior of their own bank's investors and about the behavior of other banks' investors, and where the two concerns feed into each other, is a natural and realistic description of the financial system. As far as we know, our paper is the first one to capture it in a tractable analytical framework.

With this framework in place, we turn to analyze our main research question about the effect of homogeneity on fragility. We consider a system where the fundamentals of all banks are affected by the same aggregate shock, and in addition by idiosyncratic shocks with a zero mean across the system. The idiosyncratic shocks separate banks ex post into two types: strong banks who receive a positive shock and weak banks who receive a negative shock. An increase in heterogeneity in this framework amounts to an increase in the ex-post difference between the fundamentals of the weak banks and those of the strong banks without changing the average fundamentals of the financial system.

In this framework, we show a striking result: an increase in heterogeneity makes *all* banks more stable and is thus Pareto improving, as long as cross-bank strategic uncertainties are in place, that is, as long as investors are uncertain about the run behaviours of investors in other types of banks. An increase in heterogeneity enlarges the wedge between weak and strong banks, making the former relatively more fragile than the latter. But, it pulls all banks to an equilibrium of greater stability in absolute terms due to its effect on the indirect interactions across banks in the asset market where assets are being sold. In the face of greater heterogeneity, strong banks' stability is challenged by the greater pressure that weak banks impose on liquidation prices, and weak banks' fragility is alleviated by the lower pressure that strong banks impose. The key behind the decrease in overall fragility is that the effect on weak banks dominates the one on strong banks. This is a direct result of the fact that within-bank complementarities and cross-bank complementarities feed and amplify each other. In particular, given that weak banks are more internally fragile, their investors are more strongly affected by the lower fire-sale pressure from strong banks than strong bank investors are affected by the higher fire-sale pressure from weak banks. Hence, heterogeneity leads to overall lower fragility. Importantly, in the absence of the interaction between the within- and cross-bank complementarities, increasing heterogeneity would produce zero net effect on the overall fragility. With this interaction in place, the way that lower fragility in strong banks helps calming the fragility in weak banks becomes the dominant force, making heterogeneity a stabilizing force for

all.

As mentioned above, this benefit of heterogeneity dominates only as long as cross-bank strategic uncertainties are in place, which is true when heterogeneity is not too large, or when banks are not too different from each other. Providing a full characterization of the relationship between heterogeneity and fragility, we show that when the level of heterogeneity becomes sufficiently large to cross this threshold, a further increase in it starts to make weak banks more fragile in absolute levels (while continuing to make strong banks less fragile). As banks become so different from each other, strong banks' fire-sale pressure on weak banks is negligible, since weak banks' investors are certain that when their banks are on the margin of experiencing runs, strong banks are still sound and are not going to liquidate prematurely. As a result, a further increase in heterogeneity no longer alleviates the fire-sale pressure on weak banks and only makes them more fragile by reducing their bank-specific fundamentals. Hence, our paper calls for restricting homogeneity in the banking sector from growing above the point where cross-bank strategic uncertainties emerge.

While the main analysis is carried out in a framework with two types of banks ex post, where heterogeneity is easily defined, we also provide an extension demonstrating how the spirit of the results translates more generally to a model with many types. In addition, we explore a more general information structure with noisy signals on both idiosyncratic and aggregate fundamentals, which nests our information structure in the main framework.

The main result in our paper—that homogeneity in the financial system is destabilizing—has several implications related to financial regulation and policy. First, as mentioned in the beginning, the frequent calls, e.g., in Europe, for greater diversification and lower dependence on local shocks are missing the potential destabilizing effects that will come as a result of the greater homogeneity that diversification brings. We show that having banks exposed to different local shocks helps stabilize the system, as the reduced fragility in banks experiencing positive shocks helps reduce the fragility in those experiencing negative shocks. Second, as we explore in more detail, our model sheds light on "ring-fencing" policies that received a lot of attention since the 2008 crisis. Such policies aim to split a bank into several independent subsidiaries with separate business focuses or geographical locations.[1] We show that the key benefit in such policies comes from the

[1]For instance, the UK Financial Services Act 2012 requires large banks to isolate its core retail banking business from trading and riskier activities. The Volcker rule also aims to split retail and investment banking activities by prohibiting banks from conducting proprietary trading.

resulting heterogeneity across different subsidiaries. Third, as we also explore in more detail, our model can help guide resolution mechanisms for distressed banks. Forcing strong banks to cover weak banks, which is arguably appealing to policy makers as it requires no taxpayer money, can create unintended consequences due to the reduced heterogeneity. Similarly, when the government decides to inject cash into the system, our model suggests a benefit to boosting strong banks, which can indirectly stabilize weak banks more effectively.

To the best of our knowledge, our result that creating heterogeneity across banks is Pareto improving is novel to the literature. Existing papers studying downsides of strong asset commonality emphasize a trade-off between risk sharing and systemic risk: Institutions within a homogeneous financial system are less likely to collapse in isolation but, on the other hand, might occasionally become insolvent together.[2] In such settings, banks are hit by exogenous shocks; in our model, the severity of initial "shocks" hitting banks—bank runs—depends on the cross-bank interaction on the asset market. Without a feedback loop between asset prices and run decisions of banks' investors, whether heterogeneity is beneficial for financial stability crucially depends on relative costs of systemic and individual crashes and frequency of these events (i.e., distribution of shocks). For example, in Cabrales, Gottardi and Vega-Redondo (2017), segmented financial system is socially optimal only when shocks are sufficiently fat-tailed while full integration is optimal for thin-tailed shocks.[3] Relatedly, a few papers argue that the optimal degree of heterogeneity is state dependent, i.e. in good times it is better to pool banks not to expose the weaker ones to runs while in bad times it is better to differentiate banks to save the stronger ones (e.g. Bouvard, Chaigneau and Motta, 2015 and Liu, 2018). In our setting, in contrast, a certain degree of heterogeneity is beneficial for all banks irrespective of distributional assumptions on shocks.

The interaction between heterogeneity and financial fragility has also been analyzed in a very different context by Geanakoplos (2009) and Simsek (2013). They investigate how belief heterogeneity among investors affects aggregate borrowing and then transmit into economic fluctuations. Their focus is on whether the overall level of borrowing is

[2]See, among others, Shaffer (1994), Stiglitz (2010), Ibragimov, Jaffee and Walden (2011), Wagner (2010 and 2011), and Kopytov (2019). There is also a large empirical literature studying systemic risk in the interconnected financial systems, e.g. Billio, Getmansky, Lo and Pelizzon (2012), Adrian and Brunnermeier (2016), Acharya, Pedersen, Philippon and Richardson (2017), Brownlees and Engle (2017), Cai, Eidam, Saunders and Steffen (2018), and Baum, Grazzini and Schäfer (2019).

[3]An analogous result arises in Acemoglu, Ozdaglar and Tahbaz-Salehi (2015), who consider banks connected through interbank liabilities.

excessive while ours is about the fragility of demandable debt given the leverage. We also emphasize a different type of heterogeneity, i.e. that among bank asset returns, which can simply result from random realizations of idiosyncratic productivities.

Our paper contributes to the vast literature on financial fragility created by bank runs going back to Diamond and Dybvig (1983). Our model utilizes the global-games methodology, where investors observe noisy signals about the fundamentals when they decide whether to run or not. This allows us to define the probability of runs and the level of financial fragility, as in Rochet and Vives (2004) and Goldstein and Pauzner (2005).[4] We use this to study how the heterogeneity in the financial system affects the level of fragility. An important feature of our model is the two types of interacting strategic complementarities. This is related to Goldstein (2005), who highlights how bank runs and currency attacks can reinforce each other and generate twin crises, and Liu (2016), who demonstrates the feedback between bank runs and liquidity dry-up on the interbank market.[5] In our model, runs in different banks can amplify and feedback into one another through fire-sale externality in the common asset market. Our main conclusion builds on the fact that in a heterogeneous financial system, strong banks have a larger impact on weak banks than vice versa. This is related to Sákovics and Steiner (2012), who argue that, in a single-complementarity coordination game with heterogeneous agents, optimal subsidies should target agents imposing high externalities on others.[6] However, increasing heterogeneity does not matter in their framework because there are no interactions between different types of strategic complementarities.

Two papers in the global-games literature address questions more closely related to us. Choi (2014) builds a model of regime-switching game to ask whether it is better to bolster strong or weak banks that are subject to cross-bank complementarity. His model does not feature within-bank complementarity and only features a particular asymmetric configuration of cross-bank complementarity. Hence, the mechanism behind his results is

[4]See also Frankel, Morris and Pauzner (2003), Corsetti, Dasgupta, Morris and Shin (2004), Guimaraes and Morris (2007), and Sákovics and Steiner (2012), among others.

[5]Another related paper is Uhlig (2010), who presents a model where runs on individual banks are interrelated due to fire sales.

[6]Also related is Cong, Grenadier and Hu (2019) who argue that the regulator should allocate more resource to save small financial institutions, because saving small ones generate greater informational externalities on large ones. Shen and Zou (2018) show that it is cost-efficient to design intervention policies that screen agents based on their heterogeneous information in global games. Dai and Yang (2017) look into organizational design when heterogeneous agents face coordination problems. In their setting, organizations facilitate coordinated actions at the cost of agents with extreme preferences, which limits the sustainability and the size of organizations.

different from the one described above. Liu (2018) analyzes a setting where heterogeneous banks face run risk and interact in the asset market but there is no uncertainty about the liquidation price or the aggregate fundamental (see also Eisenbach, 2017). Hence, his conclusions about fragility are driven by equilibrium multiplicity. We show that incorporating aggregate uncertainty—a salient feature of economy-wide crises (Bloom, 2009)—not only resolves the multiplicity issue but also yields sharp predictions about how fragility depends on bank heterogeneity.

The remainder of this paper is organized as follows. In Section 2 we lay out our model. Section 3 presents our results about the relationship between heterogeneity and stability in a setting with two types of banks. Section 4 generalizes our baseline model to the cases of multiple bank types and imperfect signals about bank-specific productivities. Section 5 considers two applications of our model to financial regulation. Section 6 concludes.

2 Model

The economy is populated by three types of agents: banks, bank investors, and deep-pocketed outside investors. There are three periods, $t = 0, 1, 2$. The agents and events are described in detail below.

2.1 Banks

There is a continuum of banks indexed by $i \in [0, 1]$. Banks are ex ante homogeneous and have an investment capacity of one. At $t = 0$, each bank i collects one unit of capital from a unit mass of investors in the form of demandable debt, and makes long-term investment that generates a return of θ_i at $t = 2$. We call z_i the fundamental of bank i, and it takes the following form,

$$z_i = \theta + \eta_i,$$

where θ is the aggregate component shared by all banks, and η_i is the bank-specific component. The aggregate fundamental $\theta \sim F_\theta(\cdot)$ with a bounded support $[\underline{\theta}, \bar{\theta}]$, where $\bar{\theta} > \underline{\theta} > 0$. For simplicity, we assume two groups of banks: the *strong* with a mass of w and the *weak* with a mass of $1 - w$.[7] Hence, the bank-specific shock η_i follows the

[7] In Section 4.1, we show that our main results hold in a setting with $N \geq 2$ types of banks.

distribution below,

$$\eta_i = \begin{cases} \Delta_s & \text{with probability } w, \\ \Delta_w & \text{with probability } 1-w. \end{cases}$$

Without loss of generality, we assume $\Delta_s = \Delta \geq 0$ and $\Delta_w = -\frac{w}{1-w}\Delta \leq 0$ so that the bank-specific fundamental η_i has a zero mean. The size of bank-specific shocks is restricted to be such that the overall productivity is always positive for all banks, $\underline{\theta}+\Delta_w > 0$.

Both aggregate and bank-specific productivities are realized at $t = 1$ and bank investors then may choose to withdraw their funds early. Under such circumstances, bank i needs to repay one unit of capital to each runner and thus is forced to liquidate its long-term investment early in the asset market to fulfill the needs. The liquidation process and the investors' early withdraw decisions will be specified in the next two subsections.

2.2 Outside investors and the asset market

At $t = 1$, if a mass of m_i investors withdraw their funds early from bank i, bank i needs to raise funds of amount m_i by liquidating its long-term investment. This means that bank i has to liquidate $\frac{m_i}{p_i}$ fraction of its long-term investment position given the liquidation price p_i. Below we specify the asset market and characterize the market-clearing prices $\mathbf{p} \equiv \{p_i\}_{i\in[0,1]}$.

The asset market is competitive and populated with a unit mass of deep-pocketed outside investors. In line with classic works featuring fire sales (e.g. Shleifer and Vishny, 1992 and Kiyotaki and Moore, 1997), we assume that outside investors are less efficient in managing assets than banks. In particular, under banks' management, in the absence of premature liquidations, a portfolio $\{k_i\}_{i\in[0,1]}$ generates $y \equiv \int z_i k_i di$ at $t = 2$. In contrast, if the same portfolio is managed by outside investors, the return is subject to a discount: instead of receiving y, outside investors only get $f(y)$, where $f(y) < y$ for all $y > 0$ and $f(0) = 0$. In addition, we assume that $f'(\cdot) > 0$ and $f''(\cdot) < 0$ so that outsiders' inefficiency in production increases in the amount of assets they absorb. Furthermore, we assume that $yf'(y)$ is increasing in y to guarantee equilibrium uniqueness in the asset market at $t = 1$. These assumptions on $f(\cdot)$ are typical in the literature on fire sales (Lorenzoni, 2008).

Definition 1. *Given masses of early withdrawers* $\mathbf{m} = \{m_i\}_{i\in[0,1]}$ *and bank fundamentals* $\boldsymbol{z} = \{z_i\}_{i\in[0,1]}$*, an equilibrium in the asset market consists of outside investors' demand*

functions $\{k_i(\mathbf{p}, \boldsymbol{z})\}_{i\in[0,1]}$ *and market-clearing liquidation prices* $\mathbf{p} = \{p_i(\mathbf{m}, \boldsymbol{z})\}_{i\in[0,1]}$ *such that:*

1. *Given the liquidation prices* $\mathbf{p}$*, outside investors' demand functions* $\{k_i(\mathbf{p}, \boldsymbol{z})\}_{i\in[0,1]}$ *maximize their expected payoffs:*

$$\max_{\{k_i\}_{i\in[0,1]}} f\left(\int z_i k_i di\right) - \int p_i k_i di.$$

2. *The liquidation prices satisfy the market-clearing conditions:*

$$k_i = \frac{m_i}{p_i} \quad \forall i \in [0,1].$$

The key feature of the asset market is that fire-sale externalities can spill over across banks with heterogeneous fundamentals. The direct interpretation of the feature is that banks face the same group of buyers of their assets (e.g., hedge funds). Even if the asset markets for different banks are separated, arbitrage capital might flow across these markets, leading to comoving fire-sale discounts.

The following lemma summarizes key properties of liquidation prices $\mathbf{p}$.

Lemma 1. *Given masses of early withdrawers* $\mathbf{m}$ *and bank fundamentals* $\boldsymbol{z}$*, the equilibrium liquidation price for bank* i*'s assets is*

$$p_i = p(z_i, m) = \frac{z_i}{\lambda(m)}, \qquad \forall i \in [0,1],$$

where $m \equiv \int m_i di$ *is the total mass of early withdrawers in the economy, and* $\lambda(m)$ *is a strictly increasing function.*

Proof. See Appendix A.1. □

The liquidation prices of banks' assets are proportional to their productivities z_i's and are subject to a common discount factor $\lambda(m)$. The discount factor $\lambda(m)$ increases in the total mass of early withdrawers in the financial system. Intuitively, if more bank investors withdraw their funds early, outsiders have to absorb more assets. Since the inefficiency in outsiders' asset management increases in the amount assets they manage, i.e., $f''(\cdot) < 0$, the price discount factor $\lambda(m)$ becomes larger when more bank investors withdraw early.

Such a property of the liquidation prices reflects a simple idea that banks impose fire-sale pressure on others when liquidating their own assets.

2.3 Bank investors and runs

This section describes the behavior of bank investors. For each bank i, there is a unit mass of infinitesimal investors indexed by $j \in [0,1]$. At $t=0$, each investor j contributes one unit of capital to its bank. Note that banks are ex-ante homogeneous, therefore, investors are indifferent about which bank to invest in.

At $t=1$, after the realization of bank i's fundamental z_i, each investor j of bank i observes the bank-specific fundamental η_i[8] and receives a noisy private signal s_{ij} about the aggregate fundamental θ,[9]

$$s_{ij} = \theta + \sigma\epsilon_{ij},\ \epsilon_{ij} \sim \Phi(\cdot).$$

The signal noise has a cumulative distribution function $\Phi(\cdot)$, which is differentiable and strictly increasing on its support $[\underline{\epsilon}, \bar{\epsilon}]$. A corresponding probability density function is denoted by $\phi(\cdot)$. Such an information structure follows a conventional global games setup, which allows us to pin down a unique equilibrium.

With probability $\bar{m} \in (0,1)$, investor j is "non-sleepy" and may withdraw her funds from her bank. With probability $1-\bar{m}$, investor j is "sleepy" and neglects the option to withdraw early. Therefore, bank i needs to liquidate at most $\frac{\bar{m}}{p_i}$ fraction of its assets to fulfill the maximum amount of early withdrawal $\bar{m}$. For tractability, we rule out bank failures by assuming that $\frac{\bar{m}}{p_i} \leq 1$.[10]

As mentioned previously, early withdrawers are guaranteed to get their funds back at $t=1$. At $t=2$, the investment return of bank i is equally distributed among investors who

[8] In Section 4.2, we show that our analyses remain valid as long as investors receive partially informative signals about bank-specific components η_i's.

[9] This assumption allows us to highlight the strategic uncertainties both within and across banks, i.e. investors of bank i are not sure about mass of early withdrawals in their bank m_i and in the whole financial system m. In Appendix C, we analyze an alternative setup with perfect information about the aggregate fundamental θ. In that case, multiple equilibria are possible and policy implications are blurred.

[10] Goldstein and Pauzner (2005) show that bank failure at $t=1$ creates a region of strategic substitution. Incorporating this substitution region brings technical complications while not necessarily new insights. Therefore, following Frankel, Morris and Pauzner (2003) and Chen, Goldstein and Jiang (2010), we assume investors pay limited attention, and the amount of runners is capped by $\bar{m}$.

have not liquidate their funds early a la Diamond and Dybvig (1983). Let $u_i(a_{ij})$ denote a "non-sleepy" investor j's payoff conditional on her withdraw decision $a_{ij} \in \{run,\ stay\}$.

$$u_i(a_{ij}) = \begin{cases} 1 & \text{if } a_{ij} = run \\ \dfrac{z_i\left(1 - \frac{m_i}{p_i}\right)}{1 - m_i} & \text{if } a_{ij} = stay. \end{cases} \tag{1}$$

Plugging in the market-clearing liquidation price p_i derived in Lemma 1, we can express the incremental payoff from staying as

$$\pi(z_i, m_i, m) \equiv u_i\left(stay\right) - u_i\left(run\right) = \frac{z_i - m_i\lambda(m)}{1 - m_i} - 1.$$

The incremental payoff represents an investor's incentive to run on her bank, and a "non-sleepy" investor j of bank i runs at $t = 1$ if and only if expected incremental payoff given her signal s_{ij} is negative,

$$\mathbb{E}[\pi(z_i, m_i, p_i)|s_{ij}] < 0.$$

The expression above tells us that the model features two types of strategic complementarities. First, there is a standard *within-bank strategic complementarity* that an investor has more incentive to run if more investors in her own bank do so. On top of that, the fire-sale externalities in the asset market give rise to a *cross-bank strategic complementarity* that an investor has more incentive to run if more investors in other banks do so. Formally,

$$\frac{\partial \pi(z_i, m_i, m)}{\partial m_i} < 0 \quad \text{and} \quad \frac{\partial \pi(z_i, m_i, m)}{\partial m} < 0.$$

More importantly, the two complementarities—within- and cross-bank—feed and amplify each other,

$$\frac{\partial^2 \pi(z_i, m_i, m)}{\partial m_i \partial m} < 0. \tag{2}$$

Banks that encounter more early withdrawals (higher m_i's) have to liquidate more assets to repay runners. That naturally makes the long-term payoffs of staying investors more sensitive to fluctuations in the liquidation prices, which depend on the total mass of runners in the financial system m. Later in Section 3.2, we will discuss this feature in more details and explain its importance for our results.

2.4 Timeline and equilibrium definition

Figure 1 depicts the timeline of our model.

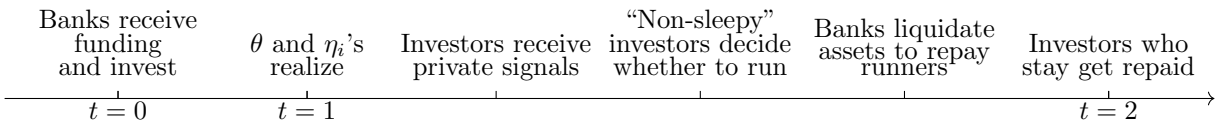


Figure 1: Timeline

Denote the strategy of a "non-sleepy" investor j in bank i as $\mathcal{A}_{ij}$ that maps bank-specific fundamental η_i and her private signal about aggregate fundamental s_{ij} to her action space $a_{ij} \in \{run, stay\}$; Outside investors' demand for bank i's assets $\{k_i(\mathbf{p}, \boldsymbol{z})\}_{i\in[0,1]}$ as functions of liquidation prices and bank fundamentals; Liquidation prices $\mathbf{p} = \{p_i(\mathbf{m}, \boldsymbol{z})\}_{i\in[0,1]}$ as functions of banks' fundamentals and masses of runners; Masses of runners $\mathbf{m} = \{m_i(\theta, \eta_i)\}_{i\in[0,1]}$ as functions of aggregate and bank-specific fundamentals.

Definition 2. *Bank investors' strategies, outside investors' demand, liquidation prices and total mass of runners constitute an equilibrium if*

1. *Given* $\mathbf{m}$ *and* $\boldsymbol{z}$, $\{k_i(\mathbf{p}, \boldsymbol{z})\}_{i\in[0,1]}$ *and* $\{p_i(\mathbf{m}, \boldsymbol{z})\}_{i\in[0,1]}$ *constitute a sub-game equilibrium in the asset market as in Definition 1;*

2. *Given* $\{p_i(\mathbf{m}, \boldsymbol{z})\}_{i\in[0,1]}$, *other investors' strategies resulting in* $\{m_i(\theta, \eta_i)\}_{i\in[0,1]}$, *each investor maximizes her expected payoff defined in Equation* (1) *conditional on her private signal and the bank-specific fundamental of her bank;*

3. $m_i(\theta, \eta_i) = \int \mathbb{I}\{a_{ij}(s_{ij}, \eta_i) = run\}\, dj$.

2.5 Global games and threshold equilibrium

In what follows, we focus on the limiting case of infinitely precise signals, $\sigma \to 0$. As we show in Appendix D, there exists a unique equilibrium, in which all bank investors follow threshold strategies. In particular, investor j of bank i run when $s_{ij} < \theta_i^*$ and stay when $s_{ij} > \theta_i^*$. An investor with signal $s_{ij} = \theta_i^*$ is indifferent between the two actions:

$$\int_0^1 \frac{\theta_i^* + \eta_i - \lambda\left(m(x)\right)\bar{m}x}{1 - \bar{m}x} dx = 1. \tag{3}$$

As in standard global-games models, a marginal investor receiving a threshold signal θ_i^* has a Laplacian belief. That is she perceives that the mass of runners within her own bank, $\bar{m}x = \bar{m}\Phi\left(\frac{\theta_i^* - \theta}{\sigma}\right)$, is uniformly distributed (Morris and Shin, 2001). Crucially, because her signal is informative about the aggregate productivity, she also makes inference about actions of investors of other banks. From her perspectives, the total mass of runners in the economy is

$$m(x) = \bar{m}\int \Phi\left(\frac{\theta_j^* - \theta}{\sigma}\right) dj = \bar{m}\int \Phi\left(\frac{\theta_j^* - \theta_i^*}{\sigma} + \Phi^{-1}(x)\right) dj, \tag{4}$$

where $\Phi^{-1}(0) = \underline{\epsilon}$ and $\Phi^{-1}(1) = \bar{\epsilon}$.

A set of equations (3) together with (4) deliver equilibrium thresholds θ_i^* for investors of all banks in the economy. Given the structure of idiosyncratic shocks, equations (3) and (4) can be rewritten as

$$\theta_s^* + \Delta = \frac{1}{\int_0^1 \frac{1}{1-\bar{m}x}dx}\left(1 + I_s(t)\right), \tag{5}$$

$$\theta_w^* - \frac{w}{1-w}\Delta = \frac{1}{\int_0^1 \frac{1}{1-\bar{m}x}dx}\left(1 + I_w(t)\right), \tag{6}$$

where $I_s(t)$ and $I_w(t)$ represent the fire-sale pressures on strong bank and weak bank investors, respectively:

$$I_s(t) \equiv \int_0^1 \lambda\left(w\bar{m}x + (1-w)\bar{m}\Phi\left(t + \Phi^{-1}(x)\right)\right)\frac{\bar{m}x}{1-\bar{m}x}dx, \tag{7}$$

$$I_w(t) \equiv \int_0^1 \lambda\left((1-w)\bar{m}x + w\bar{m}\Phi\left(-t + \Phi^{-1}(x)\right)\right)\frac{\bar{m}x}{1-\bar{m}x}dx, \tag{8}$$

and $t \equiv \lim_{\sigma\to 0}\frac{\theta_w^* - \theta_s^*}{\sigma}$ is the limiting distance between the two thresholds, which either takes a non-negative finite value or goes to infinity. If $t + \Phi^{-1}(x) \geq \bar{\epsilon}$ then threshold investors of strong banks assign probability one to the event that weak banks are going to experience runs and, thus, will have to liquidate assets prematurely. In this case, the fire-sale pressure exerted by weak banks on strong banks reaches its maximal level. Similarly, if $-t + \Phi^{-1}(x) \leq \underline{\epsilon}$ then the fire-sale pressure on weak banks from strong banks is minimized. At this point, threshold investors of weak banks consider runs on strong banks to be impossible.

2.6 Discussions

Standard bank-run models (e.g. Diamond and Dybvig, 1983 and Goldstein and Pauzner, 2005) normally focus on fragilities of individual banks and assume that the cost of premature liquidation of banks' assets is exogenous. The key distinction of our model is that the liquidation cost is set endogenously and is affected by the run behaviors of all bank investors in the economy. Our model, therefore, features mutually reinforcing within- and cross-bank complementarities (Equation 2). We argue that the interaction between the complementarities is detrimental to the stability of the banking system, and it is important to recognize the interaction when characterizing the fragility of individual banks as well as the whole banking system.

To clearly see how the interaction affects fragility, assume that the banking system is homogeneous, i.e. $\Delta = 0$. The indifference condition for threshold investors (3) is then

$$\int_0^1 \frac{\theta_0^* - \lambda(\bar{m}x)\bar{m}x}{1 - \bar{m}x} dx = 1,$$

where θ_0^* denotes the common threshold of all bank investors. In this case, the run behaviors of bank investors are synchronized. Therefore, a threshold investor expects many runs on her bank, i.e., large $\bar{m}x$, precisely at times when the fire-sale discount $\lambda(\bar{m}x)$ is high. This positive correlation between runs and fire sales exacerbates run problems within individual banks and makes the financial system more fragile.

Consider now a standard individual bank-run model where there is no such interaction, and an individual investor takes the level the fire-sale cost as given. When receiving a signal, she only updates her beliefs about the severity of the run problem within her own bank but not about the fire-sale discount. As a result, there is no correlation between runs and fire sales. The indifference condition is given by

$$\int_0^1 \frac{\hat{\theta}_0 - \bar{\lambda}\bar{m}x}{1 - \bar{m}x} dx = 1,$$

where the threshold is $\hat{\theta}_0$, and the average fire-sale cost is the same as in the previous example, $\bar{\lambda} = \int_0^1 \lambda(\bar{m}x)dx$. Despite the fact that the average fire-sale cost is the same in

both cases, banks are more fragile in the model with the interacting complementarities,[11]

$$\theta_0^* - \hat{\theta}_0 = \int_0^1 \frac{\left(\lambda\left(\bar{m}x\right) - \bar{\lambda}\right)\bar{m}x}{1-\bar{m}x}dx > 0.$$

Informally speaking, a homogeneous financial system is fragile because bank investors expect problems within their own banks and other banks to synchronize. Making banks heterogeneous is beneficial for the stability because it breaks down the simultaneity and alleviates the reinforcement between the within- and cross-bank complementarities. In the next section, we proceed to a formal analysis.

3 Analyses

In this section, we analyze our model presented in the last section. In this baseline setting, bank-specific fundamentals take only two values. In addition to making our analyses transparent, an advantage of having only two types is that we can map an increase in the degree of bank heterogeneity directly into an increase in Δ. In Section 3.1, we present our main result on the relation between bank heterogeneity and financial stability when investors of strong and weak banks exhibit nontrivial strategic interaction. In Section 3.2, we discuss the key economic force behind this result. In Section 3.3, we provide a full characterization of the relationship between heterogeneity and stability.

3.1 Bank heterogeneity and financial stability

In this section, we investigate how changes in bank asset heterogeneity affect financial stability, captured in the model by the two run thresholds θ_s^* and θ_w^*. Changes in heterogeneity affect bank investors through two channels. First, there is a direct impact on productivity of bank assets. Fixing the total amount of fire sales in the economy (i.e., total mass of runners m), an increase in Δ makes strong banks stronger and weak banks weaker. If the fire-sale terms $I_s(t)$ and $I_w(t)$ in Equations (5) and (6) were fixed, an increase in Δ undoubtedly increases the stability of strong banks (θ_s^* goes down) and reduces that of weak banks (θ_w^* goes up).

[11] To see the inequality, let x_0 be such that $\lambda(\bar{m}x_0) = \bar{\lambda}$. Then, $\lambda\left(\bar{m}x\right) - \bar{\lambda} < 0$ for $x \in [0, x_0)$, and $\lambda\left(\bar{m}x\right) - \bar{\lambda} > 0$ for $x \in (x_0, 1]$. Since $\frac{\bar{m}x}{1-\bar{m}x}$ increases in x, $\int_0^1 \frac{\left(\lambda(\bar{m}x)-\bar{\lambda}\right)\bar{m}x}{1-\bar{m}x}dx > \frac{\bar{m}x_0}{1-\bar{m}x_0}\int_0^1 \left(\lambda\left(\bar{m}x\right) - \bar{\lambda}\right)dx = 0$.

However, the total amount of fire sales is not fixed. When banks interact with each other in the asset market, the behavior of an individual investor will be affected not only by peers in her own bank but also by investors in other banks. When the increase in Δ shifts the two thresholds θ_s^* and θ_w^* away from each other, the distance between the thresholds t goes up. As a result, strong banks have to face a larger fire-sale pressure triggered by weak banks' higher likelihood to liquidate assets prematurely. In contrast, the fire-sale pressure on weak banks is alleviated. Therefore, the overall impact of heterogeneity on banks' stability becomes unclear: on the one hand, higher Δ directly improves strong banks' fundamentals, but on the other hand, it also lowers expectations of strong bank investors regarding the liquidation prices. Conversely, weak banks suffer from worse fundamentals but benefit from an alleviated fire-sale pressure.

Taking into account both within- and cross-bank strategic complementarities, the following proposition shows the key result of the paper: An increase in asset heterogeneity Δ is *Pareto improving* as long as investors of weak and strong banks remain uncertain about run behavior of each other.

Proposition 1. *Define*

$$\Delta_{min} = (1-w)\frac{1}{\int_0^1 \frac{dx}{1-\bar{m}x}}\left[I_s(\bar{\epsilon}-\underline{\epsilon}) - I_w(\bar{\epsilon}-\underline{\epsilon})\right].$$

$\forall \Delta \in (0, \Delta_{min})$*, we have*

(i) $\theta_w^*(\Delta) = \theta_s^*(\Delta) \equiv \theta^*(\Delta)$*;*

(ii) $\theta^*(\Delta)$ *decreases in* Δ*.*

Proof. See Appendix A.2. □

Δ_{min} represents the critical level of asset heterogeneity above which there is a trivial strategic interaction between investors of strong and weak banks. If $\Delta > \Delta_{min}$, the difference between the two types is so large that marginal investors in the weak banks know for sure that no one runs in strong banks ($m_i = 0$). Similarly, marginal investors in the strong banks know for sure that every "non-sleepy" investors runs in weak banks ($m_i = \bar{m}$). Therefore, when investors make their run decisions, they only need to make inference about investors in the same type of banks based on their private signals. As a result, $\theta_s^* < \theta_w^*$ and the limiting distance between the thresholds t becomes infinite. If

$\Delta < \Delta_{min}$, investors need to evaluate the run situations in both types of banks to make their run decisions. Since we have assumed that signals' noise is infinitesimal, this is possible only if the thresholds θ_s^* and θ_w^* are trivially different from each other—as what we have as result (i).[12]

Result (ii) of Proposition 1 states that when $\Delta < \Delta_{min}$ and thus strategic interactions across banks are nontrivial, the run thresholds not only stay together but also move downward as heterogeneity increases. Financial system with more heterogeneous banks is thus unambiguously more stable. This means, somewhat strikingly, that heterogeneity is beneficial even for weak banks, despite a worsening of their bank-specific fundamentals.

By setting $\theta^* = \theta_s^* = \theta_w^*$ and then substituting equation (6) into (5), we obtain:

$$\theta^*(\Delta) = \frac{1}{\int_0^1 \frac{dx}{1-\bar{m}x}} \left[w I_s(t(\Delta)) + (1-w) I_w(t(\Delta)) \right], \tag{9}$$

where $t(\Delta)$ is implicitly defined by

$$\Delta = (1-w) \frac{1}{\int_0^1 \frac{dx}{1-\bar{m}x}} \left[I_s(t(\Delta)) - I_w(t(\Delta)) \right]. \tag{10}$$

Recall that $I_s(t)$ and $I_w(t)$ represent the strengths of the fire-sale pressure in the asset market faced by strong and weak banks, respectively. These two terms, multiplied by the relative weights of strong and weak banks, determine the common run threshold, as shown in Equation (9). From (10), a marginal increase in Δ is associated with a positive change in the limiting distance t between the run thresholds. It also boosts the fire-sale pressure faced by strong banks $I_s(t)$ and alleviates that faced by weak banks $I_w(t)$. Result (ii) of Proposition 1 implies that $\theta^*(\Delta)$ declines, i.e., weak banks benefit more than strong banks suffer. The next section discusses economic forced behind this result.

[12] That the bank-type-specific thresholds cluster around the same value is a typical feature of global games with heterogeneous players and infinitely precise signals (e.g. Frankel, Morris and Pauzner, 2003). While we do rely on the assumption that $\sigma \to 0$ in order to analytically characterize the equilibrium, the main result of Proposition 1 does not require such an assumption. In Appendix B.3, we show via numerical examples that increasing heterogeneity Δ tends to make both weak and strong banks less fragile, albeit θ_s^* and θ_w^* are no longer infinitely close to each other anymore.

3.2 Role of two complementarities

The key force that lies behind result (ii) of Proposition 1 is that within- and cross-bank complementarities are mutually reinforcing, which in the model is captured by inequality (2). Weak banks are more sensitive to the cross-bank fire-sale externalities because they experience more severe bank runs and need to liquidate more assets. When $\Delta < \Delta_{min}$, the strategic interaction across different banks is nontrivial, and an increase in heterogeneity Δ alleviates cross-bank fire-sale externalities for the weak banks, and worsens those for the strong banks. However, since weak banks are more sensitive to the change, the benefit for weak banks outweighs the loss for strong banks, and the overall stability increases.

To see this point formally, we need to compare the (weighted) changes in the fire-sale pressures $I_w(t(\Delta))$ and $I_s(t(\Delta))$ due to an increase in heterogeneity Δ, as suggested by Equation (9). Recall that the limiting distance between the thresholds t is an increasing function of Δ: larger heterogeneity pushes the thresholds away from each other. Below, we describe how the expressions under integral terms in $I_w(t)$ and $I_s(t)$ change when t is increased by $dt > 0$.

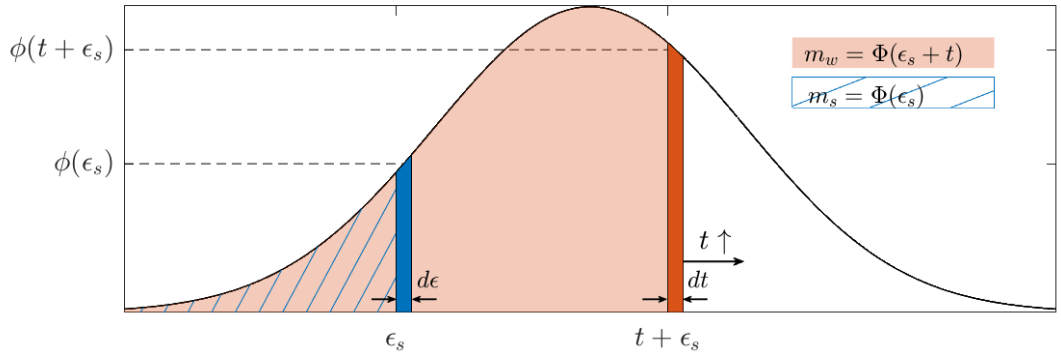
First, consider a strong bank's marginal investor receiving a threshold signal θ_s^*. She knows that if her noise realization is ϵ_s, all strong bank investors whose noise realizations are below ϵ_s are going to run (hatched area in Figure 2(a)). Therefore, the mass of runners in strong banks is $w\bar{m}\Phi(\epsilon_s)$.[13] At the same time, she expects a larger mass $(1-w)\bar{m}\Phi(t+\epsilon_s)$ of weak bank investors to run (filled area in 2(a)). Under the uniform prior about the mass of runners in her bank $\bar{m}\Phi(\epsilon_s) \sim U[0,\bar{m}]$, she assigns probability $\phi(\epsilon_s)d\epsilon$ to this event. Increasing the limiting distance between the run thresholds t by dt uniformly shifts her belief about weak investors' behavior and thus raises the perceived fraction of weak banks' runners by $(1-w)\bar{m}\phi(t+\epsilon_s)dt$. The increase in fragility, contributed by a larger fire-sale pressure faced by strong banks in this state, is given by:

$$\underbrace{w}_{\text{Weight of strong banks}} \times \underbrace{\lambda'\left(w\bar{m}\Phi(\epsilon_s)+(1-w)\bar{m}\Phi(t+\epsilon_s)\right)\times\frac{\bar{m}\Phi(\epsilon_s)}{1-\bar{m}\Phi(\epsilon_s)}}_{\text{Sensitivity of payoff to the amount of fire sales}}\times \tag{11}$$

$$\underbrace{\phi(\epsilon_s)d\epsilon}_{\text{Probability of state }\epsilon_s} \times \underbrace{(1-w)\bar{m}\phi\left(t+\epsilon_s\right)dt.}_{\text{Change in the mass of runners in weak banks}}$$

[13] In the expression for $I_s(t)$ (7), we change the variable of integration from x to ϵ, where $x = \Phi(\epsilon)$. Below, we do the same change of variable for $I_w(t)$ given in (8).

(a) Strong bank



(b) Weak bank

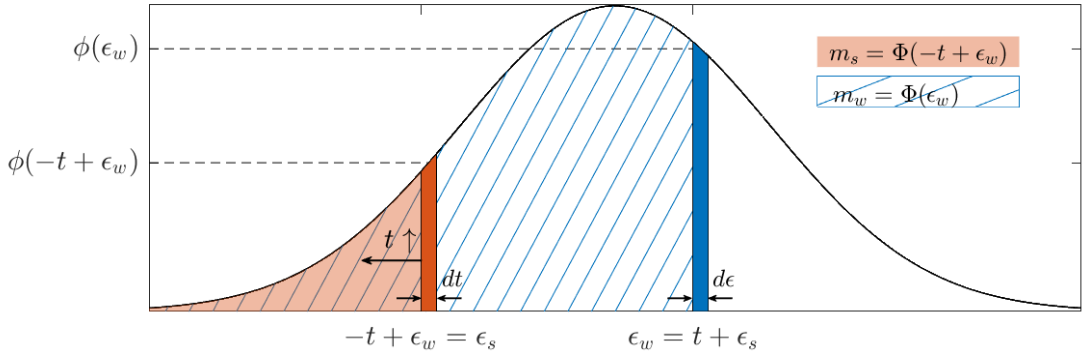


Figure 2: Impact of change in heterogeneity on the amount of fire sales

Integrating (11) over all possible realizations of noise ϵ, we obtain the full impact of elevated fire-sale pressure faced by strong banks on the overall financial fragility (i.e. the common run threshold θ^*).

Consider now a weak bank's investor receiving a threshold signal θ_w^*. From her perspective, probability of the state with the mass $w\bar{m}\Phi(\epsilon_s)$ of strong investors running is $\phi(\epsilon_w)d\epsilon{=}\phi(t+\epsilon_s)d\epsilon$, where $\epsilon_w \equiv t+\epsilon_s$. In panel (b) of Figure 2, this mass is depicted by a filled area. In this state of the world, she believes that the fraction $(1-w)\bar{m}\Phi(t+\epsilon_s)$ of weak banks' investors liquidate prematurely (hatched area in the same plot). A marginal increases in heterogeneity reduces the mass of strong banks' runners by $w\bar{m}\phi(\epsilon_s)dt$. The decrease of fragility, contributed by a smaller fire-sale pressure faced by weak banks in

this state, is given by:

$$\underbrace{1-w}_{\text{Weight of weak banks}} \times \underbrace{\lambda'\left(w\bar{m}\Phi(\epsilon_s)+(1-w)\bar{m}\Phi(t+\epsilon_s)\right)\times\frac{\bar{m}\Phi(t+\epsilon_s)}{1-\bar{m}\Phi(t+\epsilon_s)}}_{\text{Sensitivity of payoff to the amount of fire sales}}\times \tag{12}$$

$$\underbrace{\phi(t+\epsilon_s)d\epsilon}_{\text{Probability of state } \epsilon_s+t} \times \underbrace{w\bar{m}\phi\left(\epsilon_s\right)dt.}_{\text{Change in the mass of runners in strong banks}}$$

By comparing (11) and (12), one can observe that there is only one difference between strong and weak bank investors sharing the same belief about the total amount of early liquidations. Compared to weak bank investors, investors of strong banks experience a weaker within-bank complementarity problem because the mass of runners on their own bank is smaller, $\bar{m}\Phi(\epsilon_s) < \bar{m}\Phi(t+\epsilon_s)$ (hatched area is smaller in panel (a)). From condition (2), weak bank investors are therefore more sensitive to changes in t: for them, a change in the total amount of fire sales has a stronger impact on their payoff function.

Because (12) is larger than (11) state by state, i.e. for all possible realizations of noise, we have $|\partial I_w(t)/\partial t| > |\partial I_s(t)/\partial t|$. Therefore, an increase in the limiting distance t due to higher heterogeneity Δ leads to a decline in the common run threshold θ^*.

Again, this result is due to the fact that the within- and cross-bank complementarities reinforce each other. For our model, this property is a natural implication of the existence of asset fire sales at $t=1$ and a standard Diamond and Dybvig (1983) payoff structure. The following proposition formally establishes, with a more general payoff function, the importance of the mutually reinforcing complementarities for our results in Proposition 1.

Proposition 2. *Consider an investor of bank i whose net benefit of staying at $t=1$ is*

$$\pi(z_i, m_i, m) = z_i g_1(m_i) - g_2(m_i, m),$$

where $g_1(m_i)$ is increasing in m_i; $g_2(m_i, m)$ is increasing in both m_i and m; and η_i follows the same binary structure as in our model. If $\frac{\partial^2\pi}{\partial m\partial m_i} = -\frac{\partial^2 g_2}{\partial m\partial m_i} < 0$, then $\exists\ \Delta_{min} > 0$ such that $\forall\ \Delta \in (0, \Delta_{min})$, $\theta_s^*(\Delta) = \theta_w^*(\Delta) = \theta^*(\Delta)$, and $\theta^*(\Delta)$ is a decreasing function.

Corollary 1. *Under the assumptions of Proposition 2, heterogeneity does not affect the common threshold θ^* if $\frac{\partial^2\pi}{\partial m\partial m_i} = \frac{\partial^2 g_2}{\partial m\partial m_i} = 0$.*

Proof. See Appendix A.3. □

In the baseline setting, $g_1(m_i) = \frac{1}{1-m_i}$ and $g_2(m_i, m) = \frac{\lambda(m)m_i}{1-m_i}$, so $\frac{\partial^2 g_2}{\partial m \partial m_i} = \frac{\lambda'(m)}{1-m_i^2} > 0$.

Notice that if the cross-derivative $\frac{\partial^2 g_2}{\partial m \partial m_i} = 0$ then heterogeneity does not affect the common threshold θ^* (Corollary 1). This result echoes Sákovics and Steiner (2012). They show that in global games with heterogeneous agents, the weighted average strategic belief about the aggregate action (total amount of runs in the economy) is uniform. Moreover, in the absence of the interaction between the within- and cross-bank complementarities, only the weighted average belief about the aggregate action matters for the common threshold θ^*. Under the uniform average belief, the degree of heterogeneity across types does not affect θ^*.

When the two complementarities do interact, however, θ^* depends on the bank-type-specific interaction terms between the amounts of runs in the whole economy and within a particular bank. The powerful result of Sákovics and Steiner (2012) does not hold in this case, making the analysis much more cumbersome.

3.3 Heterogeneity and stability: a full characterization

In our baseline model with a binary structure of bank-specific fundamentals, we are able to analytically characterize the relation between financial stability (captured by the run thresholds θ_s^* and θ_w^*) and the degree of bank asset heterogeneity (Δ) when Δ goes beyond Δ_{min}. The following proposition, which nests Proposition 1 as case 1, characterizes fully the run thresholds as functions of bank heterogeneity Δ.

Proposition 3. $\exists\ \Delta_{max} > \Delta_{min} > 0$ *such that*

1. *If* $\Delta \in (0, \Delta_{min})$ *then* $\theta_s^* = \theta_w^* < \theta^*(0)$. $\theta_s^*(\Delta) = \theta_w^*(\Delta)$ *are decreasing functions.*
2. *If* $\Delta \in [\Delta_{min}, \Delta_{max})$ *then* $\theta_s^* < \theta_w^* < \theta^*(0)$. $\theta_s^*(\Delta)$ *and* $\theta_w^*(\Delta)$ *are decreasing and increasing functions, respectively.*
3. *If* $\Delta \geq \Delta_{max}$ *then* $\theta_s^* < \theta^*(0) \leq \theta_w^*$, *where the equality holds for* $\Delta = \Delta_{max}$. $\theta_s^*(\Delta)$ *and* $\theta_w^*(\Delta)$ *are decreasing and increasing functions, respectively.*

Proof. See Appendix A.2. □

Figure 3 depicts the results of Proposition 3. Case 1, in which $\Delta \in (0, \Delta_{min})$ and $\theta_s^* = \theta_w^*$, has been discussed in previous sections. Again, here the difference in fundamentals across

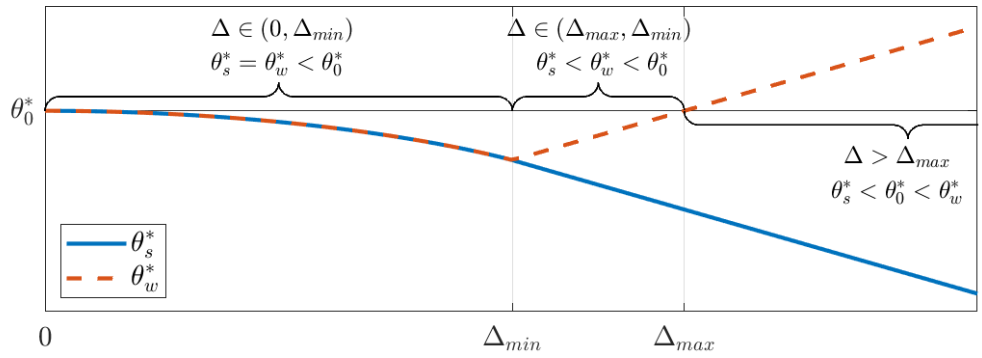


Figure 3: Run thresholds as functions of heterogeneity Δ

banks allows investors of strong and weak banks to have nontrivial strategic interactions. As bank heterogeneity increases, all banks become more stable.

In Case 2, when Δ goes beyond Δ_{min}, banks become sufficiently different. There is no strategic uncertainty across investors of different banks anymore: strong bank investors receiving a threshold signal θ_s^* are now certain that all weak bank investors are going to run and thus weak banks are going to liquidate their assets prematurely, and vice versa. θ_s^* and θ_w^* are no longer infinitely close to each other in equilibrium. A further increase in Δ does not affect the strength of the cross-bank complementarity but only further strengthens (weakens) fundamentals of strong (weak) banks. As a result, θ_s^* keeps declining while θ_w^* starts to rise. Therefore, a marginal increase in heterogeneity is no longer Pareto improving as weak banks become more fragile. However, when $\Delta \in [\Delta_{min}, \Delta_{max})$, we still have a more stable banking sector compared to one with homogeneous banks.

In Case 3, when $\Delta > \Delta_{max}$, heterogeneity becomes so large that θ_w^* exceeds the no-heterogeneity threshold $\theta^*(0)$. It is not clear whether the overall financial system is still more stable than in the case of full homogeneity: while strong banks are more stable, weak banks are more fragile.

Moving from stability to welfare is straightforward in our model. Since all agents are risk neutral, welfare can be measured as the expected output. Formally, consider a regulator who chooses the degree of heterogeneity to maximize

$$\max_{\Delta} \mathbb{E}\theta + \left[f\left(\bar{m}\lambda(\bar{m})\right) - \bar{m}\lambda(\bar{m})\right] F_\theta\left(\theta_s^*\right) + \tag{13}$$
$$\left[f\left((1-w)\bar{m}\lambda\left((1-w)\bar{m}\right)\right) - (1-w)\bar{m}\lambda\left((1-w)\bar{m}\right)\right]\left(F_\theta\left(\theta_w^*\right) - F_\theta\left(\theta_s^*\right)\right),$$

where the first term captures the overall economy's production before subtracting any fire-sale losses; the second term represents losses due to fire sales when $\theta < \theta_s^*$ and investors of strong and weak banks run; the third term represents losses due to fire sales when $\theta_s^* \leq \theta < \theta_w^*$ and only weak bank investors run.

The following corollary comes directly from Proposition 3:

Corollary 2. *It is never optimal to have a banking sector with asset heterogeneity* $\Delta < \Delta_{min}$.

Proof. Consider a banking sector with $\Delta < \Delta_{min}$. In this system, $\theta_s^* = \theta_w^* \equiv \theta^*$ and thus the regulator's objective is to minimize $F_\theta(\theta^*)$. When $\Delta < \Delta_{min}$, $\theta^*(\Delta)$ is a decreasing function. Therefore, it is never optimal to set Δ below Δ_{min}. □

Under the current criterion, it is less clear how welfare changes when Δ moves beyond Δ_{min}. On the one hand, strong banks keep becoming stronger and, thus, it is less likely to encounter systemic financial crises when all investors run. On the other hand, weak bank investors are more likely to run, and the probability of nonzero inefficient liquidations goes up. If the regulator's objective is to prevent as many runs from happening as possible, then the optimal level of heterogeneity should be set exactly to Δ_{min}. In that case, the regulator's objective boils down to $\min_\Delta F_\theta(\theta_w^*)$.

4 Robustness

In this section, we consider two extensions of our baseline model. We show that our main results are robustness to more than two types of banks and noisy information about bank-specific fundamentals.

4.1 Many types of banks

We consider a more general version of the model with $N \geq 2$ types of banks. When the number of types exceeds two, there are in principle many ways to adjust heterogeneity. Nevertheless, we are able establish two propositions that echo our results from Section 3. In particular, we first show that a homogeneous banking system is fragile, and *any*

sufficiently weak heterogeneity is beneficial for all banks. We then show that when heterogeneity becomes too strong, a further increase in heterogeneity is not unequivocally stabilizing and makes some banks more fragile.

The structure of the economy stays the same as in Section 2. The only difference is that now there are $N \geq 2$ groups of banks, where banks within the same group share idiosyncratic productivity $\eta_i = \Delta_i$. To fix notation, $\Delta_1 \leq \Delta_2 \leq \cdots \leq \Delta_N$, so that banks of group 1 have the weakest fundamentals and banks of group N have the strongest fundamentals. Weight of group i is $w_i \in (0,1)$, and by definition $\sum_{i=1}^{N} w_i = 1$. Without loss of generality, we set the average η_i to zero: $\sum_{i=1}^{N} w_i \Delta_i = 0$.

Bank investors follow threshold strategies, i.e. investors of bank i withdraw early if their signals are below θ_i^* and do not do so otherwise. An indifference condition for an investor receiving a threshold signal θ_i^* is

$$\theta_i^* + \Delta_i = \frac{1}{\int_0^1 \frac{1}{1-\bar{m}x} dx} \left(1 + \int_0^1 \lambda \left(\bar{m} \sum_{j=1}^{N} w_j \Phi \left(t_{ij} + \Phi^{-1}(x) \right) \right) \frac{\bar{m}x}{1-\bar{m}x} dx \right), \quad (14)$$

where $t_{ij} = \lim_{\sigma \to 0} \frac{\theta_j^* - \theta_i^*}{\sigma}$ is the limiting distance between the thresholds θ_i^* and θ_j^*. A system of N equations (14) is a generalized version of equations (5)-(6) of the baseline model with two types of banks.

When the banking system is fully homogeneous, $\Delta_i = 0$, $\forall i \in \{1, \ldots, N\}$. Naturally, all thresholds are then the same and $t_{ij} = 0$ for all (i,j) pairs. By continuity, all thresholds stays infinitely close to each other for any marginal change in the vector of idiosyncratic productivities. The limiting distances, however, adjust. In particular, if $\Delta_i > \Delta_k$ then $t_{ik} > 0$. Similarly to our discussion in Section 3.1, when heterogeneity is sufficiently weak, banks' behaviors on the asset market are not fully decoupled. For infinitely precise signals, it means that the distances between the thresholds have to be trivially small. The following proposition states that a homogeneous financial system is fragile: a common threshold $\theta^*(\Delta_1, \ldots, \Delta_N)$ reaches its (local) maximum when $\Delta_i = 0$ $\forall i \in \{1, \ldots, N\}$.

Proposition 4. *There exists a positive constant $\bar{\Delta} > 0$ and a non-empty set $U_{\bar{\Delta}} \in \mathbb{R}^N = \left\{ x | x_1 \leq x_2 \leq \cdots \leq x_N; \sum_{i=1}^{N} w_i x_i = 0; |x_i| < \bar{\Delta} \right\}$ such that $\forall \Delta \in U_{\bar{\Delta}}$, investors of all banks share the same threshold $\theta^*(\Delta)$. Moreover, $\theta^*(\Delta)$ reaches maximum at 0.*

Proof. See Appendix B.1. □

This result is an analogue of Proposition 1, and the intuition behind is the same. Differentiation of banks makes some of them relatively stronger (if $\Delta_i > 0$) and others relatively weaker (if $\Delta_i < 0$) ex post from the perspectives of asset performance. At the same time, the fire-sale pressure on weaker banks is alleviated, and that benefits them more than a higher fire-sale pressure hurts stronger banks. Again, this result hinges crucially on the fact that within- and cross-bank complementarities are mutually reinforcing.

Similarly to Proposition 3, increasing heterogeneity is not always unequivocally beneficial. In particular, when heterogeneity becomes sufficiently strong—in the sense that banks' behaviors on the asset market are fully decoupled and there is no strategic interactions across banks—further differentiation hurts banks with smaller Δ_i's.

Proposition 5. *There exists a positive constant $\Lambda > 0$ such that if $|\Delta_i - \Delta_k| > \Lambda$ $\forall i \neq k \in \{1, \ldots, N\}$ then all thresholds are different, $\theta_1^* < \theta_2^* < \cdots < \theta_N^*$. In this region, the threshold of bank i's investors θ_i^* depends only on bank i's productivity Δ_i. Moreover, $\theta_i^*(\Delta_i)$ is a decreasing function, and any change in bank heterogeneity increases the fragility of banks whose bank-specific fundamentals go down.*

Proof. See Appendix B.1. □

4.2 Noisy information about bank-specific fundamentals

In the baseline model, we assume that investors have perfect information about their banks' idiosyncratic fundamentals η_i. We now show that this assumption can be relaxed. As long as investors receive some useful information about η_i, the results of Proposition 1 go through.

Below, we consider an extension with noisy information about bank-specific fundamentals. In particular, investor j in bank i receives two distinct noisy signals. The first signal s_{ij} about the aggregate fundamental θ is the same as in the baseline model, i.e.,

$$s_{ij} = \theta + \sigma\epsilon_{ij}.$$

In addition, each bank i discloses bank-specific information, captured by a signal d_i about its bank-specific fundamental η_i, to its investors. d_i takes two values: G and B, with the

probability mass function below.

$$\mathbb{P}(d_i = G|\eta_i = \Delta) = \mathbb{P}\left(d_i = B \,\middle|\, \eta_i = -\frac{w}{1-w}\Delta\right) = \alpha,$$

where $\alpha \in \left[\frac{1}{2}, 1\right]$.

Denote the posterior belief about the probability of bank i being strong as p_G if $d_i = G$ and p_B if $d_i = B$. Then

$$p_G = \frac{w\alpha}{w\alpha + (1-w)(1-\alpha)} \geq w \geq p_B = \frac{w(1-\alpha)}{w(1-\alpha) + (1-w)\alpha}.$$

The equality holds if and only if $\alpha = \frac{1}{2}$, i.e., the signal d_i is uninformative about bank-specific fundamentals. Another special case is $\alpha = 1$, which is the setting in our baseline model. The proposition below generalizes Proposition 1.

Proposition 6. *The model with imperfect information about bank-specific fundamentals is equivalent to the benchmark model where bank-specific shocks take values $\Delta^{eff}(\alpha)$ with probability $w^{eff}(\alpha)$ and $-\Delta^{eff}(\alpha)\frac{w^{eff}(\alpha)}{1-w^{eff}(\alpha)}$ with probability $1 - w^{eff}(\alpha)$, where*

$$w^{eff}(\alpha) = w\alpha + (1-w)(1-\alpha),$$
$$\Delta^{eff}(\alpha) = \frac{2\alpha - 1}{w\alpha + (1-w)(1-\alpha)} w\Delta.$$

Corollary 3. *(i) If signals about bank-specific fundamentals are uninformative, $\alpha = \frac{1}{2}$, the banking system is effectively homogeneous, and $\theta^*_G(\Delta) = \theta^*_B(\Delta) = \theta^*(0)$ for any Δ.*

(ii) If signals about bank-specific fundamentals are informative, $\alpha \in (\frac{1}{2}, 1]$, there exists a $\Delta_{min}(\alpha) > 0$ such that $\theta^*_G(\Delta) = \theta^*_B(\Delta) < \theta^*(0)$ and $\theta^*_G(\Delta)$ is a decreasing function on $\Delta \in (0, \Delta_{min}(\alpha))$.

Proof. See Appendix B.2. □

The key result following Proposition 6 is that investors' perception of bank heterogeneity—rather than physical heterogeneity per se—matters for financial stability. As we show in Corollary 3, if signals about bank-specific fundamentals are not informative, then bank investors may perceive banks with very different asset holdings as homogeneous. In the extreme case of completely uninformative signals, $\alpha = \frac{1}{2}$, the effective heterogeneity

is zero, $\Delta^{eff}\left(\frac{1}{2}\right) = 0$. Run decisions of investors of strong and weak banks are then completely synchronized, and any changes in physical heterogeneity Δ do not affect run thresholds. If signals are informative, $\alpha > \frac{1}{2}$, investors are able to differentiate between banks. In that case, run decisions of investors are not perfectly synchronized.

5 Discussions and policy implications

Our analysis so far has been focused on how financial fragility depends on the degree of asset commonality in a comparative static fashion. We intentionally leave out banks' asset choices—that is, the distribution of bank-specific productivities η_i is exogenously given. This is because, to maximize transparency and highlight our key mechanism, we choose not to include some important incentives for banks to hold correlated assets, such as diversification purposes when depositors are risk-averse, or expected ex post government interventions (Acharya, 2009, Acharya and Yorulmazer, 2007 and 2008, Farhi and Tirole, 2012).[14] In either case, banks as price takers will not take into account their impact on asset prices when making individual portfolio choices. For example, while it is Pareto-improving to boost strong banks at the expense of weak banks in our setting, weak banks will not agree to support strong ones voluntarily. Existence of such pecuniary externality calls for policy interventions.

Below we consider model implications for two policies widely discussed in the aftermath of the recent global financial crisis—ring-fencing (Section 5.1) and resolution of distressed banks (Section 5.2).

5.1 Ring-fencing

In response to the "too-big-to-fail" problem, the idea of "ring-fencing" has emerged in the post-crisis policy discussions worldwide. "Ring-fencing" refers to separating a large bank's balance sheet and restricting fund reallocation across ring-fenced subsidiaries. The separation can take place according to the service divisions. For example, in the US, the Volcker Rule restricts proprietary trading by commercial banks, essentially spinning off their investment banking activities. Similarly, starting from January 1, 2019, the largest

[14] Appendix A.3 considers a general setting where banks' objective function is not linear in payoffs (e.g. due to risk aversion of banks' investors), and outlines sufficient conditions under which homogeneity is detrimental for stability.

UK banks are required to separate core businesses in retail banking from investment banking.[15] Besides service divisions, the separation can be carried out according to geographic locations. For example, the Fed required foreign banking organizations with more than $50 billion in US subsidiary assets to put all their US subsidiaries under an intermediate holding company (Kreicher and McCauley, 2018). Geographic ring-fencing has also been pursued by the European regulator through increased capital and liquidity requirements on foreign-owned subsidiaries, legal restrictions on intragroup cross-border asset transfers and limitations on the distribution of profits by foreign-owned subsidiaries (Enria, 2018).

To illustrate the effect of ring-fencing using our framework, we compare two cases. In the first case, the economy has only one large bank whose productivity equals to the aggregate productivity θ. In the second case, the large bank is divided up into different ring-fenced subsidiaries $i \in [0, 1]$. Each ring-fenced subsidiary has a productivity $z_i = \theta + \eta_i$, where $\eta_i \in \{\Delta_s, \Delta_w\}$. This corresponds to our setup in Section 2.

The comparison of the two cases yields an immediate result that dividing up the single bank into homogeneous subsidiaries has no effect. The corollary below formalizes the result. In words, simply splitting big banks into identical clones cannot reduce fragility in the financial system.[16]

Corollary 4. *If $\eta_i = 0\ \forall i$, the run thresholds for investors of all subsidiaries stay the same, i.e., $\theta_i^* = \theta^*(0)\ \forall i$.*

Proof. Corollary follows directly from Propositions 1 and 3. □

A typical argument for ring-fencing made by regulators is that it allows for balance sheet separation of different subsidiaries and, therefore, limits the degree of contagion a bank's failure imposes on the remaining financial system. However, this argument neglects the contagion through fire sales. Behaviors of homogeneous subsidiaries in the asset market are still synchronized. Therefore, fire-sale externalities across these subsidiaries essentially resemble the bank-run externalities across a large group of investors in the parent bank in the absence ring-fencing. The goal of ring-fencing should not simply be to downsize

[15] Ring-fencing was first introduced through the Financial Services (Banking Reform) Act 2013, followed by details set in further legislation. See a summary on GOV.UK at https://www.gov.uk/government/publications/ring-fencing-information/ring-fencing-information.

[16] In his public lecture, Brunnermeier (2010) makes an argument that echoes our result—"too big to fail" is not the same as "too systemic to fail".

banks' balance sheets but rather to achieve an optimal level of heterogeneity in the financial system.

The analyses and comparison of the two cases directly follow Section 3.3. In particular, following Corollary 2, the optimal degree of heterogeneity is $\Delta \geq \Delta_{min}$, beyond which cross-subsidiary strategic uncertainties vanish. To implement the optimal level of heterogeneity, the regulator could utilize different policy tools. In the example of ring-fencing that separate commercial banks from investment banks, the regulator could alter Δ with different classifications of bank assets.[17] In the example of geographic ring-fencing, the regulator could control Δ by varying the intensity of cross-region restrictions or changing the boundary or granularity of geographic divisions. Our model implies that regulators should utilize these policy tools to make sure that the ring-fenced subsidiaries are different enough such that when the weak subsidiaries are on the margin of failure, the strong ones are surviving with certainty. In this scenario, when the weak subsidiaries investors decide whether to withdraw their funds prematurely, they expect relatively high liquidation prices and, thus, are less likely to run. This, in turn, is beneficial for the overall financial stability.

5.2 Crises resolution

Our results also shed light on the optimal structuring of bank resolution policies. Within our framework, a cash injection by the regulator into bank i is equivalent to raising its bank-specific productivity. Indeed, the net benefit of not running for investors of bank i is

$$\frac{\theta + \eta_i - m_i\lambda(m) + \tau_i}{1 - m_i} - 1,$$

where τ_i is the amount of cash the government injects in bank i at $t = 2$. It is straightforward to see that a uniform support for banks (i.e. $\tau_i = \tau \ \forall i$) might not be the most efficient way to support the financial system. While such a uniform policy does make all banks more sound fundamentally, the heterogeneity across different banks stays the same. The regulator should also aim to achieve the optimal level of heterogeneity to alleviate negative externalities that banks impose on each other. When there exist nontrivial

[17] For example, legislation of the Volcker Rule went through several rounds of amendments, discussing what type of investment activities that commercial banks are prohibited to conduct. Indeed, the final version excluded some securities from the Rule.

strategic interactions across investors in different types of banks, i.e. $\Delta < \Delta_{min}$, a more active support for strong banks, even at the expense of reduced support for weak banks, benefits all financial institutions and thus the overall financial stability. Appendix A.4 formalizes this point. It states the problem of a planner allocating a bailout fund $T \geq 0$ across strong and weak banks so that $T = w\tau_s + (1-w)\tau_w$. We show (see Proposition 7), in particular, that when the bailout fund is not too large and bank investors exhibit nontrivial strategic interactions, the planner should support *only* strong banks.

In this regard, it is interesting to relate our results with the implementation of the Troubled Asset Relief Program (TARP) in 2008. The US Treasury effectively forced all major banks, irrespective of whether they were financially sound, to take the TARP money.[18] While some argue that forcing strong banks to also take the money can reduce the stigma around government intervention, our model provides a new dimension in which such action might be socially more desirable than supporting only relatively weaker banks. In the latter case, capital injections might lead to a reduction in effective heterogeneity and exacerbate cross-bank externalities.

Choi (2014) also argues that it might be more efficient to bolster strong banks. His results, however, rely on a specific model structure such that fragility of strong banks affect weak banks on the margin but not vice versa. Such cases do not arise in our model.[19] In fact, we show that when investors' payoffs follow a standard structure à la Diamond and Dybvig (1983), either both types of banks affect each other on the margin ($\Delta < \Delta_{min}$) or neither of them do ($\Delta \geq \Delta_{min}$). When different types of banks do interact nontrivially, their mutual impacts are not symmetric: strong banks impose more pronounced externalities on weak banks due to the reinforcement between the within- and cross-bank complementarities.

[18] For example, a former Wells Fargo CEO Dick Kovacevich discussed how the U.S. Treasury and the Federal Reserve were threatening major banks to take TARP money even if they didn't want it. See at https://www.cnbc.com/2013/09/13/tarp-ruined-banks-former-wells-fargo-ceo-kovacevich-says.html.

[19] Choi (2014) considers a regime-switching game with a binary payoff structure. In his setting, what matters for fragility is the total mass of runners in the economy when the value of the fundamental hits the type-specific regime-switching point. In our case, there is no such regime-switching points; fragility is shaped by investors' expectations about all possible realizations of the fundamental. In addition, Choi (2014) focuses on the case when bank heterogeneity Δ is large but below Δ_{min}. In his model, if bank heterogeneity is smaller, it might be optimal to support either type of banks or only weak banks. The policy prescriptions are, therefore, blurred.

6 Conclusion

In this paper, we analyze a model where banks not only face conventional risks of early withdrawals by their creditors, but also interact with one another in the asset market. We show that a homogeneous banking system is fragile because it suffers from the mutual reinforcement between two complementarities—runs on individual banks and fire-sale externality in the asset market. One may think that increasing heterogeneity could hurt some banks as their asset performances can become weaker. However, under fairly general assumptions, we show that maintaining a certain level of heterogeneity enhances the stability of *all* banks. By considering ring-fencing and distress resolution policies, we highlight the importance of our mechanism for the design of regulatory tools.

To focus on the analysis of complex interactions between probabilities of panics and asset prices in an interconnected financial system, we have left ex ante banks' portfolio choices out of consideration. In the future work, our setting can be incorporated into a richer dynamic model, featuring nontrivial asset and leverage choices.

References

Acemoglu, Daron, **Asuman Ozdaglar** and **Alireza Tahbaz-Salehi**, "Systemic Risk and Stability in Financial Networks", *American Economic Review*, 2015, *105*(2), 564–608.

Acharya, Viral V., "A Theory of Systemic Risk and Design of Prudential Bank Regulation", *Journal of Financial Stability*, 2009, *5*(3), 224–255.

Acharya, Viral V., **Lasse H. Pedersen**, **Thomas Philippon** and **Matthew Richardson**, "Measuring Systemic Risk", *Review of Financial Studies*, 2017, *30*(1), 2–47.

Acharya, Viral V. and **Tanju Yorulmazer**, "Too Many to Fail - An Analysis of Time-Inconsistency in Bank Closure Policies", *Journal of Financial Intermediation*, 2007, *16*(1), 1–31.

Acharya, Viral V. and **Tanju Yorulmazer**, "Cash-in-the-market Pricing and Optimal Resolution of Bank Failures", *Review of Financial Studies*, 2008, *21*(6), 2705–2742.

Adrian, Tobias and **Markus K. Brunnermeier**, "CoVaR", *American Economic Review*, 2016, *106*(7), 1705–1741.

Baum, Christopher F., **Caterina Forti Grazzini** and **Dorothea Schäfer**, "Banking Sector Diversity and Bank Stability", Unpublished Working Paper, 2019.

Billio, Monica, **Mila Getmansky**, **Andrew W. Lo** and **Loriana Pelizzon**, "Econometric Measures of Connectedness and Systemic Risk in the Finance and Insurance Sectors", *Journal of Financial Economics*, 2012, *104*(3), 535–559.

Bloom, Nicholas, "The Impact of Uncertainty Shocks", *Econometrica*, 2009, *77*(3), 623–685.

Bouvard, Matthieu, **Pierre Chaigneau** and **Adolfo De Motta**, "Transparency in the Financial System: Rollover Risk and Crises", *Journal of Finance*, 2015, *70*(4), 1805–1837.

Brownlees, Christian and **Robert F. Engle**, "SRISK: A Conditional Capital Shortfall Measure of Systemic Risk", *Review of Financial Studies*, 2017, *30*(1), 48–79.

Brunnermeier, Markus K., "Efficient Market Hypothesis, Bubbles and Liquidity", *Public lecture at Kings College, Cambridge University*, 2010.

Cabrales, Antonio, **Piero Gottardi** and **Fernando Vega-Redondo**, "Risk Sharing and Contagion in Networks", *Review of Financial Studies*, 2017, *30*(9), 3086–3127.

Cai, Jian, **Frederik Eidam**, **Anthony Saunders** and **Sascha Steffen**, "Syndication, Interconnectedness, and Systemic Risk", *Journal of Financial Stability*, 2018, *34*, 105–120.

Chen, Qi, **Itay Goldstein** and **Wei Jiang**, "Payoff Complementarities and Financial Fragility: Evidence from Mutual Fund Outflows", *Journal of Financial Economics*, 2010, *97*(2), 239–262.

Choi, Dong Beom, "Heterogeneity and Stability: Bolster the Strong, Not the Weak", *Review of Financial Studies*, 2014, *27*(6), 1830–1867.

Cong, Lin William, **Steven R. Grenadier** and **Yunzhi Hu**, "Dynamic interventions and informational linkages", *Journal of Financial Economics*, 2019.

Corsetti, Giancarlo, **Amil Dasgupta**, **Stephen Morris** and **Hyun Song Shin**, "Does One Soros Make a Difference? A Theory of Currency Crises with Large and Small Traders", *Review of Economic Studies*, 2004, *71*(1), 87–113.

Dai, Liang and **Ming Yang**, "Organizations and Coordination in a Diverse Population", *Ssrn*, 2017.

Diamond, Douglas W. and **Philip H. Dybvig**, "Bank Runs, Deposit Insurance, and Liquidity", *Journal of Political Economy*, 1983, *91*(3), 401–419.

Eisenbach, Thomas M., "Rollover risk as market discipline: A two-sided inefficiency", *Journal of Financial Economics*, 2017, *126*(2), 252–269.

Enria, Andrea, "Fragmentation in Banking Markets: Crisis Legacy and the Challenge of Brexit", *Speech delivered at BCBS-FSI High Level Meetign for Europe on Banking Supervision*, 2018.

Farhi, Emmanuel and **Jean Tirole**, "Collective Moral Hazard, Maturity Mismatch, and Systemic Bailouts", *American Economic Review*, 2012, *102*(1), 60–93.

Frankel, David M., **Stephen Morris** and **Ady Pauzner**, "Equilibrium Selection in Global Games with Strategic Complementarities", *Journal of Economic Theory*, 2003, *108*(1), 1–44.

Geanakoplos, John, "The Leverage Cycle", *NBER Macroeconomics Annual*, 2009, *24*(February), 1–65.

Goldstein, Itay, "Strategic Complementarities and the Twin Crises", *Economic Journal*, 2005, *115*(503), 368–390.

Goldstein, Itay and **Ady Pauzner**, "Demand-Deposit Contracts and the Probability of Bank Runs", *Journal of Finance*, 2005, *60*(3), 1293–1327.

Guimaraes, Bernardo and **Stephen Morris**, "Risk and Wealth in a Model of Self-Fulfilling Currency Attacks", *Journal of Monetary Economics*, 2007, *54*(8), 2205–2230.

Haldane, Andrew G., "Rethinking the Financial Network", *Speech delivered at the Financial Student Association, Amsterdam*, 2009.

Haldane, Andrew G. and **Robert M. May**, "Systemic Risk in Banking Ecosystems", *Nature*, 2011, *469*(7330), 351–355.

Ibragimov, Rustam, **Dwight Jaffee** and **Johan Walden**, "Diversification Disasters", *Journal of Financial Economics*, 2011, *99*(2), 333–348.

Kiyotaki, Nobuhiro and **John Moore**, "Credit Cycles", *Journal of Political Economy*, 1997, *105*(2), 211–248.

Kopytov, Alexandr, "Booms, Busts, and Common Risk Exposures", Unpublished Working Paper, 2019.

Kreicher, Lawrence L. and **Robert N. McCauley**, "The New US Intermediate Holding Companies: Reducing or Shifting Assets?", *BIS Quarterly Review*, 2018.

Liu, Xuewen, "Interbank market freezes and creditor runs", *Review of Financial Studies*, 2016, *29*(7), 1860–1910.

Liu, Xuewen, "Diversification and Systemic Bank Runs", Unpublished Working Paper, 2018.

Lorenzoni, Guido, "Inefficient Credit Booms", *Review of Economic Studies*, 2008, *75*(3), 809–833.

Morris, Stephen and **Hyun Song Shin**, "Global Games: Theory and Application Stephen", Unpublished Working Paper, 2001.

Rochet, Jean-Charles and **Xavier Vives**, "Coordination Failures and the Lender of Last Resort: Was Bagehot Right After All?", *Journal of the European Economic Association*, 2004, *2*(6), 1116–1147.

Sákovics, József and **Jakub Steiner**, "Who Matters in Coordination Problems?", *American Economic Review*, 2012, *102*(7), 3439–3461.

Shaffer, Sherrill, "Pooling Intensifies Joint Failure Risk", *Research in Financial Services*, 1994, *6*, 249–280.

Shen, Lin and **Junyuan Zou**, "Intervention with Screening in Global Games", Unpublished Working Paper, 2018.

Shleifer, Andrei and **Robert W. Vishny**, "Liquidation Values and Debt Capacity : A Market Equilibrium Approach", *Journal of Finance*, 1992, *47*(4), 1343–1366.

Simsek, Alp, "Belief Disagreements and Collateral Constraints", *Econometrica*, 2013, *81*(1), 1–53.

Stiglitz, Joseph E., "Risk and Global Economic Architecture: Why Full Financial Integration May be Undesirable", *American Economic Review: Papers and Proceedings*, 2010, *100*(2).

Uhlig, Harald, "A Model of a Systemic Bank Run", *Journal of Monetary Economics*, 2010, *57*(1), 78–96.

Wagner, Wolf, "Diversification at Financial Institutions and Systemic Crises", *Journal of Financial Intermediation*, 2010, *19*(3), 373–386.

Wagner, Wolf, "Systemic Liquidation Risk and the Diversity-Diversification Trade-Off", *Journal of Finance*, 2011, *66*(4), 1141–1175.

Yellen, Janet, "Interconnectedness and Systemic Risk: Lessons from the Financial Crisis and Policy Implications", *Speech delivered at the AEA/AFA Joint Luncheon, San Diego*, 2013.

A Proofs

A.1 Liquidation price

We prove Lemma 1 in this section.

Proof. The first order condition of the outside investors' problem implies that liquidation price of bank i's assets p_i is proportional to its fundamental z_i, i.e.,

$$p_i = \frac{\partial f(y)}{\partial y} z_i \quad \forall i \in [0,1], \tag{15}$$

where $y \equiv \int z_i k_i di$. After imposing the market clearing conditions, $k_i = \frac{m_i}{p_i}$ $\forall i \in [0,1]$, we obtain

$$m_i = z_i k_i \frac{\partial f(y)}{\partial y} \Rightarrow m \equiv \int m_i di = y \frac{\partial f(y)}{\partial y}.$$

Since by assumption $yf'(y)$ is an increasing function of y, there is a unique solution y to the equation above. We denote the unique solution as $y = h(m)$, where $h'(\cdot) > 0$. Plug this into (15) and we obtain

$$p_i(z_i, m) = z_i \frac{m}{h(m)} = \frac{z_i}{\lambda(m)},$$

where $\lambda(m) \equiv \frac{h(m)}{m}$.

Notice that the liquidation price is a decreasing function of the total mass of early withdrawers m. Indeed, using (15), we can write

$$p_i(z_i, m) = z_i \frac{\partial f(y)}{\partial y}\bigg|_{y=h(m)} \Rightarrow \frac{\partial p_i}{\partial m} = z_i \frac{\partial^2 f(y)}{\partial y^2}\bigg|_{y=h(m)} \times \frac{\partial h(m)}{\partial m} < 0.$$

□

A.2 Main results on heterogeneity and stability

We prove Propositions 1 and 3 in this section.

Proof. The thresholds for investors of strong and weak banks are implicitly defined by

Equations (5) and (6).

$$\int_0^1 \frac{\theta_s^* + \Delta - \lambda\left(w\bar{m}x + (1-w)\bar{m}\Phi\left(t + \Phi^{-1}(x)\right)\right)\bar{m}x}{1-\bar{m}x}dx = 1, \tag{16}$$

$$\int_0^1 \frac{\theta_w^* - \frac{w}{1-w}\Delta - \lambda\left((1-w)\bar{m}x + w\bar{m}\Phi\left(-t + \Phi^{-1}(x)\right)\right)\bar{m}x}{1-\bar{m}x}dx = 1, \tag{17}$$

where $t \equiv \lim_{\sigma\to 0}\frac{\theta_w^* - \theta_s^*}{\sigma}$.

Notice that $t > 0$ if $\Delta > 0$. If $\Delta = 0$ then $t = 0$ and $\theta_s^* = \theta_w^* = \theta^*(0)$, where $\theta^*(0)$ is a common threshold shared by investors of all banks in the absence of any heterogeneity across banks.

Case 1: Weak heterogeneity, $0 \leq \Delta < \Delta_{min}$.

First, consider the case that $0 \leq t < \bar{\epsilon} - \underline{\epsilon}$. Then investors of strong banks receiving threshold signal θ_s^* assign nonzero probability to the event that investors of weak banks will not withdraw funds early: $\exists \underline{x} \in (0,1] : \ \forall x \in [0, \underline{x}], \Phi\left(t + \Phi^{-1}(x)\right) < 1$. Similarly, investors of weak banks receiving threshold signal θ_w^* assign nonzero probability to the event that investors of strong banks will withdraw funds early: $\exists \bar{x} \in [0,1) : \ \forall x \in [\bar{x}, 1], \Phi\left(-t + \Phi^{-1}(x)\right) > 0$.

For finite values of t, it is true that thresholds θ_s^* and θ_w^* converge to the same value θ^* in the limit: $\theta^* = \lim_{\sigma\to 0}\theta_w^* = \lim_{\sigma\to 0}\theta_s^*$. Essentially, weak and strong banks are equally fragile. Equate θ_s^* and θ_w^* and take the difference between Equations (5) and (6) to obtain

$$\frac{1}{1-w}\Delta\int_0^1 \frac{dx}{1-\bar{m}x} = I_s(t) - I_w(t), \tag{18}$$

where

$$I_s(t) = \int_0^1 \lambda\left(w\bar{m}x + (1-w)\bar{m}\Phi\left(t + \Phi^{-1}(x)\right)\right)\frac{\bar{m}x}{1-\bar{m}x}dx, \tag{19}$$

$$I_w(t) = \int_0^1 \lambda\left((1-w)\bar{m}x + w\bar{m}\Phi\left(-t + \Phi^{-1}(x)\right)\right)\frac{\bar{m}x}{1-\bar{m}x}dx. \tag{20}$$

Define Δ_{min} as

$$\begin{aligned}\Delta_{min} =&(1-w)\frac{1}{\int_0^1 \frac{dx}{1-\bar{m}x}}\left(I_s(\bar{\epsilon}-\underline{\epsilon})-I_w(\bar{\epsilon}-\underline{\epsilon})\right)= \\ &(1-w)\frac{1}{\int_0^1 \frac{dx}{1-\bar{m}x}}\int_0^1 \frac{\left[\lambda\left(w\bar{m}x+(1-w)\bar{m}\right)-\lambda\left((1-w)\bar{m}x\right)\right]\bar{m}x}{1-\bar{m}x}dx.\end{aligned}$$

Since $I_s(t)$ increases in t, and $I_s(t)$ decreases in t, it is easy to verfity that $t(\Delta)$ is a continuous and increasing function on $[0,\Delta_{min})$. Moreover, $t(0)=0$ and $\lim_{\Delta\to\Delta_{min}^-} t(\Delta)=\bar{\epsilon}-\underline{\epsilon}$.

We now show that the common threshold θ^* is decreasing in t if $t\in(0,\bar{\epsilon}-\underline{\epsilon})$. In other words, in the region of weak heterogeneity, additional heterogeneity—captured by an increase in Δ—leads to a less fragile financial system. Eliminating Δ from (16) and (17), we obtain

$$\theta^*(t)=\frac{1}{\int_0^1 \frac{dx}{1-\bar{m}x}}\left[wI_s(t)+(1-w)I_w(t)\right],$$

where $I_s(t)$ and $I_w(t)$ are given by (19) and (20). Rewrite $I_s(t)$ as

$$\begin{aligned}I_s(t)=&\int_{\underline{\epsilon}}^{\bar{\epsilon}-t}\lambda\left(w\bar{m}\Phi(\epsilon)+(1-w)\bar{m}\Phi\left(t+\epsilon\right)\right)\frac{\bar{m}\Phi(\epsilon)}{1-\bar{m}\Phi(\epsilon)}d\Phi(\epsilon)+\\ &\int_{\bar{\epsilon}-t}^{\bar{\epsilon}}\lambda\left(w\bar{m}\Phi(\epsilon)+(1-w)\bar{m}\right)\frac{\bar{m}\Phi(\epsilon)}{1-\bar{m}\Phi(\epsilon)}d\Phi(\epsilon),\end{aligned}$$

where we change the variable of integration $x=\Phi(\epsilon)$. Differentiating $I_s(t)$ with respect to t, we obtain

$$\frac{\partial I_s}{\partial t}=(1-w)\bar{m}\int_{\underline{\epsilon}}^{\bar{\epsilon}-t}\lambda'\left(w\bar{m}\Phi(\epsilon)+(1-w)\bar{m}\Phi\left(t+\epsilon\right)\right)\frac{\bar{m}\Phi(\epsilon)}{1-\bar{m}\Phi(\epsilon)}\phi(t+\epsilon)\phi(\epsilon)d\epsilon.$$

Similarly, derivative of $I_w(t)$ with respect to t is

$$\begin{aligned}\frac{\partial I_w}{\partial t}&=-w\bar{m}\int_{\underline{\epsilon}+t}^{\bar{\epsilon}}\lambda'\left((1-w)\bar{m}\Phi(\epsilon)+w\bar{m}\Phi\left(-t+\epsilon\right)\right)\frac{\bar{m}\Phi(\epsilon)}{1-\bar{m}\Phi(\epsilon)}\phi(-t+\epsilon)\phi(\epsilon)d\epsilon\\ &=-w\bar{m}\int_{\underline{\epsilon}}^{\bar{\epsilon}-t}\lambda'\left(w\bar{m}\Phi(\epsilon)+(1-w)\bar{m}\Phi\left(t+\epsilon\right)\right)\frac{\bar{m}\Phi(t+\epsilon)}{1-\bar{m}\Phi(t+\epsilon)}\phi(t+\epsilon)\phi(\epsilon)d\epsilon.\end{aligned}$$

The last equality is obtained by changing the variable of integration, $\epsilon\to\epsilon+t$.

It is now easy to see that

$$\frac{\partial \theta^*}{\partial t} = \frac{1}{\int_0^1 \frac{dx}{1-\bar{m}x}} \left[w \frac{\partial I_s}{\partial t} + (1-w) \frac{\partial I_w}{\partial t} \right] < 0$$

whenever $t \in (0, \bar{\epsilon} - \underline{\epsilon})$ because $\frac{\bar{m}x}{1-\bar{m}x}$ is an increasing function of x. Moreover, for $t = 0$ $\frac{\partial \theta^*}{\partial t} = 0$.

We have shown that if $\Delta \in (0, \Delta_{min})$ then $t(\Delta)$ and $\theta^*(t)$ are increasing and decreasing functions of their arguments, respectively. Therefore, $\theta^*(t(\Delta))$ is decreasing in Δ in the weak-heterogeneity region.

Case 2: Medium heterogeneity, $\Delta_{min} \leq \Delta < \Delta_{max}$.

If $\Delta > \Delta_{min}$ then (18) does not have solutions. Therefore, thresholds θ_s^* and θ_w^* do not converge to the same value when $\sigma \to 0$. Equations (16) and (17) can be rewritten as

$$\int_0^1 \frac{\theta_s^* + \Delta - \lambda\left(w\bar{m}x + (1-w)\bar{m}\right)\bar{m}x}{1-\bar{m}x} dx = 1,$$

$$\int_0^1 \frac{\theta_w^* - \frac{w}{1-w}\Delta - \lambda\left((1-w)\bar{m}x\right)\bar{m}x}{1-\bar{m}x} dx = 1,$$

Clearly, when $\Delta > \Delta_{min}$, $\theta_s^*(\Delta)$ is decreasing in Δ while $\theta_w^*(\Delta)$ is increasing in Δ. However, both thresholds are below the common threshold in the absence of any heterogeneity, $\theta^*(0)$, as long as $\Delta < \Delta_{max}$, where Δ_{max} is defined as

$$\Delta_{max} = \frac{1-w}{w} \frac{1}{\int_0^1 \frac{dx}{1-\bar{m}x}} \left[\int_0^1 \left[\lambda(\bar{m}x) - \lambda((1-w)\bar{m}x)\right] \frac{\bar{m}x}{1-\bar{m}x} dx \right].$$

Case 3: Strong heterogeneity, $\Delta \geq \Delta_{max}$.

Finally, if $\Delta > \Delta_{max}$ then θ_w^* surpasses $\theta^*(0)$, and we have $\theta_s^* < \theta^*(0) < \theta_w^*$.

□

A.3 Role of two complementarities

We prove Proposition 2 in this section.

Consider a generalized payoff function $g(z_i, m_i, m)$ of an investor of bank i that chooses

not to withdraw her funds early. It depends on her bank's productivity z_i, mass of early withdrawers from her bank m_i, and overall mass of early withdrawers in the whole economy m. Somewhat abusing notation, in this appendix, we denote partial derivatives of the $g(\cdot,\cdot,\cdot)$ function by subscripts. We assume that $g_1 \equiv \frac{\partial g}{\partial z_i} > 0$, $g_2 \equiv \frac{\partial g}{\partial m_i} < 0$, $g_3 \equiv \frac{\partial g}{\partial m} < 0$.

We focus on the case of weak heterogeneity, when $t \equiv \lim_{\sigma\to 0} \frac{\theta_w^* - \theta_s^*}{\sigma} \in (0, \bar{\epsilon} - \underline{\epsilon})$ and $\theta^* = \lim_{\sigma\to 0} \theta_w^* = \lim_{\sigma\to 0} \theta_s^*$. The analogues of Equations (5) and (6) from the main text are

$$\eta(\theta^*, \Delta, t) = \begin{pmatrix} \eta^1(\theta^*, \Delta, t) \\ \eta^2(\theta^*, \Delta, t) \end{pmatrix} \equiv \begin{pmatrix} \int_0^1 g\left(\theta^* + \Delta, \bar{m}x, w\bar{m}x + (1-w)\bar{m}\Phi\left(t + \Phi^{-1}(x)\right)\right) dx - 1 \\ \int_0^1 g\left(\theta^* - \frac{w}{1-w}\Delta, \bar{m}x, (1-w)\bar{m}x + (1-w)\bar{m}\Phi\left(-t + \Phi^{-1}(x)\right)\right) dx - 1 \end{pmatrix} = 0.$$

Define

$$m_{tot}(x, t, w) = w\bar{m}x + (1-w)\bar{m}\Phi\left(t + \Phi^{-1}(x)\right),$$

and notice that

$$m_{tot}(x, t, w) > \bar{m}x > m_{tot}(x, -t, 1-w). \tag{21}$$

Compute Jacobian of η with respect to θ^* and Δ:

$$J = \begin{pmatrix} \int_0^1 g_1\left(\theta^* + \Delta, \bar{m}x, m_{tot}(x, t, w)\right) dx & \int_0^1 g_1\left(\theta^* + \Delta, \bar{m}x, m_{tot}(x, t, w)\right) dx \\ \int_0^1 g_1\left(\theta^* - \frac{w\Delta}{1-w}, \bar{m}x, m_{tot}(x, -t, 1-w)\right) dx & \frac{-w}{1-w}\int_0^1 g_1\left(\theta^* - \frac{w\Delta}{1-w}, \bar{m}x, m_{tot}(x, -t, 1-w)\right) dx \end{pmatrix}.$$

Notice that $J = \{J_{ij}\}_{i,j=1}^2$ is invertible, and its determinant $D \equiv J_{11}J_{22} - J_{12}J_{21} < 0$. Moreover,

$$J^{-1} = \frac{1}{D}\begin{pmatrix} J_{22} & -J_{12} \\ -J_{21} & J_{11} \end{pmatrix}.$$

By the implicit function theorem there exists an interval $U \in \mathbb{R}$ containing t such that

there exists unique $h : U \in \mathbb{R}^2$ such that $h^1(t) = \theta^*$, $h^2(t) = \Delta$, and $\eta(\theta^*(t), \Delta(t), t) = 0$ $\forall t \in U$. Moreover, partial derivatives of h can be computed as

$$\begin{pmatrix} \frac{\partial h^1}{\partial t} \\ \frac{\partial h^2}{\partial t} \end{pmatrix} = -J^{-1} \begin{pmatrix} \frac{\partial \eta^1}{\partial t} \\ \frac{\partial \eta^2}{\partial t} \end{pmatrix}.$$

Partial derivatives of η with respect to t are computed in the same fashion as in Appendix A.2:

$$\frac{\partial \eta}{\partial t} = \begin{pmatrix} (1-w) \int_0^{\Phi(\bar{\epsilon}-t)} g_3\left(\theta^* + \Delta, \bar{m}x, m_{tot}(x,t,w)\right) \phi\left(t + \Phi^{-1}(x)\right) dx \\ -w \int_{\Phi(\underline{\epsilon}+t)}^1 g_3\left(\theta^* - \frac{w}{1-w}\Delta, \bar{m}x, m_{tot}(x,-t,1-w)\right) \phi\left(-t + \Phi^{-1}(x)\right) dx \end{pmatrix}.$$

It is easy to see that $\frac{\partial h^2}{\partial t} > 0$:

$$\frac{\partial h^2}{\partial t} = -\underbrace{\frac{1}{D}}_{<0} \left(-\underbrace{J_{21}}_{>0} \underbrace{\frac{\partial \eta^1}{\partial t}}_{<0} + \underbrace{J_{11}}_{>0} \underbrace{\frac{\partial \eta^2}{\partial t}}_{>0} \right) > 0.$$

$\Delta(t) = h^2(t)$ is an increasing function. This result is natural: higher heterogeneity corresponds to larger distance between the thresholds $t = \lim_{\sigma \to 0} \frac{\theta_w^* - \theta_s^*}{\sigma}$.

The sign of $\frac{\partial h^1}{\partial t}$ is in principle unclear,

$$\frac{\partial h^1}{\partial t} = -\underbrace{\frac{1}{D}}_{<0} \left(\underbrace{J_{22}}_{<0} \underbrace{\frac{\partial \eta^1}{\partial t}}_{<0} - \underbrace{J_{12}}_{>0} \underbrace{\frac{\partial \eta^2}{\partial t}}_{>0} \right),$$

The following lemma provides sufficient conditions for $\frac{\partial h^1}{\partial t} < 0$.

Lemma 2. *If cross-derivatives $g_{23} \leq 0$, $g_{13} \geq 0$, $g_{11} \geq 0$, with one of the inequalities holding strictly, then $\frac{\partial h^1}{\partial t} < 0$.*

Proof. Notice that

$$\frac{\partial h^1}{\partial t} \propto -\frac{J_{12}}{J_{22}} \frac{\frac{\partial \eta^2}{\partial t}}{\frac{\partial \eta^1}{\partial t}} + 1,$$

where $A \propto B$ means that A and B are the same up to a positive multiplicative term.

Write $-\frac{J_{12}}{J_{22}}$ as

$$-\frac{J_{12}}{J_{22}} = \frac{\int_0^1 g_1\left(\theta^* + \Delta, \bar{m}x, m_{tot}(x,t,w)\right) dx}{\frac{w}{1-w}\int_0^1 g_1\left(\theta^* - \frac{w}{1-w}\Delta, \bar{m}x, m_{tot}(x,-t,1-w)\right) dx} \overset{g_{13}\geq 0}{\geq}$$
$$\frac{\int_0^1 g_1\left(\theta^* + \Delta, \bar{m}x, m_{tot}(x,-t,1-w)\right) dx}{\frac{w}{1-w}\int_0^1 g_1\left(\theta^* - \frac{w}{1-w}\Delta, \bar{m}x, m_{tot}(x,-t,1-w)\right) dx} \overset{g_{11}\geq 0}{\geq} \frac{1-w}{w},$$

where the first inequality holds due to (21).

Next, consider $\frac{\frac{\partial \eta^2}{\partial t}}{\frac{\partial \eta^1}{\partial t}}$:

$$\frac{\frac{\partial \eta^2}{\partial t}}{\frac{\partial \eta^1}{\partial t}} = \frac{-w\int_{\Phi(\underline{\epsilon}+t)}^1 g_3\left(\theta^* - \frac{w}{1-w}\Delta, \bar{m}x, m_{tot}(x,-t,1-w)\right)\phi\left(-t+\Phi^{-1}(x)\right)dx}{(1-w)\int_0^{\Phi(\bar{\epsilon}-t)} g_3\left(\theta^*+\Delta, \bar{m}x, m_{tot}(x,t,w)\right)\phi\left(t+\Phi^{-1}(x)\right)dx}$$
$$= \frac{-w\int_{\Phi(\underline{\epsilon}+t)}^1 g_3\left(\theta^* - \frac{w}{1-w}\Delta, \bar{m}x, m_{tot}(x,-t,1-w)\right)\phi\left(-t+\Phi^{-1}(x)\right)dx}{(1-w)\int_{\Phi(\underline{\epsilon}+t)}^1 g_3\left(\theta^*+\Delta, \bar{m}\Phi\left(-t+\Phi^{-1}(x)\right), m_{tot}(x,-t,1-w)\right)\phi\left(-t+\Phi^{-1}(x)\right)dx}$$
$$\overset{g_{13}\geq 0}{\leq} \frac{-w\int_{\Phi(\underline{\epsilon}+t)}^1 g_3\left(\theta^*+\Delta, \bar{m}x, m_{tot}(x,-t,1-w)\right)\phi\left(-t+\Phi^{-1}(x)\right)dx}{(1-w)\int_{\Phi(\underline{\epsilon}+t)}^1 g_3\left(\theta^*+\Delta, \bar{m}\Phi\left(-t+\Phi^{-1}(x)\right), m_{tot}(x,-t,1-w)\right)\phi\left(-t+\Phi^{-1}(x)\right)dx}$$
$$\overset{g_{23}\leq 0}{\leq} \frac{-w\int_{\Phi(\underline{\epsilon}+t)}^1 g_3\left(\theta^*+\Delta, \bar{m}x, m_{tot}(x,-t,1-w)\right)\phi\left(-t+\Phi^{-1}(x)\right)dx}{(1-w)\int_{\Phi(\underline{\epsilon}+t)}^1 g_3\left(\theta^*+\Delta, \bar{m}x, m_{tot}(x,-t,1-w)\right)\phi\left(-t+\Phi^{-1}(x)\right)dx} = -\frac{w}{1-w}.$$

Therefore, $-\frac{J_{12}}{J_{22}}\frac{\frac{\partial \eta^2}{\partial t}}{\frac{\partial \eta^1}{\partial t}} + 1 \leq -\frac{1-w}{w}\frac{w}{1-w} + 1 \leq 0 \Rightarrow \frac{\partial h^1}{\partial t} \leq 0$. It is easy to see that if one of the inequalities $\{g_{23} \leq 0, g_{13} \geq 0, g_{11} \geq 0\}$ holds strictly, then $\frac{\partial h^1}{\partial t} < 0$. □

$\frac{\partial h^1}{\partial t} = \frac{\partial \theta^*(t)}{\partial t} < 0$ and $\frac{\partial h^2}{\partial t} = \frac{\partial \Delta^*(t)}{\partial t} > 0$ together imply that stronger heterogeneity reduces fragility of all banks in the economy. As shown in Lemma 2, there are three sufficient conditions that guarantee this result.

Intuitively, the direct impact of an increase in heterogeneity is that strong banks become stronger and weak banks become weaker. If the payoff function is convex in productivity, $g_{11} > 0$, then the increase in heterogeneity benefits strong banks disproportionately more than it hurts weak banks. At the same time, higher heterogeneity implies that investors of weak banks assign lower probability that strong banks will fire sale. Therefore, expected payoffs of weak banks' long-term technology increases due to $g_3 < 0$. Conversely, investors of strong banks assign higher probability that weak banks will have to liquidate

prematurely, which reduces expected payoffs of strong banks for the same reason. If $g_{13} > 0$ then benefit from reduction in the expected amount of fire sales from the perspective of weak bank investors outweighs the cost faced by strong bank investors.

Finally, $g_{23} < 0$ means that within- and cross-bank complementarities reinforce each other. As we discuss in the main text, an increase in heterogeneity simultaneously diminishes expected fire sales from the perspective of weak bank investors and rises expected fire sales from the perspective of strong bank investors. Under $g_{23} < 0$ the former effect dominates. Analogously to the baseline case, under the same beliefs about the total amount of premature liquidations, weak bank investors expect more premature liquidations within their banks than strong bank investors. Therefore, if $g_{23} < 0$ then the adjustment in the amounts of premature liquidations due to stronger heterogeneity benefits weak banks more than it hurts strong banks.

Notice that in the micro-founded case considered in the main text $g_{11} = 0$, $g_{13} = 0$, and $g_{23} < 0$. Therefore, the crucial underlying economic mechanism behind Propositions 1 and 2 is mutually reinforcing within- and cross-bank complementarities.

A.4 Crises resolution

Consider a social planner who has access to a bailout fund of exogenous size $T > 0$. At $t = 0$, the planner commits to a policy according to which bank of type i receives a nonnegative cash injection $\tau_i \geq 0$. We assume that she can provide different support to different types of banks but cannot discriminate banks within the same type. The planner's objective is to maximize expected output (or, equivalently, minimize fire-sale losses as in (13)):

$$\begin{aligned}
&\max_{\tau_s,\tau_w} \mathbb{E}\theta + \left[f\left(\bar{m}\lambda(\bar{m})\right) - \bar{m}\lambda(\bar{m})\right] F_\theta\left(\theta_s^*\right) + \\
&\qquad \left[f\left((1-w)\bar{m}\lambda\left((1-w)\bar{m}\right)\right) - (1-w)\bar{m}\lambda\left((1-w)\bar{m}\right)\right]\left(F_\theta\left(\theta_w^*\right) - F_\theta\left(\theta_s^*\right)\right), \\
&\text{s.t. } \tau_s w + \tau_w(1-w) = T,\ \tau_s \geq 0,\ \tau_w \geq 0,
\end{aligned}$$

where the run thresholds θ_s^* and θ_w^* are given by

$$\int_0^1 \frac{\theta_s^* + \Delta + \tau_s - \lambda\left(w\bar{m}x + (1-w)\bar{m}\Phi\left(t + \Phi^{-1}(x)\right)\right)\bar{m}x}{1 - \bar{m}x} dx = 1,$$

$$\int_0^1 \frac{\theta_w^* - \frac{w}{1-w}\Delta + \tau_w - \lambda\left((1-w)\bar{m}x + w\bar{m}\Phi\left(-t + \Phi^{-1}(x)\right)\right)\bar{m}x}{1-\bar{m}x}dx = 1.$$

Define $\hat{\tau}_i = \tau_i - T \geq -T$. Then, using the planner's budget constraint, $\hat{\tau}_w = -\hat{\tau}_s \frac{w}{1-w}$. Clearly, $w\hat{\tau}_s + (1-w)\hat{\tau}_w = 0$. Therefore, the planner provides a uniform support to all banks of size T and, at the same time, changes the degree of bank heterogeneity by simultaneously varying $\hat{\tau}_s$ and $\hat{\tau}_w$. As a result of the intervention, the effective degree of heterogeneity is $\Delta + \hat{\tau}_s$. The maximum level of heterogeneity the planner is able to achieve is, however, bounded above by $\Delta + \frac{1-w}{w}T$ because she is not allowed to tax banks (i.e., $\tau_w \geq 0$). Following the discussion in Section 3.3, the planner should aim to set it to at least Δ_{min}.

Proposition 7. *Given the size of the bailout fund T and the degree of bank heterogeneity Δ, the planner's optimal support policy is as follows.*

(i) Small bailout fund and weak heterogeneity, $\Delta + \frac{1-w}{w}T < \Delta_{min}$: Only strong banks receive support, $\tau_s = \frac{1}{w}T$ and $\tau_w = 0$. Effective heterogeneity is $\Delta + \frac{1-w}{w}T < \Delta_{min}$.

(ii) Large bailout fund and/or strong heterogeneity, $\Delta_{min} \leq \Delta + \frac{1-w}{w}T$: Strong banks receive a support $\tau_s \geq \max\{0, \Delta_{min} - \Delta + T\}$. Effective heterogeneity is at least Δ_{min}.

Proposition 7 parallels with Corollary 2. As long as there is strategic uncertainty across investors of different banks, it is beneficial to increase heterogeneity between them. The planner is able to achieve it by a disproportionate support of strong banks. However, bolstering strong banks too much (so that $\Delta + \hat{\tau}_s > \Delta_{min}$) might be socially harmful because it might result in excessively fragile weak banks.

B Robustness

B.1 Many types of banks

We start by proving Proposition 4.

Proof. Denote by t_{ij} the limiting difference between banks' thresholds:

$$t_{ij} = \lim_{\sigma \to 0} \frac{\theta_j^* - \theta_i^*}{\sigma}.$$

In case of weak heterogeneity, $t_{ij} < \bar{\epsilon} - \underline{\epsilon} \ \ \forall i, j$, so that banks share the same threshold θ^* in the limit of infinitely precise signals ($\sigma \to 0$).

The existence of the weak heterogeneity region—i.e. the existence of $\bar{\Delta} > 0$ and the set $U_{\bar{\Delta}}$—follows from two facts. First, when $\Delta_i = 0 \ \forall i \in \{1, \ldots, N\}$, all investors share the same threshold $\theta_i^* = \theta^*(0)$ and $t_{ij} = 0 \ \forall i, j \in \{1, \ldots, N\}$. Second, $\lambda(\cdot)$ and $\Phi(\cdot)$ are continuous functions.

From (14), we obtain the expression for the common threshold θ^*:

$$\theta^* = -\Delta_i + \frac{1}{\int_0^1 \frac{1}{1-\bar{m}x} dx} \left(1 + \int_0^1 \lambda \left(\bar{m} \sum_j w_j \Phi \left(t_{ij} + \Phi^{-1}(x) \right) \right) \times \frac{\bar{m}x}{1 - \bar{m}x} dx \right).$$

Notice that t_{kj} can be written as $t_{kj} = t_{ij} - t_{ik}$, hence

$$\Delta_i > \Delta_k \Leftrightarrow t_{ik} > 0 \text{ and } \Delta_i = \Delta_k \Leftrightarrow t_{ik} = 0. \tag{22}$$

Since $\sum_{i=1} w_i \Delta_i = 0$, the common threshold can be written as

$$\theta^* = \frac{1}{\int_0^1 \frac{1}{1-\bar{m}x} dx} \left(1 + \int_0^1 \left\{ \frac{\bar{m}x}{1 - \bar{m}x} \times \sum_i \left[w_i \lambda \left(\bar{m} \sum_j w_j \Phi \left(t_{ij} + \Phi^{-1}(x) \right) \right) \right] \right\} dx \right).$$

Denote

$$I(t_{12}, \ldots, t_{1N}) = \int_0^1 c(x) \times \sum_i w_i \lambda \left(\bar{m} \sum_j w_j \Phi \left(t_{ij} + \Phi^{-1}(x) \right) \right) dx,$$

where $c(x) = \frac{\bar{m}x}{1-\bar{m}x}$, $c'(x) > 0$. Notice that any t_{kj} can be written using only $\{t_{12}, \ldots, t_{1N}\}$ because $t_{kj} = t_{1j} - t_{1k}$ and $t_{jj} = 0$. In what follows, we prove that $I(t_{12}, \ldots, t_{1N})$ reaches its maximum at zero by proving that zero is the only critical point and that the Hessian is negative definite at zero.

Necessary conditions

We first find critical points by equating all first-order partial derivatives to zero:

$$\frac{\partial I}{\partial t_{1k}} = \bar{m} w_k \sum_i w_i J_{ik}, \text{ where } J_{ik} = \int_0^1 \left[c\left(\Phi\left(-t_{ik} + \Phi^{-1}(x)\right)\right) - c(x) \right] \times \lambda' \left(\bar{m} \sum_j w_j \Phi\left(t_{kj} + \Phi^{-1}(x)\right) \right) \times \phi\left(-t_{ik} + \Phi^{-1}(x)\right) dx.$$

It is easy to see that $t_{jk} = 0 \ \forall j, k \in \{1, \ldots, N\}$ is a critical point. Notice that it is also the only critical point. Assume not and banks are not fully homogeneous. Then consider $k = N$, where by assumption $\Delta_N \geq \Delta_i \ \forall i$. Because there is some heterogeneity and by (22), $t_{iN} \leq 0 \ \forall i$ and $\exists j \neq N$ such that $t_{jN} < 0$. Consequently, $J_{iN} \geq 0 \ \forall i$ and $J_{jN} > 0$. Hence $\frac{\partial I}{\partial t_{1N}} > 0$, which is a contradiction.

Sufficient conditions

The Hessian matrix of $I(t_{12}, \ldots, t_{1N})$ at zero is

$$H = -B \times \begin{bmatrix} w_2(1-w_2) & -w_2 w_3 & \ldots & -w_2 w_N \\ -w_3 w_2 & w_3(1-w_3) & \ldots & -w_3 w_N \\ \ldots & \ldots & \ldots & \ldots \\ -w_N w_2 & -w_N w_3 & \ldots & w_N(1-w_N) \end{bmatrix},$$

where $B = \bar{m} \times \int_0^1 f'(x) \times \left[\phi\left(\Phi^{-1}(x)\right)\right]^2 \times \lambda'(\bar{m}x)\, dx > 0$. The Hessian matrix is negative definite because

$$x^T H x = -B \left\{ \sum_{i=2}^N w_i x_{i-1}^2 - \left(\sum_{i=2}^N w_i x_{i-1} \right)^2 \right\} < 0$$

for any $x \in \mathbb{R}^{N-1} \backslash \{0\}$. Indeed, consider $\tilde{x} \in \mathbb{R}^N$ with $\tilde{x}_j = x_{j-1} \ \forall j \in \{2, \ldots, N\}$ and $\tilde{x}_1 = 0$. Then

$$x^T H x = -B \left\{ \sum_{i=1}^N w_i \tilde{x}_i^2 - \left(\sum_{i=2}^N w_i \tilde{x}_i \right)^2 \right\} = -B \mathbb{V}\chi < 0,$$

where χ is a random variable that takes values $\{\tilde{x}_i\}_{i=1}^N$ with probabilities $\{w_i\}_{i=1}^N$. The variance $\mathbb{V}\chi$ is strictly positive because $\tilde{x}_1 = 0$ and $\exists j \in \{2, \ldots, N\}$ such that $\tilde{x}_j \neq 0$.

Hence, we have established that zero is the only critical point of $I(t_{12}, \ldots, t_{1N})$ and

the function reaches maximum at it. By (22), $t_{12} = \cdots = t_{1N} = 0$ if and only if $\Delta_1 = \cdots = \Delta_N$, so in the absence of any heterogeneity the threshold θ^* is at its maximum. Any heterogeneity that is sufficiently weak—so that banks share the same threshold in the limit of infinitely precise signals—therefore reduces fragility of all banks.

□

We now prove Proposition 5.

Proof. Define Λ as

$$\Lambda \equiv \frac{1}{\int_0^1 \frac{1}{1-\bar{m}x}dx} \int_0^1 \left(\lambda(\bar{m}) - \lambda(0)\right) \frac{\bar{m}x}{1-\bar{m}x}dx$$

and assume that $|\Delta_i - \Delta_k| > \Lambda \ \forall i \neq k \in \{1, \ldots, N\}$.

Rewrite (14) as

$$\theta_i^* - \theta_k^* = -\Delta_i + \Delta_k + \frac{1}{\int_0^1 \frac{1}{1-\bar{m}x}dx} \times$$
$$\int_0^1 \left[\lambda\left(\bar{m}\sum_{j=1}^N w_j \Phi\left(t_{ij} + \Phi^{-1}(x)\right)\right) - \lambda\left(\bar{m}\sum_{j=1}^N w_j \Phi\left(t_{kj} + \Phi^{-1}(x)\right)\right)\right] \frac{\bar{m}x}{1-\bar{m}x}dx.$$

Because $|\Delta_i - \Delta_k| > \Lambda$, the right hand side of this expression cannot be zero. Moreover, if $\Delta_i > \Delta_k$ then $\theta_i^* < \theta_k^*$. Therefore, (14) simplifies to

$$\theta_i^* + \Delta_i = \frac{1}{\int_0^1 \frac{1}{1-\bar{m}x}dx} \left(1 + \int_0^1 \lambda\left(\bar{m}w_i x + \bar{m}\sum_{j>i} w_j\right) \frac{\bar{m}x}{1-\bar{m}x}dx\right).$$

Clearly, in this region θ_i^* only depends on Δ_i. Moreover, θ_i^* is a decreasing function of Δ_i. Any change in heterogeneity under the zero-mean constraint $\sum_{i=1}^N w_i \Delta_i = 0$ necessarily implies that there is a bank type i for which Δ_i goes down and, hence, θ_i^* goes up.

□

B.2 Noisy information about bank-specific fundamentals

We prove Proposition 6 in this section.

Proof. Define $t = \lim_{\sigma \to 0} \frac{\theta_B^* - \theta_G^*}{\sigma}$. The indifference conditions for investors receiving signals G and B are respectively

$$\theta_G^* + p_G\Delta - (1-p_G)\frac{w}{1-w}\Delta = \frac{1}{\int_0^1 \frac{1}{1-\bar{m}x}dx}\left(1 + \int_0^1 \frac{\lambda(m_G(x,t))\bar{m}x}{1-\bar{m}x}dx\right),$$
$$\theta_B^* + p_B\Delta - (1-p_B)\frac{w}{1-w}\Delta = \frac{1}{\int_0^1 \frac{1}{1-\bar{m}x}dx}\left(1 + \int_0^1 \frac{\lambda(m_B(x,t))\bar{m}x}{1-\bar{m}x}dx\right),$$

where $m_G(x,t) = (w\alpha + (1-w)(1-\alpha))\bar{m}x + (w(1-\alpha) + (1-w)\alpha)\bar{m}\Phi(t + \Phi^{-1}(x))$ and $m_B(x,t) = (w\alpha + (1-w)(1-\alpha))\bar{m}\Phi(-t + \Phi^{-1}(x)) + (w(1-\alpha) + (1-w)\alpha)\bar{m}x$.

Define *effective* weight and heterogeneity as

$$w^{eff}(\alpha) = w\alpha + (1-w)(1-\alpha),$$
$$\Delta^{eff}(\alpha) = p_G\Delta - (1-p_G)\frac{w}{1-w}\Delta = \frac{2\alpha - 1}{w\alpha + (1-w)(1-\alpha)}w\Delta.$$

It straightforward to see that the model described in Section 4.2 boils down to our baseline setting with redefined type weights and the measure of heterogeneity. Therefore, the results of Section 3.1 generalize to the case of noisy idiosyncratic signals.

□

B.3 Finite noise precision

In this section, we briefly describe how the run thresholds depend on heterogeneity when banks' investors signals are not infinitely precise. For concreteness, we assume that the aggregate fundamental θ is drawn from a uniform distribution. Under this assumption, investors of strong and weak banks follow threshold strategies and withdraw prematurely if and only if their signals are below θ_s^* and θ_w^*.[20] The run thresholds solve

$$\theta_s^* = -\Delta + \frac{1}{\int_0^1 \frac{dx}{1-\bar{m}x}}\left(1 + \int_0^1 \frac{\lambda\left(w\bar{m}x + (1-w)\bar{m}\Phi\left(\frac{\theta_w^* - \theta_s^*}{\sigma} + \Phi^{-1}(x)\right)\right)\bar{m}x}{1-\bar{m}x}dx\right),$$

[20] Under uniform distribution of θ, the proof that the threshold equilibrium is a unique equilibrium is very similar to the one provided in Appendix D and is therefore omitted. For a general prior, it is only possible to show that the threshold equilibrium is unique when signals are sufficiently (but not necessarily infinitely) precise (see Morris and Shin, 2001).

$$\theta_w^* = \frac{w}{1-w}\Delta + \frac{1}{\int_0^1 \frac{dx}{1-\bar{m}x}}\left(1 + \int_0^1 \frac{\lambda\left((1-w)\bar{m}x + w\bar{m}\Phi\left(\frac{\theta_s^*-\theta_w^*}{\sigma} + \Phi^{-1}(x)\right)\right)\bar{m}x}{1-\bar{m}x}dx\right).$$

When σ is finite, θ_s^* and θ_w^* do not need to be infinitely close to each other for strong and weak bank investors to have nontrivial strategic interaction. As a result, even for $\Delta < \Delta_{min}$, $\theta_s^* \neq \theta_w^*$. Although in this case analytical characterization of $\theta_s^*(\Delta)$ and $\theta_w^*(\Delta)$ becomes cumbersome, we verify via numerical examples that the main results of Proposition 1 hold even if signals have finite precision.

In particular, we assume that $f(x) = \log(1+x)$, $w = 0.5$, $\bar{m} = 0.5$, and signals are normally distributed. For these parameter values, $\Delta_{min} = 0.075$ and $\Delta_{max} = 0.103$. Figure 4 shows $\theta_s^*(\Delta)$ and $\theta_w^*(\Delta)$ for the case of infinitely precise signals ($\sigma \to 0$, panel (a)) and finitely precise signals ($\sigma = 0.01$, panel (b)). Panel (b) shows that even when signals have finite precision, both thresholds tend to decline when heterogeneity Δ goes up. In comparison with the case when $\sigma \to 0$, the region of nontrivial strategic interaction widens (when $\sigma \to 0$, this region is $0 \leq \Delta < \Delta_{min}$). As a result, $\theta_w^*(\Delta)$ keeps decreasing even when Δ crosses Δ_{min}.

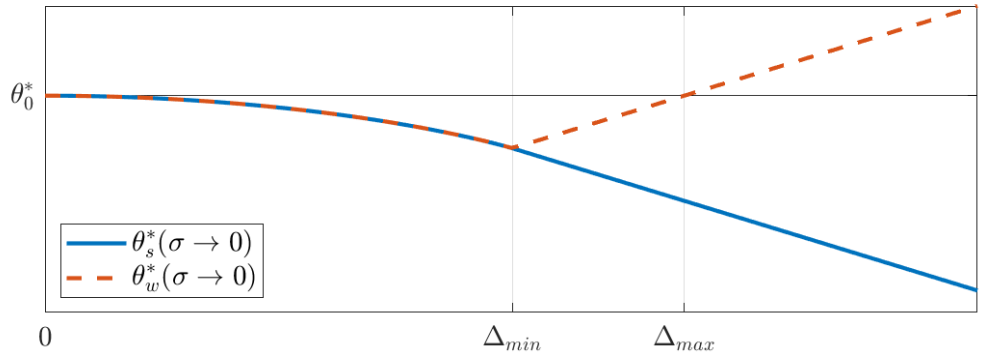
C Role of aggregate uncertainties

In the baseline model, we assume that the aggregate fundamental θ is not perfectly observable to investors. As a result, they are not certain about the total mass of early withdrawers m and the liquidation costs. Arguably, aggregate uncertainties lie at the heart of the 2007-2008 financial crisis. Financial institutions and investors limited their market participation and reduced liquidity provision because of uncertainties about systemic failure in the financial system.

From a theoretical standpoint, imperfect signals on θ help individual investors both within and across banks to coordinate, which gives rise to a unique equilibrium and allows us to conduct a sensible analysis of financial stability. If, on the contrary, investors perfectly observe θ, the cross-bank complementarity might give rise to multiple equilibria. As we argue below, a meaningful analysis of financial stability then become unachievable.

Consider an alternative information structure with no aggregate uncertainty. Specifically, at $t = 1$, the aggregate fundamental θ is a common knowledge. In addition, we introduce

(a) Infinitely precise signals



(b) Finitely precise signals

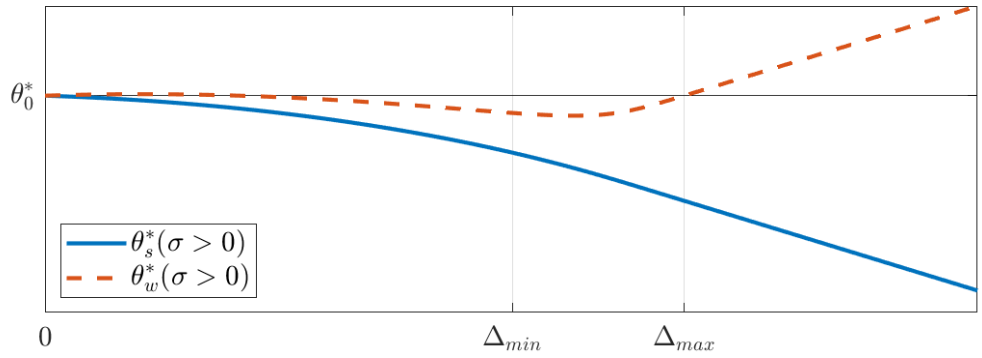


Figure 4: Run thresholds as function of heterogeneity Δ: infinitely precise signals (panel a) and finitely precise signals (panel b)

noises to bank-specific information to show that idiosyncratic uncertainties alone are insufficient to resolve equilibrium multiplicity. In particular, we assume that bank i's bank-specific fundamental η_i is characterized by its position x_i, uniformly distributed on $[0, 1]$, as follows:

$$\eta_i = \begin{cases} \Delta & \text{if } x_i > 1 - w, \\ -\frac{w}{1-w}\Delta & \text{otherwise.} \end{cases}$$

This setting is consistent with our baseline model setup that $\eta_i = \Delta$ with probability w and $\eta_i = -\frac{w}{1-w}\Delta$ with probability $1 - w$. At $t = 1$, investor j in bank i receives a noisy signal about its position x_i,[21]

$$s_{ij}^{\eta} = x_i + \sigma\epsilon_{ij},$$

[21] Under this information structure, there is no uncertainty about the distribution of bank-specific fundamentals. Investors know that bank-specific fundamental can only take two values, Δ and $-\frac{w}{1-w}\Delta$, but they do not know whether their bank is strong or weak.

where ϵ_{ij} has zero mean and follows an independent and identical distribution with a strictly increasing on its support $[\underline{\epsilon}, \bar{\epsilon}]$ cumulative distribution function $\Phi(\cdot)$. Again, we focus on the limit when information friction vanishes, i.e., $\sigma \to 0$. Everything else in the setup stays the same as in our baseline model presented in Section 2.

Under this alternative setup, we focus on threshold equilibria in which all investors follow threshold strategies $x^*(\theta)$. For a given aggregate fundamental θ, a marginal investor who receives signal $s_{ij}^{\eta} = x^*(\theta)$ is indifferent between staying and withdrawing early:

$$\int_0^1 \frac{\theta + \mathbb{E}(\eta_i | x^*) - \lambda(m(x^*))\bar{m}x}{1 - \bar{m}x} dx - 1 = 0, \tag{23}$$

where

$$m(x^*) = \bar{m} \int_0^1 \Phi\left(\frac{x^* - x_i}{\sigma}\right) dx_i.$$

As x^* increases, two opposing forces are at play. On the one hand, the expected bank-specific fundamental of the marginal investor $\mathbb{E}(\eta_i | x^*)$ increases, raising the expected payoff from staying. Thus, if the total amount of runs in the system m was fixed then Equation (23) would have a unique solution for x^*. With flexible m, on the other hand, an increase in x^* leads to more runs in the whole economy, which reduces the expected return from staying for the marginal investor. As a result, the left-hand side of Equation (23) is in principle nonmonotone, which might give rise to multiple solutions for the threshold x^*.

Perfect information about the aggregate fundamental leaves investors with no uncertainty about the cross-bank fire-sale externalities, captured by $\lambda(m(x^*))$. Common knowledge about $m(x^*)$ introduces a feedback loop resulting in multiple equilibria. If investors follow a high (low) run threshold x^*, the mass of withdrawers $m(x^*)$ is high (low), aggravating (alleviating) the fire-sale externalities. Strong (weak) fire-sale externalities, in turn, give more (less) incentive for investors to run on their banks, justifying a high (low) run threshold x^*. Therefore, there can exist multiple run thresholds $x^*(\theta)$ for a given θ.

In particular, when the aggregate fundamental is sufficiently good, i.e., $\theta > \theta_L$, there exists a "low-run threshold" $x_L^*(\theta) < 1 - w$. We say the threshold is low because runs only occur in weak banks with $x_i < x_L^*(\theta)$. When $\theta < \theta_H$, there exists a "high-run threshold" $x_H^*(\theta) > 1 - w$. In this case, runs occur in all weak banks as well as strong

banks with $x_i < x_H^*(\theta)$. If $\theta_L < \theta_H$, both equilibria coexist when $\theta \in (\theta_L, \theta_H)$.[22]

Proposition 8. *$\theta_L(\Delta)$ is an increasing function and $\theta_H(\Delta)$ is a decreasing function.*

Proof. First, we show that a marginal investor believes that the mass of withdrawers in its bank is uniformly distributed. Given the equilibrium run threshold $x^*(\theta)$, the mass of withdrawers in bank i is

$$m_i(x^*) = \bar{m}\Phi\left(\frac{x^* - x_i}{\sigma}\right).$$

Therefore, for a marginal investor in bank i, the cumulative distribution of the mass of withdrawers in her bank can be expressed as follows,

$$\mathbb{P}(m_i(x^*) \leq m | x^*) = \mathbb{P}\left(x_i \geq x^* - \sigma\Phi^{-1}\left(\frac{m}{\bar{m}}\right) \middle| x^*\right) = \Phi\left(\frac{x^* - x^* + \sigma\Phi^{-1}\left(\frac{m}{\bar{m}}\right)}{\sigma}\right) = \frac{m}{\bar{m}}.$$

Hence, for a marginal investor, the mass of withdrawers in its bank is uniformly distributed on $[0, \bar{m}]$.

Next, we write out the expected payoff from staying for a marginal investor given the aggregate fundamental θ as follows,

$$V(x^*, \theta) = \int_0^1 \frac{\theta + \mathbb{E}(\eta_i | x^*) - \lambda(m(x^*))\bar{m}x}{1 - \bar{m}x} dx,$$

where

$$\mathbb{E}(\eta_i | x^*) = \Delta\frac{1}{1-w}\Phi\left(\frac{x^* - (1-w)}{\sigma}\right) - \frac{w}{1-w}\Delta,$$

and

$$m(x^*) = \bar{m}\int_0^1 \Phi\left(\frac{x^* - x}{\sigma}\right) dx.$$

We consider a limiting case of inifinitely precise signals, $\sigma \to 0$. Below, we characterize all possible $x^*(\theta)$.

Case 1: $\lim_{\sigma\to 0}\frac{x^*-(1-w)}{\sigma} = d \in (-\infty, \infty)$.

In this case, only investors of weak banks run. The indifference condition for a threshold

[22] There exists a third possible equilibrium where weak bank investors run and strong bank investors do not, so that $x^* = 1 - w$. See the proof of Proposition 8 below.

investor is

$$\int_0^1 \frac{\theta + \Delta\frac{1}{1-w}\Phi(d) - \frac{w}{1-w}\Delta - \lambda(\bar{m}(1-w))\bar{m}x}{1-\bar{m}x}dx = 1.$$

An equilibrium outcome featuring investors of only weak banks running can emerge when $\theta \in (\theta_1; \theta_2)$, where $\theta_1 = -\Delta + v_1$, $\theta_2 = \frac{w}{1-w}\Delta + v_1$, and $v_1 = \frac{1}{\int_0^1 \frac{dx}{1-\bar{m}x}}\left[1 + \int_0^1 \frac{\lambda(\bar{m}(1-w))\bar{m}x}{1-\bar{m}x}dx\right]$.

Case 2: $\lim_{\sigma\to 0}\frac{x^*-(1-w)}{\sigma} = \infty$.

The indifference condition for a threshold investor is

$$\int_0^1 \frac{\theta + \Delta - \lambda(\bar{m}x^*)\bar{m}x}{1-\bar{m}x}dx = 1.$$

Define $\theta_H = -\Delta + \frac{1}{\int_0^1 \frac{dx}{1-\bar{m}x}}\left[1 + \int_0^1 \frac{\lambda(\bar{m})\bar{m}x}{1-\bar{m}x}dx\right]$. When $\theta < \theta_H$, a "high-run threshold" exists. For these values of θ an equilibrium outcome with all weak bank investors and at least some strong bank investors running is possible. Clearly, $\theta_H(\Delta)$ is a decreasing function.

Case 3: $\lim_{\sigma\to 0}\frac{x^*-(1-w)}{\sigma} = -\infty$.

The indifference condition for a threshold investor is

$$\int_0^1 \frac{\theta - \Delta\frac{w}{1-w} - \lambda(\bar{m}x^*)\bar{m}x}{1-\bar{m}x}dx = 1.$$

Define $\theta_L = \Delta\frac{w}{1-w} + \frac{1}{\int_0^1 \frac{dx}{1-\bar{m}x}}\left[1 + \int_0^1 \frac{\lambda(0)\bar{m}x}{1-\bar{m}x}dx\right]$. When $\theta > \theta_L$, a "low-run threshold" exists. For these values of θ an equilibrium outcome with at least some weak bank investors not running is possible. Clearly, $\theta_L(\Delta)$ is an increasing function.

□

With stronger heterogeneity, there are fewer states of the world where investors can in principle follow strategies with a "low-run threshold" (θ_L goes up). This happens because weak banks become weaker, and higher values of the aggregate fundamental θ are required to prevent some weak bank investors from running. Similarly, a "high-run threshold" moves upwards because strong banks are becoming stronger. Increase in Δ, therefore, yields ambiguous implications about financial stability.

D Global games proofs

In this section, we prove that our baseline model features a unique equilibrium. In the unique equilibrium, investors of all banks follow threshold strategies, where the value of thresholds are given by (5) and (6). Our proofs are based on Morris and Shin (2001) but are modified to address bank heterogeneity and cross-bank interaction.

Consider the setting of Section 2. For investors of bank i, the net benefit of not withdrawing funds early is $g(\theta + \eta_i, m_i, m)$. Here m_i is the total mass of runners on bank i, and $m = \int w_i m_i di$ is the total mass of runners in the whole economy. Idiosyncratic productivity η_i takes value of Δ with probability $w_s = w$ and $-\Delta\frac{w}{1-w}$ with probability $w_w = 1 - w$. To reduce notational clutter, we assume that all banks are able to run at $t = 1$, $\bar{m} = 1$.

Investor j of bank i receives a signal about the aggregate fundamental θ,

$$s_{ij} = \theta + \sigma\epsilon_{ij},$$

where ϵ_{ij} are identically and independently across investors distributed noise with cumulative distribution function $\Phi(\cdot)$ which is differentiable, has a finite first moment and is strictly increasing on its support $[\underline{\epsilon}, \bar{\epsilon}]$. Fundamental θ follows a prior distribution with cumulative distribution function $F_\theta(\cdot)$ with support $[\underline{\theta}, \bar{\theta}]$.

We make the following set of standard assumptions, all of which are satisfied in our baseline model.

Assumption 1. *$g(\theta, m_i, m)$ is a continuous function that increases in θ, decreases in m_i and m.*[23]

Assumption 2. *Strict Laplacian State Monotonicity: There exists a unique pair $(\theta_s^*; \theta_w^*)$ solving*

$$\int_0^1 g\left(\theta_i^* + \Delta_i, m, \sum_{j\in\{s,w\}} w_j \Phi\left(t_{ij} + \Phi^{-1}(m)\right)\right) dm = 0,$$

where $t_{ij} \equiv \lim_{\sigma\to 0} \frac{\theta_j^* - \theta_i^*}{\sigma}$.

[23] In principle, the equilibrium uniqueness result does not require strict monotonicity and continuity assumptions on g. However, we make these assumptions because throughout the paper we work with smooth payoff functions.

Assumption 3. *There exists $\theta^{LDR} > \underline{\theta}$, $\theta^{UDR} < \bar{\theta}$, and $\delta > 0$ such that $g(\theta^{LDR}+\eta, 0, 0) \leq -\delta$ and $g(\theta^{UDR}+\eta, 1, 1) \geq \delta$ for all viable realizations of idiosyncratic shock η.*

In this appendix, we work with the uniform prior on θ. Moreover, we assume that the net benefit function g depends on investors' signals but not directly on θ. These assumptions are innocuous when signals are infinitely precise $\sigma \to 0$ (Proposition 2.2 of Morris and Shin, 2001).

Proposition 9. *Let θ_i^*, $i \in \{s, w\}$ be defined as in Assumption 2. The unique strategies surviving iterated deletion of dominated strategies is such that investors of bank of type i withdraw early if their signals are below θ_i^* and do not withdraw early otherwise.*

Proof. Define $\pi^s(s, k_s, k_w)$ as the expected payoff from not running on her bank for a strong bank's investor that observes signal s and knows that investors of strong/weak banks run if they observe signals below k_s/k_w. Define $\pi^w(s, k_s, k_w)$ in an analogous way but for a weak bank's investor.

$$\pi^i(s, k_s, k_w) = \tag{24}$$
$$\int_{s-\sigma\bar{\epsilon}}^{s-\sigma\underline{\epsilon}} g\left(s + \Delta_i, \Phi\left(\frac{k_i - \theta}{\sigma}\right), \sum_{j\in\{w,s\}} w_j \Phi\left(\frac{k_j - \theta}{\sigma}\right)\right) \frac{1}{\sigma}\phi\left(\frac{s-\theta}{\sigma}\right) d\theta,$$

where $i \in \{w, s\}$.

$\pi^i(s, k_s, k_w)$ is continuous, increasing in s, decreasing in k_s and k_w. Below, we describe the process of iterated deletion of dominated strategies. Fix $\underline{k}_s^0 = \underline{k}_w^0 \in (\underline{\theta}, \theta^{LDR})$ and $\bar{k}_s^0 = \bar{k}_w^0 \in (\theta^{UDR}, \bar{\theta})$. Define

$$\underline{k}_i^{n+1} = \max\{s : \pi^i(s, \underline{k}_s^n, \underline{k}_w^n) = 0\},$$
$$\bar{k}_i^{n+1} = \min\{s : \pi^i(s, \bar{k}_s^n, \bar{k}_w^n) = 0\}.$$

Then for investor of bank i a strategy survives n rounds of iterated deletion of dominated strategies if and only if she withdraws early when $s < \underline{k}_i^n$ and does not withdraw early when $s > \bar{k}_i^n$.

Notice that $\underline{k}_i^n$ and $\bar{k}_i^n$ are increasing and decreasing sequences, respectively, due to monotonicity properties of $\pi^i(s, k_s, k_w)$. Hence, $\underline{k}_i^n \to \underline{k}_i$ and $\bar{k}_i^n \to \bar{k}_i$. By continuity of $\pi^i(s, k_s, k_w)$, it must be that $\pi^i(\bar{k}_i, \bar{k}_s, \bar{k}_w) = \pi^i(\underline{k}_i, \underline{k}_s, \underline{k}_w) = 0$.

Change variables in (24), $m = \Phi\left(\frac{k_i - \theta}{\sigma}\right)$, to obtain

$$\pi^s(\theta_s^*, \theta_s^*, \theta_w^*) = \int_0^1 g\left(\theta_s^* + \Delta, m, wm + (1-w)\Phi\left(\frac{\theta_w^* - \theta_s^*}{\sigma} + \Phi^{-1}(m)\right)\right) dm,$$

$$\pi^w(\theta_w^*, \theta_s^*, \theta_w^*) = \int_0^1 g\left(\theta_w^* - \Delta\frac{w}{1-w}, m, (1-w)m + w\Phi\left(\frac{\theta_s^* - \theta_w^*}{\sigma} + \Phi^{-1}(m)\right)\right) dm.$$

These are Equations (5) and (6) from the main text (when $\sigma \to 0$). They have a unique solution by the Laplacian monotonicity property. □

It is worth mentioning here that this proof can be straightforwardly extended to the case with $N > 2$ types banks. Finally, we omit the proof of the uniqueness result for non-uniform prior on θ and fundamental-dependent payoff functions. This proof is standard and follows Morris and Shin (2001) directly.