Increasing Returns to Scale and Financial Fragility*

Jiahong Gao Robert R. Reed

August 15, 2022

Abstract

How do liquidity creation and financial fragility depend on increasing returns to scale? We study this question in a version of the Diamond and Dybvig (1983) model with limited commitment. When depositors' preferences are of the constant-relative-risk-aversion variety and scale economies are absent, the degree of financial fragility is independent of the level of investment and aggregate welfare is proportional to the size of the bank's deposit base. Yet, once the minimum scale effect is present, these conventional results may be overturned. In particular, the bank facing an intermediate scale threshold is able to deliver the highest expected utility to depositors but is most prone to panics, especially if runs are highly unlikely. Such findings indicate that failing to account for scale economies in banking activity leaves a significant void in policy debates regarding the stability and efficiency of the financial system.

JEL Classification: G21, G51
Keywords: Bank Runs, Increasing Returns to Scale, Limited Commitment

*We thank conference and seminar participants at the 2021 European (Winter), China, and Asian Meetings of the Econometric Society and Zhongnan University of Economics and Law for their helpful comments. Jiahong Gao: Wenlan School of Business, Zhongnan University of Economics and Law. E-mail address: jiahong.gao@zuel.edu.cn; Tel.: +86 13063038299. Please address correspondence to Robert R. Reed: Department of Economics, Finance and Legal Studies, University of Alabama. E-mail address: rreed@cba.ua.edu; Tel.: +1 (205) 348-8667.

1 Introduction

In recent years, there has been an abundance of empirical research that shows that economies of scale play an important role in banking activity. For instance, using data on all U.S. banks during the period 1984-2006, Wheelock and Wilson (2012) find evidence that banks over a large range of sizes faced increasing returns to scale. In addition, Wheelock and Wilson (2018) present new estimates of returns to scale for U.S. banks for both 2006 and 2015. They find that the largest banks continued to operate under increasing returns to scale in 2015 and that scale economies were stronger for many of the largest institutions between 2006 and 2015. Moreover, the percentage of banks that experienced increasing returns to scale increased from 35% in 2006 to 43% in 2015. In other work, for 103 European listed banks over 2000-2011, Beccali et al. (2015) show that economies of scale were widespread across different size classes of banks and were especially large for the biggest banks. Given the existence and vital role of increasing returns to scale in the banking sector, it is important to develop a framework to investigate how scale economies affect banking activity and financial stability. Notably, while a burgeoning literature in the tradition of the Diamond and Dybvig (1983) model of bank runs examines the consequences of financial instability, less attention has been paid to increasing returns to scale. Yet, by influencing the degree of maturity transformation, economies of scale arguably impact intermediaries' exposure to liquidity risk and thus fragility.

The objective of this paper is to examine the role of increasing returns to scale for liquidity transformation, risk-sharing, and financial fragility. To this end, we build on a version of Diamond and Dybvig (1983) with limited commitment as in Ennis and Keister (2009) where the bank subject to instability is unable to commit to a course of actions but issues payments which adjust to available information at each point in time. In this setting, we incorporate scale economies as in Andolfatto and Nosal (2018) where the higher productive asset return is obtained if the quantity of investment held to maturity is above a minimum level (also referred to as the scale/size threshold). Given that increasing returns reflect critical characteristics of investment opportunities facing banks, our approach can be viewed as a generalization of the standard specification used in the existing literature.[1] More importantly, introducing economies of scale also allows us to study the endogenous level of deposits in a pre-deposit game, which is relevant to financial fragility and

[1]To be clear, the standard Diamond-Dybvig model assumes that the asset return is linear in the amount of bank investment. Yet, the analysis in our structure is based upon a form of economies of scale in which the asset return is higher over the larger investment size. As we will discuss later, the low return can be thought of as that on debt securities whereas the high return on large-scale investment portfolios. However, we also acknowledge that empirical studies regarding scale economies may suggest other forms of increasing returns. The return on assets, for example, may be a smooth, strictly convex function of the amount invested or it may have combined features where over a minimum scale, the rate of return on assets is an increasing and strictly convex function of the size of bank investment. While allowing for these possibilities may produce more patterns of banking behavior and thus financial fragility, it is first useful to look at the impact of a scale threshold with a relatively high degree of tractability.

welfare.

Prior to the exposition of our results, we first discuss the adopted modeling strategy and highlight some important features of our setup. As a starting point, in order to show how the effects of increasing returns to scale on financial stability depend on the potential for consumer sentiment to deteriorate, runs are modeled as positive-probability events. In this regard, our framework seeks to study how the possibility of a crisis along with the potential to reap the benefits of scale economies affects banking behavior – in particular, how the bank structures its liabilities which impact the incentives for self-fulfilling runs to emerge. It is straightforward that one can also study this issue in a much simplified environment by confining attention to a specific case where panics occur unexpectedly as in Diamond and Dybvig (1983) and Ennis and Keister (2009). Yet, such an approach does not account for bank behavior in a relatively more crisis-prone economy. Further, due to limited commitment and the predictable contingency that banks will be able to infer panics, if any, from the flow of withdrawals, intermediaries resort to rescheduling payoffs on deposits upon the identification of a run. That is, other types of ex-post interventions such as a fixed payoff schedule without a suspension clause or a simple deposit freeze are suboptimal or may be nested in payment rescheduling. Consequently, these analyses would only offer limited insights into the impact of scale economies on liquidity creation and financial fragility.

In this setting, we conduct analysis by focusing on equilibrium outcomes of a competitive bank whose objective is to maximize depositors' expected utility. Yet, as a benchmark for understanding the role of increasing returns, we first study how maturity transformation and depositors' incentives to panic are affected if the bank only has access to a constant-returns-to-scale technology. Notably, absent scale economies, consumers would certainly pool their full endowments in the intermediary as long as the bank is able to provide a higher level of expected utility than would be possible to achieve at the individual level. Under this benchmark setting, there exists a linear relationship between the amount of bank investment and welfare. Further, the degree of instability is independent of the size of the bank's deposit base. Intuitively, payments on deposits issued by the bank are only scaled in proportion to the size of the intermediary's attracted deposits. As a result, changes in the amount of bank investment alter the sum of depositors' expected utilities proportionally but do not affect a bank's susceptibility to a crisis. In this vein, all intermediaries would be predicted to behave in the same manner and fragility is not affected by the bank's investment scale. In contrast, once scale economies are taken into consideration, consumers may make partial deposits in a bank to achieve increasing returns in normal times only as risk-sharing cannot be improved if more endowments are pooled in the intermediation. More importantly, a non-monotonicity of financial fragility and aggregate welfare with respect to the minimum scale may arise. In particular, the bank facing an intermediate size threshold is capable of delivering the highest expected utility to depositors but is most prone to crises, especially when runs are highly unlikely.

We proceed to provide more details of our results. To begin, if the minimum scale requirement is sufficiently low, the bank best responds by anticipating that it will experience economies of scale regardless of whether a run occurs. In this way, it provides the highest level of aggregate welfare. If the size threshold is higher, however, the intermediary will issue liabilities conservatively so that it is still able to take advantage of increasing returns even in the event of a crisis. For this reason, a bank facing a somewhat higher size threshold is more stable. Nevertheless, such an intermediary with a larger minimum scale delivers lower welfare to depositors, implying that there exhibits a trade-off between the benefits of risk-sharing and financial fragility across scale cutoffs.

By comparison, over intermediate size thresholds, it may be desirable for the bank not to qualify for a high asset return in the presence of a panic. Given such incentives, the intermediary focuses more on promoting risk-sharing by offering high deposit returns to consumers who withdraw before the bank infers whether a panic has occurred. Should a poor realization of the sunspot signal happen, however, the intermediary will sharply reduce payments on deposits due to low asset returns, exacerbating the degree of fragility. Yet, aggregate welfare can be as high as that under the bank facing a sufficiently low scale requirement as long as bank run risk is negligible. Interestingly, along relatively lower size thresholds where the bank pursues the higher asset return in normal times only, consumers deposit part of their endowments with the bank. This is the case because the implementation of the bank's best-response allocation requires a lower investment scale and individuals resort to direct investing for the rest of funds as they have access to one bank only.

Furthermore, if threshold levels are high but scale economies are weak, e.g., the spread between the high and low asset returns is small, the way that the intermediary behaves again revolves around the minimum scale. To be precise, over relatively lower cutoffs of high scale requirements, the bank fulfills redemption requests prudently by lowering payments prior to identifying the realized aggregate state in order to qualify for increasing returns in normal times. Yet, if the intermediary faces a comparatively high scale threshold, the welfare loss from offering low early returns on deposits to achieve the minimum scale in normal times is significant. In particular, it dominates the welfare gain from providing slightly more consumption ex-post in a weak scale economy. In this vein, it would be desirable for the bank to best respond by not taking advantage of economies of scale, implying that there exists an endogenous threshold minimum size below which the bank issues short-term obligations cautiously. In general, unveiling these aspects of vulnerability and welfare indicates that failing to account for scale economies in banking activity leaves a significant void in policy debates regarding the stability and efficiency of the financial system.

The remainder of the paper is organized as follows. In the next section, we turn to the literature review. Section 3 presents the baseline model and discusses its key assumptions and the definition of financial fragility. In Section 4, we analyze how increasing returns to scale shape the bank's best responses to an incipient crisis and study the implications of banking activity for financial

instability and welfare. Section 5 concludes.

2 Related literature

The model presented in this paper relates to a strand of literature that studies the consequences of fragility in the Diamond-Dybvig tradition. In particular, Ennis and Keister (2009) highlight the role of limited commitment for inducing a bank run equilibrium, which would otherwise be prevented by an off-equilibrium suspension of convertibility. Building on this type of model that recovers multiplicity of equilibria, we study the impact of scale economies, which have been shown to be an important component of banking activity, on maturity transformation and financial fragility.

Our framework is also closely related to the work by Andolfatto and Nosal (2018) who build on the Green and Lin (2000, 2003) model with aggregate uncertainty about the total fundamental demand for early consumption. Specifically, the authors study the design of bank contracts and the possibility of run equilibria where bank-funded investments are subject to scale economies. They find that in such a setting, runs can occur without a sequential service constraint. In this way, the authors argue that their analysis rationalizes the fragility of a wholesale shadow bank absent sequential service. Moreover, the expected cost of eliminating bank run equilibria is decreasing in bank size, implying that larger banks are more stable. By comparison, we maintain sequential service and other important features such as limited commitment that generate bank run equilibria as in Ennis and Keister (2009) and ask how increasing returns affect fragility and aggregate welfare. Notably, we unveil the novel possibility that there exhibits a non-linear relationship across the relative bank's investment scale, instability, and social welfare.

Other papers analyze how asset returns affect financial stability directly in the absence of scale economies. Bertolai et al. (2014), for instance, examine a setting with a finite population similar to Green and Lin (2003) and Peck and Shell (2003). They show that a bank run is off the equilibrium path of play if asset returns are high enough but is on if asset returns are sufficiently low. Following Ennis and Keister (2010), Li (2017) studies how a change in the term structure of interest rates affects the vulnerability of banks to a run. By varying the value of asset returns, the author shows that the relationship between interest rates and banking instability is non-monotone. Again, while these are valuable contributions, the presence of scale economies is overlooked.

Last but not least, our model also features an endogenous level of deposits, which is absent in the previously highlighted literature that exogenously assumes that all of consumers' endowments are pooled in the bank. In a similar spirit, Ennis and Keister (2003), Gu et al. (2020), and Peck and Setayesh (2022), among others, study the endogenous level of deposits but the modeling strategies are different from the present paper. That is, the decisions of how much to deposit are not driven by economies of scale.

3 The model

We augment the Ennis and Keister (2009) version of the Diamond and Dybvig (1983) model to incorporate scale economies as in Andolfatto and Nosal (2018). Moreover, in contrast to previous work, we also endogenize consumers' deposit decisions in a pre-deposit game. We begin by describing the physical environment and defining financial fragility in this setting.

3.1 The environment

There are three time periods indexed by $t = 0, 1, 2$. The economy is populated by a $[0, 1]$ continuum of ex-ante identical consumers indexed by i. Each consumer is equally endowed with one unit of the all-purpose good at the start of period 0 and has preference of the form:

$$u\left(c_1, c_2; \omega_i\right) = \frac{\left(c_1 + \omega_i c_2\right)^{1-\gamma}}{1-\gamma},$$

where c_t denotes the consumption level in period $t = 1, 2$. For consumer i, ω_i represents her type and is a privately observed binomial random variable with support $\Omega \equiv \{0, 1\}$. With probability π, an individual i is impatient, i.e., $\omega_i = 0$, and only values consumption in period 1; with probability $1 - \pi$, an individual i is patient, i.e., $\omega_i = 1$, and values consumption equally in both periods. According to the law of large numbers, the fraction of impatient or patient consumers in the population is also π or $1-\pi$. Following Diamond and Dybvig (1983), the coefficient of relative risk-aversion γ is greater than unity.

As in Andolfatto and Nosal (2018), there is a single increasing-returns-to-scale investment technology that transforms k units of goods in period 0 into $R_A k$ units of period-2 output:

$$R_A k = \begin{cases} R_L k & \text{if } k \in [0, \kappa) \\ R_H k & \text{if } k \in [\kappa, 1) \end{cases},$$

where $1 < R_L < R_H$ and $\kappa \in (0, 1)$.[2] That is, the higher productive asset return R_H is realized if and only if the level of investment held to maturity exceeds a certain scale requirement κ. In practice, R_L can be thought of as the rate of return on debt securities whereas R_H is the rate of return on large-scale investment portfolios. Yet, if assets are liquidated early in period 1, only the face value of investment is received. In addition, for a given lower productive asset return R_L, the greater the higher productive asset return R_H, the stronger economies of scale. Overall,

[2]One could also generalize the analysis to permit multiple threshold levels and associated rates of return, as discussed in Andolfatto and Nosal (2018). In this setting, we restrict our attention to a single threshold. Adopting this approach simplifies notation and allows us to focus more clearly on how increasing returns affect liquidity creation and financial instability.

such a technology can be viewed as important generalization to the standard specification used in the existing literature, which assumes $\kappa = 0$ so that $R_A(\kappa = 0)k = R_L k = R_H k$ for all $R_L = R_H > 1$ and $k > 0$.[3] It is also worth noting that though the minimum scale effect is present, among cases where the size threshold is higher than the amount of bank investment, i.e., $k \in [0, \kappa)$, the intermediary will not be able to operate under increasing returns.

To insure themselves against idiosyncratic liquidity risk, consumers may form coalitions referred to as banks whose objective is to maximize the expected utility of depositors. As in Ennis and Keister (2009), the bank lacks commitment and offers a flexible demandable deposit scheme which adjusts to available information at each point in time. Further, as is common in the literature for the purpose of simplifying the exposition, we assume that a consumer can only access one bank.[4] Importantly, welfare derived from the banking service should not be lower than the value of consumers' outside options. That is, resources are not pooled if the bank yields lower welfare to depositors than financial autarky. To determine consumers' reservation utility, note that an agent is represented by a point in the unit interval $[0, 1]$ in the sense that the measure of an individual is arbitrarily close to zero. Then, if consumer i chooses not to meet at $t = 0$, it would not be possible for her to meet the positive, minimum scale requirement $\kappa > 0$. As a result, an agent would only be able to obtain the expected asset return R_L. In this manner, let W^A denote the aggregate reservation utility derived from one unit of goods invested under autarky, which is given by:

$$W^A = \pi u(1) + (1 - \pi) u(R_L). \tag{1}$$

3.2 Bank panics

Following Cooper and Ross (1998) and Peck and Shell (2003), among others, we introduce an extraneous sunspot variable as a device for coordinating depositors' actions. In particular, there are two possible states of the world, $s \in S \equiv \{\alpha, \beta\}$, with respective probabilities $\{1 - q, q\}$. The value of q is assigned strictly between 0 and 1 and hence the potential for run behavior occurs with a positive probability. Note that the realization of the sunspot state is revealed to depositors

[3] It is worth noting that the interpretation of the scale economy in Andolfatto and Nosal (2018) is different from ours. Andolfatto and Nosal interpret the scale economy in their setting as costly liquidation where any remaining investment yields a rate of return lower than unity if the level of liquidation exceeds a threshold. In this manner, the authors attempt to rationalize instability of the wholesale financial intermediation absent sequential service. Nevertheless, in a related setting of Gao and Reed (2021) where costly liquidation occurs if long-term assets are recouped prematurely, the degree of instability can be shown to be independent of the size of attracted deposits. As a result, we do not incorporate costly liquidation but assume that increasing returns to scale are a property of bank investment.

[4] Note that given the potential advantage of economies of scale described above, all consumers would tend to pool resources at only one intermediary. Note also that as will be shown later, the optimal structure of liabilities by the bank may not require that all individuals deposit their full endowments. Under these scenarios, agents are assumed to resort to the option of direct investing for the rest of goods as they have access to one bank only. Our main mechanisms and results would still hold if consumers are allowed to pool the remaining resources to form another coalition.

immediately at the beginning of period 1 whereas the bank needs to infer the state from the flow of withdrawals.

Note also that since a fraction π of depositors in a bank are impatient and have an urgent need to consume in period 1, they will choose to withdraw early with certainty. Thus, if only π depositors show up in period 1, no panics occur and, under such circumstances, the bank would resume its operations normally. Nonetheless, the intermediary experiences runs if more than a fraction π of depositors request withdrawals before long-term assets mature. In these situations, the bank will choose the of amount of productive assets to liquidate and reschedules the remaining payments to maximize returns to consumers who have not contacted the bank. That is, the intermediary will implement the efficient continuation allocation of resources.

3.3 Withdrawal strategies and partial runs

As discussed above, at the start of period 1, depositors decide whether to contact the bank in period 1 or to wait until period 2 after observing their preference type and the aggregate state. In particular, for each combination of the consumers' preference type and the sunspot state, the function for a withdrawal strategy for an agent i at the bank is:

$$y_i : \{0,1\} \times \{\alpha, \beta\} \mapsto \{0,1\}.$$

Specifically, $y_i = 0$ and $y_i = 1$ correspond to withdrawing in periods 1 and 2 respectively.

To proceed, without loss of generality, depositors withdraw funds according to their preference types in state α, which is labeled the good aggregate state. Nevertheless, runs may occur in state β, which is called the panic or crisis state. We focus on actions taken by patient depositors, since impatient agents only value consumption in period 1 and will choose to withdraw early with certainty. Then, we consider a specific partial-run strategy profile for depositors at the bank as follows:

$$\begin{aligned} & y_i(\omega_i, \alpha) = \omega_i \qquad \text{for all } i \text{ and} \\ & y_i(\omega_i, \beta) = \left\{ \begin{array}{c} 0 \\ \omega_i \end{array} \right\} \quad \text{for} \left\{ \begin{array}{l} i \in [0, \pi] \\ i \in (\pi, 1] \end{array} \right\}. \end{aligned} \tag{2}$$

In this profile, the patient depositor with $i \leq \pi$ at the bank chooses to withdraw in period 1 in the panic state. As we will show below in our analysis, once the remaining consumption levels are rescheduled after the first fraction π of withdrawals, all remaining patient depositors will have no incentive to run.[5] In this regard, we next state the formal definition of financial fragility.

[5]Note that we confine our attention to a single wave run to simplify notation. One could generalize the model

Definition 1. *If the strategy profile* (2) *is part of an equilibrium, the bank is said to be fragile.*

Note also that this form of fragility is different from that in Diamond and Dybvig (1983) where depositors run en masse and there is no payment rescheduling. However, the interpretation of relative fragility across different minimum scales in the present paper would be the same as that in a model where full bank runs are possible. That is, if the parameter space in which run equilibria exist is larger, then a bank is said to be more fragile.

3.4 Timeline of events

The timing of events is summarized in Figure 1. In period 0, after receiving their endowments, consumers decide whether to pool resources in a bank. If the expected utility of the risk-sharing arrangement offered by the intermediary is not worse than autarky, then agents deposit their full or partial funds.

At the beginning of period 1, depositors choose which period to withdraw after learning their liquidity needs and the realized aggregate state. If individuals decide to withdraw in period 1, they will arrive one at a time in the order given by their index within a bank and must consume immediately upon arrival. Following Wallace (1988, 1990), the bank is subject to this sequential service constraint where it only observes that a depositor has come and thus the payment to each withdrawing individual can only depend on the number of total withdrawals that have been made. As discussed above, the bank with limited commitment offers a flexible payout scheme. That is, once the bank infers the aggregate state after a fraction π of depositors have withdrawn, payments to the remaining depositors will be rescheduled based on the updated information.

In period 2, depositors who have not withdrawn contact the bank and consume.

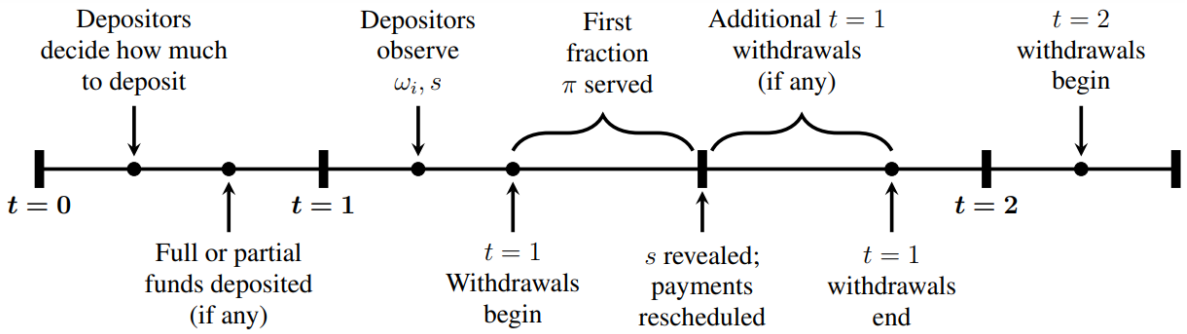

Figure 1: Timeline of events

by introducing multiple waves of runs in an environment of a richer sunspot variable as in Ennis and Keister (2010). Employing such an approach, however, does not alter the main mechanisms and results in the model qualitatively.

4 Equilibrium, banking fragility, and welfare

We proceed to examine how the bank performs liquidity transformation and responds to idiosyncratic short-term withdrawals, which in turn affects an intermediary's exposure to liquidity risk and hence financial stability. Note that a burgeoning literature extending Ennis and Keister (2009) typically assumes that agents always pool their full endowments to form a coalition, which only has access to a constant-returns-to-scale technology. Yet, in the current context, we incorporate increasing returns as in Andolfatto and Nosal (2018). Furthermore, in comparison to prior contributions such as Gao and Reed (2021), we explicitly study the amount of funds consumers would place in the bank in a pre-deposit game, which allows us to characterize financial fragility more precisely. Note also that the endogenous level of deposits critically depends on how both consumers and the bank are anticipated to respond in periods 1 and 2. Consequently, we first assume that agents deposit a certain level of funds denoted by $\lambda \in (0, 1]$ with a bank in period 0 and analyze backwards.

In the next subsection, we derive the best responses of a bank to the strategy profile (2) by adopting a backwards induction approach. In particular, we first investigate how the intermediary adjusts its payments to depositors after the sunspot state is revealed. Subsequently, we look at what the bank would choose to offer to the first fraction π of depositors. We then study an agent's decision of how much to deposit before turning to address how increasing returns to scale affect the bank's susceptibility to a run along with aggregate welfare.

4.1 The best-response allocations under the scale threshold

4.1.1 Ex-post allocation of consumption

We start by discussing the amount of the bank's remaining resources in the ex-post allocation where the intermediary attracts $\lambda \leq 1$ units of funds initially. The first fraction π of depositors who withdraw in period 1 will receive an amount $c_1(R_A; R_A)$, where the first and second R_A in the parentheses represent the ex-post asset returns in states α and β respectively. The reason for specifying ex-post asset returns in the bank's early payment is as follows. When the intermediary redeems the first π withdrawals, it has not inferred the aggregate state. Further, in the absence of scale economies, the return on assets would be independent of the bank's investment scale and thus the early payoff on deposits could simply be written as c_1. However, because the realization of asset returns varies across scales, the early payment would be conditional across the anticipated ex-post returns in different aggregate states. Thus, the early payout is referred to as $c_1(R_A; R_A)$, and $\lambda - \pi c_1(R_A; R_A)$ denotes the amount of resources left for the remaining $1 - \pi$ redemptions.

To proceed, after the bank has inferred the sunspot signal, it can calculate the fraction of re-

maining depositors who are impatient and will contact the bank in period 1 based on the strategy profile (2). In particular, since runs do not occur in the good state and all impatient depositors have withdrawn from their bank, no remaining depositors will contact intermediaries. Yet, an extra fraction $(1-\pi)\pi$ of depositors will still demand early consumption if a run is underway. In this manner, upon the realization of the panic state, the intermediary will choose to give a common amount $c_{1\beta}(R_A)$ to each remaining impatient depositor in period 1. In addition, each remaining patient depositor who withdraws in period 2 will receive a common amount $c_{2s}(R_A)$ for $s=\alpha,\beta$. Importantly, the value of these payments depends on the ex-post asset returns available and will be chosen to maximize depositors' welfare:

$$V_\alpha(R_A) \equiv \max_{\{c_{2\alpha}(R_A)\}} (1-\pi)\, u(c_{2\alpha}(R_A)) \quad \text{and} \tag{3}$$

$$V_\beta(R_A) \equiv \max_{\left\{c_{1\beta}(R_A),c_{2\beta}(R_A)\right\}} (1-\pi)\left[\pi u(c_{1\beta}(R_A)) + (1-\pi)\, u(c_{2\beta}(R_A))\right] \tag{4}$$

It is worth noting that the solution to this problem provides an efficient allocation of resources conditional on the feasibility of ex-post asset returns in each state. In this vein, we can write the following set of the overall resource constraints:

$$(1-\pi)\, c_{2\alpha}(R_A) = R_A \underbrace{\left[\lambda - \pi c_1(R_A;R_A)\right]}_{\equiv k_\alpha} \quad \text{and} \tag{5}$$

$$(1-\pi)^2 c_{2\beta}(R_A) = R_A \underbrace{\left[\lambda - \pi c_1(R_A;R_A) - (1-\pi)\pi c_{1\beta}(R_A)\right]}_{\equiv k_\beta}. \tag{6}$$

Constraints (5) and (6) state that investment held to maturity denoted by k_s will earn an associated rate of return in the respective sunspot state and the period-2 payoff is determined by the standard pro rata division of the matured assets. Notably, k_s is equal to the amount of resources left upon the revelation of the aggregate state after subtracting payments to any of the remaining impatient consumers in period 1. Due to the extra demand for early consumption after a run is identified as discussed above, the remaining assets kept in the investment technology will be lower in state β than in state α.

Thus, in solving this optimization problem, we divide the above constraint set into three distinct cases according to the ex-post asset returns as described in Table 1 below. Under Case A, if the intermediary keeps a large amount of investment, a high asset return is available for the bank regardless of the realization of the sunspot state. Under Case B, the bank earns a high asset return in the good state but a low return in the event of a run. Under Case C, only low asset returns are possible in both states.

	Asset return R_A	
	State α	State β
Case A	R_H	R_H
Case B	R_H	R_L
Case C	R_L	R_L

Table 1: Three ex-post cases of asset returns

Letting $\mu_{2s}(R_A)$ for $s = \alpha, \beta$ represent the multiplier on the resource constraints with the asset return R_A for $R_A = R_H, R_L$ in equations (5) and (6), the first-order conditions characterizing the payments of $c_{2\alpha}(R_A)$, $c_{1\beta}(R_A)$, and $c_{2\beta}(R_A)$ are:

$$u'(c_{2\alpha}(R_A)) = \mu_{2\alpha}(R_A) \text{ and } u'(c_{1\beta}(R_A)) = R_A u'(c_{2\beta}(R_A)) = R_A \mu_{2\beta}(R_A), \quad (7)$$

where under Case A, $R_A = R_H$ for $s = \alpha, \beta$; under Case B, $R_A = R_H$ for $s = \alpha$ whereas $R_A = R_L$ for $s = \beta$; under Case C, $R_A = R_L$ for $s = \alpha, \beta$. These optimal ex-post risk-sharing conditions indicate that runs, if any, will stop since withdrawing in period 2 becomes a strictly dominant strategy for the remaining patient depositors. Moreover, as is shown in the Appendix, the condition $c_{2\alpha}(R_A) > c_{2\beta}(R_A)$ holds for all cases, implying that period-2 bank payoffs on deposits to consumers should no runs occur are not lower than those in the event of a crisis.

4.1.2 Early payments

Moving backwards, the bank needs to decide how much consumption to give to each of the first fraction π of depositors, recognizing that ex-post it would reallocate the remaining resources according to (7). Since the bank has not inferred the sunspot signal and depositors are risk averse, it will choose to offer the same payment $c_1(R_A; R_A)$ to maximize depositors' expected utility. Note that under the three cases in Table 1, there are five possible scenarios when the bank determines the early payment, as summarized in Table 2. Specifically, Scenarios I and II correspond to Case A where the latter scenario differs from the former one in the sense that the scale threshold is merely satisfied in times of crisis. In a similar manner, Scenarios III and IV are under Case B where Scenario IV differs from III in that the minimum scale requirement is just met in normal times. Lastly, Scenario V is the only situation under Case C.

Case A	Scenario I	$k_\alpha > k_\beta > \kappa$
	Scenario II	$k_\alpha > k_\beta = \kappa$
Case B	Scenario III	$k_\alpha > \kappa > k_\beta$
	Scenario IV	$k_\alpha = \kappa > k_\beta$
Case C	Scenario V	$\kappa > k_\alpha > k_\beta$

Table 2: Five possible allocations of resources

As a start, we concentrate on scenarios where the ex-post asset returns are taken as given when the bank structures its liabilities. In these situations, the value of $c_1(R_A; R_A)$ is chosen to solve:

$$W \equiv \max_{\{c_1(R_A;R_A)\}} \pi u\left(c_1\left(R_A;R_A\right)\right) + \left(1-q\right) V_\alpha\left(R_A\right) + qV_\beta\left(R_A\right)$$

subject to the resource constraints (5) and (6). Based upon the different realized returns on investment, the first-order conditions for this problem are:

$$u'\left(c_1\left(R_H;R_H\right)\right) = R_H\left[\left(1-q\right)\mu_{2\alpha}\left(R_H\right) + q\mu_{2\beta}\left(R_H\right)\right], \tag{8}$$

$$u'\left(c_1\left(R_H;R_L\right)\right) = \left(1-q\right)R_H\mu_{2\alpha}\left(R_H\right) + qR_L\mu_{2\beta}\left(R_L\right), \text{ and} \tag{9}$$

$$u'\left(c_1\left(R_L;R_L\right)\right) = R_L\left[\left(1-q\right)\mu_{2\alpha}\left(R_L\right) + q\mu_{2\beta}\left(R_L\right)\right], \tag{10}$$

which apply to Scenarios I, III, and V respectively. Such optimality conditions suggest that anticipating how much it will earn on assets in state s, the bank balances the marginal value of resources before the state is known against the expected future value of resources in both states.

Yet, the early payment may be chosen to merely achieve the minimum scale. In particular, under Scenario II, the bank chooses $c_1(R_H; R_H)$ to solve:

$$\lambda - \pi c_1\left(R_H;R_H\right) - \left(1-\pi\right)\pi c_{1\beta}\left(R_H\right) = \kappa. \tag{11}$$

That is, after meeting the total withdrawal demand in period 1 should runs occur, the bank invests just enough to qualify for the high asset return in state β. Further, under Scenario IV, the early payment is determined such that the bank is still able to operate under scale economies in state α:

$$\lambda - \pi c_1\left(R_H;R_L\right) = \kappa. \tag{12}$$

4.2 Equilibrium allocations and the existence of run equilibria

The preceding analysis demonstrates the bank's reaction to the strategy profile (2). As previously discussed, there are five possible scenarios in total. Letting the superscripts $I, II, ..., V$ designate

the outcomes under each scenario, we can express the best-response allocation as follows:

$$c^* \equiv \left\{c_1^*\left(R_A; R_A\right), c_{2\alpha}^*\left(R_A\right), c_{1\beta}^*\left(R_A\right), c_{2\beta}^*\left(R_A\right)\right\} \text{ for } * = I, II, ..., V \text{ and } R_A = R_H, R_L.^6$$

Note that the consumption levels specified above are jointly defined by conditions (7) to (12), and the resource constraints (5) and (6). Moreover, let W^* for $* = I, II, ..., V$ represent the equilibrium expected utility delivered by the bank.

Prior to discussing under what conditions the strategy profile (2) is part of an equilibrium, let y denote a complete profile of strategies for all depositors at the bank and χ the repayment strategy for the bank. Notably, we study equilibria of such a withdrawal game where: (i) each depositor determines y_i to maximize her expected utility $\sigma_i\left(\chi, (y_i, y_{-i})\right)$, (ii) the bank chooses χ to maximize the expected utility of its depositors $U\left(\chi, y\right)$, and (iii) each agent's strategy is a best response to the strategies of others. Therefore, the formal definition of an equilibrium is as follows:

Definition 2. *The equilibrium of the withdrawal game at the bank is the profile of strategies* (χ^*, y^*) *such that:*

i. $\sigma_i\left(\chi^*, \left(y_i^*, y_{-i}^*\right)\right) \geq \sigma_i\left(\chi^*, \left(y_i, y_{-i}^*\right)\right)$ *for all* y_i *and all* i;

ii. $U\left(\chi^*, y^*\right) \geq U\left(\chi, y^*\right)$ *for all* χ.

In what follows, we demonstrate conditions under which the profile (2) emerges as part of an equilibrium. To begin, we only focus on actions taken by patient depositors since impatient agents with an urgent need to consume will choose to withdraw in period 1 with certainty. We then begin by considering the behavior of patient individuals with $i > \pi$. Specifically, as there is no uncertainty once the aggregate state is identified, the bank sets returns on deposits such that $c_{2\beta}^*\left(R_A\right) > c_{1\beta}^*\left(R_A\right)$ under the ex-post allocation of consumption. As a result, withdrawing in period 2 is a strictly dominant strategy for a patient depositor $i > \pi$.

To proceed, we consider separately two possible cases for patient agents who have the chance to withdraw before the state of nature is inferred by the bank. In particular, a patient depositor $i \leq \pi$ receives $c_{2\alpha}^*\left(R_A\right)$ should the good aggregate state be realized and $c_1^*\left(R_A; R_A\right)$ if she contacts the bank in period 1. In this regard, we confine our attention to situations in which $c_{2\alpha}^*\left(R_A\right) > c_1^*\left(R_A; R_A\right)$ holds across all scenarios and thus no runs occur in normal times. By comparison, if the panic state is realized at the start of period 1, a patient agent $i \leq \pi$ anticipates to receive $c_{2\beta}^*\left(R_A\right)$ if she waits until period 2 but $c_1^*\left(R_A; R_A\right)$ if she joins the run. Consequently, the partial-run strategy profile specified in (2) is part of an equilibrium if and only if $c_1^*\left(R_A; R_A\right) \geq c_{2\beta}^*\left(R_A\right)$.

[6] See the Appendix for the complete best-response allocations in detail.

We now move on to determine the equilibrium allocation. In particular, we assume that the bank will follow one important criterion. To be specific, in an economy characterized by the parameters $(R_L, R_H, \pi, \gamma, q)$, if multiple best-response allocations are feasible for a given scale threshold κ, the bank would choose the one in which it delivers the highest expected utility to depositors.

To discuss the equilibrium allocation in greater detail, it is helpful to note that based upon the possible allocations described above and the initial deposit level λ, the sum of all depositors' welfare at the bank can be characterized by:

$$\begin{cases} W^I = \frac{\lambda^{1-\gamma}(\psi_{HH}+\pi)^\gamma}{1-\gamma} \\ W^{II} = \frac{\pi^\gamma\left(\lambda-\left[\frac{\pi}{1-\pi}(R_H)^{\frac{\gamma-1}{\gamma}}+1\right]\kappa\right)^{1-\gamma}+(\psi_{HH})^\gamma\left(\left[\frac{\pi}{1-\pi}(R_H)^{\frac{\gamma-1}{\gamma}}+1\right]\kappa\right)^{1-\gamma}}{1-\gamma} \\ W^{III} = \frac{\lambda^{1-\gamma}(\psi_{HL}+\pi)^\gamma}{1-\gamma} \\ W^{IV} = \frac{\pi^\gamma(\lambda-\kappa)^{1-\gamma}+(\psi_{HL})^\gamma\kappa^{1-\gamma}}{1-\gamma} \\ W^V = \frac{\lambda^{1-\gamma}(\psi_{LL}+\pi)^\gamma}{1-\gamma} \end{cases}, \tag{13}$$

where ψ_{HH}, ψ_{HL}, and ψ_{LL}, which are functions of $\phi_s(R_A)$, are defined to be:

$$\psi_{HH} = \left[(1-q)\,\phi_\alpha(R_H) + q\phi_\beta(R_H)\right]^{\frac{1}{\gamma}}, \tag{14}$$

$$\psi_{HL} = \left[(1-q)\,\phi_\alpha(R_H) + q\phi_\beta(R_L)\right]^{\frac{1}{\gamma}}, \tag{15}$$

$$\psi_{LL} = \left[(1-q)\,\phi_\alpha(R_L) + q\phi_\beta(R_L)\right]^{\frac{1}{\gamma}}, \tag{16}$$

$$\phi_\alpha(R_A) = \left[(1-\pi)\,(R_A)^{\frac{1-\gamma}{\gamma}}\right]^\gamma, \text{ and } \phi_\beta(R_A) = \left((1-\pi)\left[\pi + (1-\pi)\,(R_A)^{\frac{1-\gamma}{\gamma}}\right]\right)^\gamma. \tag{17}$$

In seeking to determine the equilibrium allocation, we next consider the feasibility and desirability for the bank to implement the different scenarios by looking at various exogenous thresholds which are required to achieve economies of scale. To begin, first note that as the bank seeks to maximize depositors' welfare and there exists the potential advantage of scale economies, consumers will pool their full endowments, if doing so would allow them to obtain higher expected utility.

Next, consider the amount of investment $k_\beta^I = \frac{\lambda\left(\frac{\psi_{HH}}{\psi_{HH}+\pi}\right)}{\frac{\pi}{1-\pi}(R_H)^{\frac{\gamma-1}{\gamma}}+1}$. Further, define a scale threshold κ_L which is just equal to the value of k_β^I for $\lambda = 1$. In this vein, if individuals deposit all of their funds and $\kappa < \kappa_L$, allocation c^I is implementable and the bank can best respond by pursuing higher asset returns in both sunspot states. More importantly, though other allocations, such as c^{II} to c^V, are feasible, these payout schedules are inferior as the bank delivers the highest expected utility to

depositors by implementing allocation c^I. Thus, we have the following results in the lemma below.

Lemma 1. *Suppose that* $\kappa \in (0, \kappa_L)$. *Under this condition, allocation* c^I *is implemented and* $\lambda = 1$.

Moving forward, we now look to higher minimum scales where $\kappa \geq \kappa_L$ and thus allocation c^I is ruled out. As a start to examine other possibilities, we determine when implementable arrangements include c^{II} to c^V. Yet, prior to discussing the feasibility of allocation c^{II}, it would be useful to examine scale thresholds in which c^{III} can emerge in equilibrium. Such an analysis, as we will discuss below, also sheds light on the implementability of c^{II} and other schedules. In this regard, note that $k_\alpha^{III} = \lambda\left(\frac{\psi_{HL}}{\psi_{HL}+\pi}\right)$ and, as long as $\kappa < k_\alpha^{III}$, allocation c^{III} is feasible. Consequently, we define the maximum size threshold with which c^{III} can be obtained whenever individuals deposit all of their funds at the bank to be $\kappa_m \equiv \frac{\psi_{HL}}{\psi_{HL}+\pi}$.

By comparison, $k_\beta^{III} = \frac{\lambda\left(\frac{\psi_{HL}}{\psi_{HL}+\pi}\right)}{\frac{\pi}{1-\pi}(R_L)^{\frac{\gamma-1}{\gamma}}+1}$. Further, consider a threshold κ_M, which is equal to k_β^{III} for $\lambda = 1$, where $\kappa_M \in (\kappa_L, \kappa_m)$. Notably, along the minimum scale $\kappa \in (\kappa_L, \kappa_M]$ in which c^{III} may be optimal, consumers only deposit a fraction $\frac{\kappa}{\kappa_M}$ of total endowments while the amount of the remaining resources $1 - \frac{\kappa}{\kappa_M}$ will be invested directly under financial autarky. In particular, the reason for partial deposits is that only a lower investment scale is required for the bank to achieve allocation c^{III}; otherwise, the resulting arrangement will be such that $k_\alpha^{III} > k_\beta^{III} > \kappa$ should a larger proportion of funds than $\frac{\kappa}{\kappa_M}$ be deposited. Consequently, this allocation of resources would be inconsistent with the schedule under Scenario III where $k_\alpha^{III} > \kappa > k_\beta^{III}$. Instead, individuals resort to the option of direct investing for the rest of endowments $1 - \frac{\kappa}{\kappa_M}$ as they have access to one coalition only.

We now turn to study the feasibility of other repayment schemes, which builds upon the previous arguments. First, the implementation of c^{II} entails that $k_\alpha^{II} = \left[\frac{\pi}{1-\pi}(R_H)^{\frac{\gamma-1}{\gamma}} + 1\right]\kappa$ for any $\kappa \in [\kappa_L, \kappa_m)$. Given that c^{III} can be exercised over such a range of minimum scales, c^{II} is achievable as well for reasons as follows. To be precise, if $\kappa \in (\kappa_L, \kappa_M]$ where the bank takes full deposits to implement c^{II} but a fraction $\frac{\kappa}{\kappa_M}$ of total endowments to obtain c^{III}, $k_\alpha^{II} > k_\alpha^{III} = \frac{\kappa}{\kappa_M}\left(\frac{\psi_{HL}}{\psi_{HL}+\pi}\right)$. If instead $\kappa \in (\kappa_M, \kappa_m)$ where all funds are pooled for the allocation under each scenario, $k_\alpha^{II} > k_\alpha^{III} = \frac{\psi_{HL}}{\psi_{HL}+\pi}$ for any $\kappa > \frac{\frac{\psi_{HL}}{\psi_{HL}+\pi}}{\frac{\pi}{1-\pi}(R_H)^{\frac{\gamma-1}{\gamma}}+1}$, which necessarily holds as $\kappa_M = \frac{\frac{\psi_{HL}}{\psi_{HL}+\pi}}{\frac{\pi}{1-\pi}(R_L)^{\frac{\gamma-1}{\gamma}}+1} > \frac{\frac{\psi_{HL}}{\psi_{HL}+\pi}}{\frac{\pi}{1-\pi}(R_H)^{\frac{\gamma-1}{\gamma}}+1}$. In addition, should c^{II} and c^{III} be feasible, c^{IV} and c^V can also be implemented. Overall, all payoff schedules, except for allocation c^I, are attainable.

To proceed, which allocation of resources will be on the equilibrium path of play? To address this question of interest, it is straightforward that c^{IV} is dominated by c^{III}. The intuition is as follows. While the bank that just qualifies for scale economies experiences increasing returns during

normal times and is able to raise ex-post consumption levels accordingly, it issues excessively high short-term liabilities, resulting in lower expected utility. Furthermore, delivering allocation c^{III} yields higher welfare than c^{V} as the intermediary best responds by earning higher productive asset returns in normal times. As a result, the bank only makes a choice between c^{II} and c^{III}.

Interestingly, it can easily be shown that if the size threshold κ is close to κ_L, allocation c^{II} dominates c^{III} irrespective of the likelihood of a panic due to the higher level of initial deposits, i.e., $1 > \frac{\kappa_L}{\kappa_M}$, and asset return should a run occur, i.e., $R_H > R_L$. Conversely, suppose that the minimum scale is relatively high where κ approaches κ_M. Under this circumstance, $\lim_{\kappa\to\kappa_M, q\to 0} W^{III} > \lim_{\kappa\to\kappa_M, q\to 0} W^{II}$ as $\lim_{q\to 0}\psi_{HL} = \lim_{q\to 0}\psi_{HH}$. To see this, once the run probability is low enough, the impact of a low asset return on the earliest payoff under Scenario III is absent. Thus, depositors' expected utilities derived from short-term obligations and ex-post consumption contingent on state α are higher than under Scenario II where the inefficiently low early payment is issued. Hence, c^{III} is dominant at this moderately high scale cutoff if bank run risk is negligible. By comparison, as the prospect of a run attains the highest value that is consistent with the strategy profile (2), either allocation may be optimal depending on parameter values.

To formalize the above argument, we next pin down the scope of q for the implementability of c^{II} and c^{III} by examining conditions under which the profile (2) emerges as part of an equilibrium at the threshold κ_M. To start, note that, under Scenario II, $c_1^{II}(R_H; R_H) \geq c_{2\beta}^{II}(R_H)$ if $\kappa \leq \tilde{\kappa}_M = \frac{1-\pi}{\frac{\pi}{1-\pi}R_H + \left[(1-\pi)+\pi(R_H)^{\frac{\gamma-1}{\gamma}}\right]}$. Thus, to ensure that there exists a parameter space in which the incentive to run condition associated with c^{II} is satisfied at κ_M, $\tilde{\kappa}_M \geq \kappa_M$ must hold. In particular, this condition requires:

$$q \leq \frac{(R_L)^{\gamma-1}\left(R_H + (1-\pi)\left[(R_H)^{\frac{\gamma-1}{\gamma}} - (R_L)^{\frac{\gamma-1}{\gamma}}\right]\right)^{-\gamma}\phi_\beta(R_L) - \phi_\alpha(R_H)}{\phi_\beta(R_L) - \phi_\alpha(R_H)} = q_F^{II}. \tag{18}$$

Notably, $q_F^{II} > 0$ if $\pi > \pi_F^{II} = \frac{(R_H)^{\frac{1}{\gamma}} - \left(\frac{R_H}{R_L}\right)^{\frac{1-\gamma}{\gamma}}}{(R_L)^{\frac{\gamma-1}{\gamma}} - \left(\frac{R_H}{R_L}\right)^{\frac{1-\gamma}{\gamma}}}$. That is, at this critical cutoff κ_M, the intermediary behaves more cautiously if q is higher and hence consumers have an incentive to panic if the run probability is relatively low, i.e., $q \leq q_F^{II}$. Further, when the fraction of impatient depositors is large enough, there is room for the bank to be subject to instability since $q_F^{II} > 0$. This is the case as the intermediary infers the sunspot state after the first fraction π of redemptions have been served and if there are more individuals with a pressing demand for their consumption, panics, if any, will be identified relatively late. In this vein, the bank tends to be prone to fragility over higher values of π as it only has few resources left for ex-post payoffs.

In a similar manner, under Scenario III, $c_1^{III}(R_H; R_L) \geq c_{2\beta}^{III}(R_L)$ if:

$$q \leq \frac{\frac{1}{R_L}\phi_\beta(R_L) - \phi_\alpha(R_H)}{\phi_\beta(R_L) - \phi_\alpha(R_H)} = q_F^{III}, \tag{19}$$

where $q_F^{III} > 0$ if $\pi > \pi_F^{III} = \frac{(R_L)^{\frac{1}{\gamma}}(R_H)^{\frac{1-\gamma}{\gamma}} - (R_L)^{\frac{1-\gamma}{\gamma}}}{1-(R_L)^{\frac{1-\gamma}{\gamma}}}$. Hence, when consumers' degree of exposure to liquidity risk is large enough, intermediaries are subject to instability over lower run probabilities. It can further be shown that $\pi_F^{III} < \pi_F^{II} < 1$ if $\gamma > 2$ and $R_L > (R_H)^{\frac{1}{\gamma-1}}$, implying that $q_F^{II} < q_F^{III}$. Therefore, under these conditions, q_F^{II} is the maximum run probability with which banks are unstable under Scenarios II and III.

To summarize the properties of the equilibrium allocation over intermediate minimum scales more generally, we offer the following lemma by confining to situations where the probability of a run is relatively low. For notational convenience, we define q_L^*, q_L^{**}, κ_L^*, and κ_L^{**} in the Appendix.

Lemma 2. *Suppose that* $\kappa \in [\kappa_L, \kappa_m)$, $\gamma \in (2, \infty)$, $R_L \in \left((R_H)^{\frac{1}{\gamma-1}}, R_H\right)$, $\pi \in \left(\pi_F^{II}, 1\right)$, *and* $q \in (0, q_L^*]$. *Under these conditions:*

i. If $\kappa \in [\kappa_L, \kappa_L^*]$, *allocation* c^{II} *is implemented and* $\lambda = 1$*;*

ii. (a) If $\kappa \in (\kappa_L^*, \min\{\kappa_L^{**}, \kappa_M\}]$ *and* $q \in (q_L^{**}, q_L^*]$, *allocation* c^{II} *is implemented and* $\lambda = 1$*;*
(b) If $\kappa \in (\kappa_L^*, \kappa_M]$ *and* $q \in (0, q_L^{**}]$, *allocation* c^{III} *is implemented and* $\lambda = \frac{\kappa}{\kappa_M}$*;*

iii. If $\kappa \in (\kappa_M, \kappa_m)$, *allocation* c^{III} *is implemented and* $\lambda = 1$.

The lemma highlights how both the scale threshold and the likelihood of a run jointly determine the optimal allocation in equilibrium. First, note that when the minimum scale requirement is low enough, i.e., $\kappa \in [\kappa_L, \kappa_L^*]$, the welfare level to depositors would necessarily be higher if the intermediary reacts to qualify for higher asset returns in both aggregate states. By comparison, if the scale threshold is relatively high, i.e., $\kappa \in (\kappa_L^*, \kappa_M]$, offering sufficiently low early payments on deposits to implement allocation c^{II} may not be desirable, especially in a less crisis-prone economy, i.e., $q \in (0, q_L^{**}]$. That is, when runs are less probable, the effect that only a low asset return is realized during a panic is insignificant and hence the bank finds it optimal to achieve allocation c^{III} without reducing payoffs before the aggregate state is inferred. The same is true over relatively higher size thresholds, i.e., $\kappa \in (\kappa_M, \kappa_m)$.

We now turn to examine the equilibrium allocation if $\kappa \geq \kappa_m$. We show that the size threshold is again the key determinant of the intermediary's liabilities as to whether to qualify for increasing returns should no runs occur. For a relatively high degree of tractability, we further assume that the extent of scale economies is weak and define R_H^{IV}, π_F^{IV}, q_F^{IV}, κ_S, and $\tilde{\kappa}_S$ in the Appendix.

Lemma 3. *Suppose that* $\kappa \in [\kappa_m, 1)$, $\gamma \in (2, \infty)$, $R_H \in \left(R_L, R_H^{IV}\right)$, $\pi \in \left(\pi_F^{IV}, 1\right)$, *and* $q \in \left(0, q_F^{IV}\right]$. *Under these conditions, there exists a cutoff* $\kappa_S^* \in [\kappa_S, \tilde{\kappa}_S]$ *such that:*

i. If $\kappa \in [\kappa_m, \kappa_S^*]$, *allocation* c^{IV} *is implemented and* $\lambda = 1$*;*

ii. If $\kappa \in (\kappa_S^*, 1)$, *allocation* c^V *is implemented and* $\lambda = 1$.

When the minimum scale requirement is not too high, it is always optimal for the bank to act to earn the higher asset return in normal times. By doing so, the bank is able to augment the consumption levels for the remaining depositors in state α and thus the welfare level would be higher than that if only low asset returns are available in both sunspot states. If the minimum scale is large enough while scale economies are weak, however, it would certainly be efficient for the intermediary to pursue lower asset returns regardless of the realization of the sunspot state. Intuitively, the welfare gain from providing slightly more consumption should no runs occur is dominated by the welfare loss from offering sufficiently lower payments to depositors prior to identifying the realized aggregate state.

Lastly, upon concluding the equilibrium allocation, we proceed to summarize the existence of run equilibria in an active bank where the results in Lemmas 1 to 3 hold. To begin, we confine to circumstances in which $\kappa \in (0, \kappa_L)$ and therefore it would be optimal for the bank to pursue higher productive asset returns in both sunspot states by implementing allocation c^I. Notably, the fragility condition $c_1^I(R_H; R_H) \geq c_{2\beta}^I(R_H)$ holds if and only if:

$$q \leq \frac{\frac{1}{R_H}\phi_\beta(R_H) - \phi_\alpha(R_H)}{\phi_\beta(R_H) - \phi_\alpha(R_H)} = q_F^I, \tag{20}$$

where $q_F^I > 0$ if $\pi > \pi_F^I = \frac{(R_H)^{\frac{2-\gamma}{\gamma}} - (R_H)^{\frac{1-\gamma}{\gamma}}}{1-(R_H)^{\frac{1-\gamma}{\gamma}}}$ and $\pi_F^I < 1$ if $\gamma > 2$.

In addition to consumers' inclination to panic, it is also important to acknowledge that run equilibria will never occur unless the bank is active. In particular, the bank is able to offer a desirable level of services to agents and therefore deposits are attracted in the initial period when the propensity to run is relatively low. To be precise, $W^I \geq W^A$ if and only if:

$$q \leq \frac{\left(\left[\pi + (1-\pi)(R_L)^{1-\gamma}\right]^{\frac{1}{\gamma}} - \pi\right)^\gamma - \phi_\alpha(R_H)}{\phi_\beta(R_H) - \phi_\alpha(R_H)} = q_A^I \tag{21}$$

and thus run behavior may be observed if $q \in \left(0, \min\left\{q_F^I, q_A^I\right\}\right]$. Prior to characterizing run equilibria, it is helpful to note that $q_F^I < q_A^I$ if $R_L < R_L^I$, where $R_L^I = \left(\frac{\left[(R_H)^{-\frac{1}{\gamma}}\left(\phi_\beta(R_H)\right)^{\frac{1}{\gamma}} + \pi\right]^\gamma - \pi}{1-\pi}\right)^{\frac{1}{1-\gamma}}$

and, as demonstrated in the Appendix, there exists a unique threshold high asset return $\underline{R}_H^I > 1$ below (above) which $R_L^I > 1$ (R_L^I is undefined).

Proposition 1. *Suppose that* $\kappa \in (0, \kappa_L)$, $\gamma \in (2, \infty)$, *and* $R_H \in \left(R_L, \underline{R}_H^I\right)$.

i. If $\pi \in \left(\pi_F^I, \frac{1}{\gamma-1}\right]$, *then* $q_F^I < q_A^I$ *and run equilibria exist* $\forall q \in \left(0, q_F^I\right]$.

ii. If $\pi \in \left(\frac{1}{\gamma-1}, 1\right)$, *there exists a cutoff* R_H^I *such that: (a) if* $R_L \in \left(1, R_L^I\right)$ *and* $R_H \in \left(R_L, \min\left\{\underline{R}_H^I, R_H^I\right\}\right)$, *then* $q_F^I < q_A^I$ *and run equilibria exist* $\forall q \in \left(0, q_F^I\right]$; *(b) if* $R_L \in \left(R_L^I, R_H\right)$ *and* $R_H \in \left(R_L, \min\left\{\underline{R}_H^I, R_H^I\right\}\right)$, *then* $q_A^I < q_F^I$ *and run equilibria exist* $\forall q \in \left(0, q_A^I\right]$; *(c) if* $R_H \in \left(R_H^I, \underline{R}_H^I\right)$, *then* $q_F^I < q_A^I$ *and run equilibria exist* $\forall q \in \left(0, q_F^I\right]$.

Notably, when depositors' risk aversion is high enough, the bank may be susceptible to instability depending on the degree of liquidity risk and the extent of scale economies. On one hand, suppose that liquidity risk is intermediate, i.e., $\pi \in \left(\pi_F^I, \frac{1}{\gamma-1}\right]$, where only a relatively smaller amount of investment is liquidated for the first π withdrawals and large-scale assets are held to maturity. Among these cases, irrespective of the extent of scale economies, the bank is active though the potential for run behavior is higher, indicating that run equilibria exist for any $q \in \left(0, q_F^I\right]$.

On the other hand, suppose that liquidity risk is high enough, i.e., $\pi \in \left(\frac{1}{\gamma-1}, 1\right)$, such that only few resources are left for consumption of the remaining depositors. Then, over lower levels of the high productive asset return, i.e., $R_H \in \left(R_L, \min\left\{\underline{R}_H^I, R_H^I\right\}\right)$, the bank is (not) able to provide sufficient liquidity insurance over higher run probabilities if scale economies are somewhat stronger (weaker), i.e., $R_L \in \left(1, R_L^I\right)$ and $R_H \in \left(R_L, \min\left\{\underline{R}_H^I, R_H^I\right\}\right)$ ($R_L \in \left(R_L^I, R_H\right)$ and $R_H \in \left(R_L, \min\left\{\underline{R}_H^I, R_H^I\right\}\right)$). Given that consumers only have an incentive to panic if the probability of a crisis is low enough, run equilibria exist $\forall q \in \left(0, q_F^I\right]$ ($\forall q \in \left(0, q_A^I\right]$).

By comparison, in alternative situations where the bank earns a sufficiently high productive asset return, i.e., $R_H \in \left(R_H^I, \underline{R}_H^I\right)$, the relatively small investment base still translates into more matured output. In this vein, regardless of the low productive asset return R_L, the intermediary is always capable of offering enough liquidity insurance over a wider range of the prospect of a crisis and therefore run equilibria exist $\forall q \in \left(0, q_F^I\right]$.

We proceed to examine the existence of run equilibria under scenarios where the bank implements allocation c^{II}. To this end, recall from Lemma 2 that as long as $\gamma > 2$, $R_L > (R_H)^{\frac{1}{\gamma-1}}$, and $\pi > \pi_F^{II}$, the fragility condition $c_1^{II}(R_H; R_H) \geq c_{2\beta}^{II}(R_H)$ holds if $\kappa \in [\kappa_L, \kappa_L^*]$ and $q \in (0, q_L^*]$ or $\kappa \in (\kappa_L^*, \kappa_M]$ and $q \in (q_L^{**}, q_L^*]$. In addition, the bank is able to deliver higher welfare than autarky, i.e., $W^{II} \geq W^A$, if and only if:

$$q \leq \frac{\left(\left[\pi+(1-\pi)(R_L)^{1-\gamma}\right]-\pi^{\gamma}\left[1-\left(\phi_\beta(R_H)\right)^{\frac{1}{\gamma}}(R_H)^{\frac{\gamma-1}{\gamma}}\kappa(1-\pi)^{-2}\right]^{1-\gamma}\right)\left[\left(\phi_\beta(R_H)\right)^{\frac{1}{\gamma}}(R_H)^{\frac{\gamma-1}{\gamma}}\kappa(1-\pi)^{-2}\right]^{\gamma-1}-\phi_\alpha(R_H)}{\phi_\beta(R_H)-\phi_\alpha(R_H)} = q_A^{II}. \tag{22}$$

We then offer the following proposition.

Proposition 2. *Suppose that* $\gamma \in (2,\infty)$, $R_L \in \left((R_H)^{\frac{1}{\gamma-1}}, R_H\right)$, *and* $\pi \in \left(\pi_F^{II}, 1\right)$. *Under these conditions:*

i. If $\kappa \in [\kappa_L, \kappa_L^*]$, *run equilibria exist* $\forall q \in \left(0, \min\left\{q_L^*, q_A^{II}\right\}\right]$;

ii. If $\kappa \in (\kappa_L^*, \min\{\kappa_L^{**}, \kappa_M\}]$, *run equilibria exist* $\forall q \in \left(q_L^{**}, \min\left\{q_L^*, q_A^{II}\right\}\right]$.

That is, the above results indicate that either q_L^* or q_A^{II} is the maximum probability of a crisis with which depositors play the run strategy profile under allocation c^{II}. Yet, it is also worth noting that there may exist a lower bound probability of a panic due to parameter restrictions imposed for the implementation of such an allocation in equilibrium.

Next, we study the existence of run equilibria when allocation c^{III} is on the equilibrium path of play. Note that irrespective of whether consumers make partial or full deposits, i.e., $\lambda = \frac{\kappa}{\kappa_M}$ or $\lambda = 1$, the incentive to run condition $c_1^{III}(R_H; R_L) \geq c_{2\beta}^{III}(R_L)$ holds if and only if $q \leq q_F^{III}$, where q_F^{III} is defined in (19). Moreover, recall that if $\pi > \pi_F^{III}$, $q_F^{III} \in (0,1)$ and it can easily be verified that $\pi_F^{III} < \pi_F^I < \pi_F^{II}$ for any $\gamma > 2$ and $R_L > (R_H)^{\frac{1}{\gamma-1}}$. To understand why, note that though only a low productive asset return is available in the event of a crisis, the bank still anticipates to earn a high return in normal times. As a result, it provides more liquidity insurance to the first fraction π of consumers by maintaining a relatively high level of short-term liabilities. Yet, should a poor realization of the sunspot signal occur, the intermediary would reduce payments on deposits significantly, leaving itself excessively prone to fragility. In this manner, the minimum liquidity risk to induce run behavior under allocation c^{III} is not as high as those associated with allocations c^I and c^{II}. In addition, given that $\pi_F^{III} < \pi_F^I < \pi_F^{II}$, $q_L^{**} \leq q_L^* \leq q_F^{II} < q_F^I < q_F^{III}$ $\forall \pi > \pi_F^{II}$.

Similarly, it is non-trivial that consumers must have an incentive to deposit their funds at the bank in order for allocation c^{III} to be implemented. To be precise, for $\lambda = \frac{\kappa}{\kappa_M}$ or $\lambda = 1$, $W^{III} \geq W^A$ if and only if:

$$q \leq \frac{\left(\left[\pi + (1-\pi)(R_L)^{1-\gamma}\right]^{\frac{1}{\gamma}} - \pi\right)^{\gamma} - \phi_\alpha(R_H)}{\phi_\beta(R_L) - \phi_\alpha(R_H)} = q_A^{III} \tag{23}$$

and hence runs may emerge in equilibrium $\forall q \in \left(0, \min\left\{q_L^{**}, q_A^{III}\right\}\right]$ if $\kappa \in (\kappa_L^*, \kappa_M]$ or $\forall q \in \left(0, \min\left\{q_L^*, q_A^{III}\right\}\right]$ if $\kappa \in (\kappa_M, \kappa_m]$. Note also that $\pi_F^{II} \lesseqgtr \frac{1}{\gamma-1}$ if $R_L \gtreqless R_L^{II} = \left(\frac{(R_H)^{\frac{\gamma-1}{\gamma}} + (\gamma-2)}{R_H(\gamma-1)}\right)^{\frac{\gamma}{1-\gamma}}$, where $R_L^{II} \in \left((R_H)^{\frac{1}{\gamma-1}}, R_H\right)$, we then summarize the existence of run equilibria below:

Proposition 3. *Suppose* $\gamma \in (2,\infty)$ *and* $\pi \in \left(\pi_F^{II}, 1\right)$.

i. Further, suppose that $R_L \in \left((R_H)^{\frac{1}{\gamma-1}}, R_L^{II}\right)$*. Under these conditions, there exists a threshold* R_L^{III} *such that: (a) if* $R_L \in \left((R_H)^{\frac{1}{\gamma-1}}, \min\left\{R_L^{II}, R_L^{III}\right\}\right)$ *and* $\kappa \in (\kappa_L^*, \kappa_M]$ *(*$\kappa \in (\kappa_M, \kappa_m]$*), then* $q_A^{III} < q_F^{III}$ *and run equilibria exist* $\forall q \in \left(0, \min\left\{q_L^{**}, q_A^{III}\right\}\right]$ *(*$\forall q \in \left(0, \min\left\{q_L^{*}, q_A^{III}\right\}\right]$*); (b) if* $R_L \in \left(\max\left\{(R_H)^{\frac{1}{\gamma-1}}, R_L^{III}\right\}, R_L^{II}\right)$ *and* $\kappa \in (\kappa_L^*, \kappa_M]$ *(*$\kappa \in (\kappa_M, \kappa_m]$*), then* $q_F^{III} < q_A^{III}$ *and run equilibria exist* $\forall q \in (0, q_L^{**}]$ *(*$\forall q \in (0, q_L^{*}]$*).*

ii. Suppose instead that $R_L \in \left(R_L^{II}, R_H\right)$*. Then, if* $\pi \in \left(\pi_F^{II}, \frac{1}{\gamma-1}\right]$ *and* $\kappa \in (\kappa_L^*, \kappa_M]$ *(*$\kappa \in (\kappa_M, \kappa_m]$*), then* $q_F^{III} < q_A^{III}$ *and run equilibria exist* $\forall q \in (0, q_L^{**}]$ *(*$\forall q \in (0, q_L^{*}]$*). Yet, if* $\pi \in \left(\frac{1}{\gamma-1}, 1\right)$*, there exists a threshold* R_L^{III} *such that: (a) if* $R_L \in \left(R_L^{II}, \min\left\{R_L^{III}, R_H\right\}\right)$ *and* $\kappa \in (\kappa_L^*, \kappa_M]$ *(*$\kappa \in (\kappa_M, \kappa_m]$*), then* $q_A^{III} < q_F^{III}$ *and run equilibria exist* $\forall q \in \left(0, \min\left\{q_L^{**}, q_A^{III}\right\}\right]$ *(*$\forall q \in \left(0, \min\left\{q_L^{*}, q_A^{III}\right\}\right]$*); (b) if* $R_L \in \left(\max\left\{R_L^{II}, R_L^{III}\right\}, R_H\right)$ *and* $\kappa \in (\kappa_L^*, \kappa_M]$ *(*$\kappa \in (\kappa_M, \kappa_m]$*), then* $q_F^{III} < q_A^{III}$ *and run equilibria exist* $\forall q \in (0, q_L^{**}]$ *(*$\forall q \in (0, q_L^{*}]$*).*

When agents are sufficiently risk-averse and the population of impatient depositors is relatively large, i.e., $\gamma \in (2, \infty)$ and $\pi \in \left(\pi_F^{II}, 1\right)$, the existence of run equilibria again hinges on the extent of scale economies. On the one hand, over lower levels of R_L, if economies of scale are stronger (weaker), i.e., $R_L \in \left((R_H)^{\frac{1}{\gamma-1}}, \min\left\{R_L^{II}, R_L^{III}\right\}\right)$ ($R_L \in \left(\max\left\{(R_H)^{\frac{1}{\gamma-1}}, R_L^{III}\right\}, R_L^{II}\right)$), the bank is able to offer higher welfare than autarky in a more (less) crisis-prone economy. As a result, run equilibria exist $\forall q \in \left(0, \min\left\{q_L^{**}, q_A^{III}\right\}\right]$ if $\kappa \in (\kappa_L^*, \kappa_M]$ while $\forall q \in \left(0, \min\left\{q_L^{*}, q_A^{III}\right\}\right]$ if $\kappa \in (\kappa_M, \kappa_m]$ ($\forall q \in (0, q_L^{**}]$ if $\kappa \in (\kappa_L^*, \kappa_M]$ while $\forall q \in (0, q_L^{*}]$ if $\kappa \in (\kappa_M, \kappa_m]$).

In opposite scenarios where scale economies are relatively weak, i.e., $R_L \in \left(R_L^{II}, R_H\right)$, consumers' degree of exposure to liquidity risk also plays an important role for the existence of run equilibria. In particular, over intermediate degrees of liquidity risk, i.e., $\pi \in \left(\pi_F^{II}, \frac{1}{\gamma-1}\right]$, the bank can still augment depositors' expected utility above autarkic welfare when the propensity to run is high, indicating that run equilibria exist $\forall q \in (0, q_L^{**}]$ if $\kappa \in (\kappa_L^*, \kappa_M]$ or $\forall q \in (0, q_L^{*}]$ if $\kappa \in (\kappa_M, \kappa_m]$. Nevertheless, if the level of investment held to maturity is low enough, i.e., $\pi \in \left(\frac{1}{\gamma-1}, 1\right)$, the intermediary is active over higher run probabilities should the low asset return R_L be somewhat strong, i.e., $R_L \in \left(\max\left\{R_L^{II}, R_L^{III}\right\}, R_H\right)$. In this vein, runs may appear over a wider range of the potential for consumer sentiment to deteriorate, i.e., $\forall q \in (0, q_L^{**}]$ if $\kappa \in (\kappa_L^*, \kappa_M]$ or $\forall q \in (0, q_L^{*}]$ if $\kappa \in (\kappa_M, \kappa_m]$.

Moving forward, we now turn to examine the existence of run equilibria when allocation c^{IV} is obtained. Recall from Lemma 3 that if $\kappa \in [\kappa_m, \kappa_S^*]$, $\gamma > 2$, $\pi > \pi_F^{IV}$, $q \leq q_F^{IV}$, and $R_H < R_H^{IV}$, the fragility condition $c_1^{IV}(R_H; R_L) \geq c_{2\beta}^{IV}(R_L)$ is satisfied. In addition, $W^{IV} \geq W^A$ if:

$$q \leq \frac{\left(\left[\pi + (1-\pi)(R_L)^{1-\gamma}\right] - \pi^{\gamma}(1-\kappa)^{1-\gamma}\right)\kappa^{\gamma-1} - \phi_{\alpha}(R_H)}{\phi_{\beta}(R_L) - \phi_{\alpha}(R_H)} = q_A^{IV}. \tag{24}$$

We then summarize the existence of run equilibria associated with allocation c^{IV} below:

Proposition 4. *Suppose that* $\kappa \in [\kappa_m, \kappa_S^*]$, $\gamma \in (2, \infty)$, $\pi \in \left(\pi_F^{IV}, 1\right)$, *and* $R_H \in \left(R_L, R_H^{IV}\right)$. *Under these conditions, run equilibria exist* $\forall q \in \left(0, \min\left\{q_F^{IV}, q_A^{IV}\right\}\right]$.

Again, the above proposition suggests that either q_F^{IV} or q_A^{IV} is the cutoff likelihood of a crisis below which runs may be observed.

Lastly, we discuss the existence of run equilibria should allocation c^V be on the equilibrium path of play. Specifically, the incentive to run condition $c_1^V\left(R_L; R_L\right) \geq c_{2\beta}^V\left(R_L\right)$ is not violated if and only if $q \leq q_F^{IV}$. Moreover, $W^V \geq W^A$ if:

$$q \leq \frac{\left(\left[\pi + (1-\pi)\left(R_L\right)^{1-\gamma}\right]^{\frac{1}{\gamma}} - \pi\right)^{\gamma} - \phi_\alpha\left(R_L\right)}{\phi_\beta\left(R_L\right) - \phi_\alpha\left(R_L\right)} = q_A^V. \tag{25}$$

As a result, run equilibria exist $\forall q \in \left(0, \min\left\{q_F^{IV}, q_A^V\right\}\right]$:

Proposition 5. *Suppose that* $\kappa \in (\kappa_S^*, 1)$, $\gamma \in (2, \infty)$, *and* $R_H \in \left(R_L, R_H^{IV}\right)$.

i. If $\pi \in \left(\pi_F^{IV}, \frac{1}{\gamma-1}\right]$, *then* $q_F^{IV} < q_A^V$ *and run equilibria exist* $\forall q \in \left(0, q_F^{IV}\right]$;

ii. If $\pi \in \left(\frac{1}{\gamma-1}, 1\right)$, *there exists a threshold* R_L^V *such that: (a) if* $R_L \in \left(1, \min\left\{R_L^V, R_H\right\}\right)$, *then* $q_A^V < q_F^{IV}$ *and run equilibria exist if* $q \in \left(0, q_A^V\right]$; *(b) if* $R_L \in \left(R_L^V, R_H\right)$, *then* $q_F^{IV} < q_A^V$ *and run equilibria exist* $\forall q \in \left(0, q_F^{IV}\right]$.

If liquidity risk is not too high, for example, the bank still holds a relatively large amount of investment and will be able to offer better welfare than autarky even though panics are more likely to occur. As a consequence, the threshold crisis probability below which bank run equilibria exist is q_F^{IV}. Nevertheless, suppose that there are more depositors who have an immediate demand to consume early. Then, the bank only has a small amount of the remaining resources kept in investment after meeting the total need for early consumption. If the return R_L is relatively low, the underprovision of risk-sharing services in a more crisis-prone economy implies that the bank is only active when the prospect of a panic is low and hence run equilibria exist if and only if $q \in \left(0, q_A^V\right]$.

4.3 The role of scale economies for fragility and welfare

We now compare instability and aggregate welfare across different minimum scale requirements and also discuss their trade-offs. As a baseline, it would be useful to understand the implications of constant-returns-to-scale technologies for financial fragility and depositors' expected utility. To

this end, note that the inverse of the minimum scale can serve as a proxy for the relative amount of bank investment. Note also that constant-returns-to-scale technologies imply that the scale requirement κ is absent where regardless of the size of the bank's attracted deposits, the intermediary will behave in the same manner. Alternatively, $R_A = R_L = R_H$ and the best-response allocation is equivalent to allocation c^V. Given that $W^V = \frac{\lambda^{1-\gamma}(\psi_{LL}+\pi)^\gamma}{1-\gamma}$ for a level of bank deposits λ and the satisfaction of the incentive to run condition $c_1^V(R_L; R_L) \geq c_{2\beta}^V(R_L)$ does not depend on κ, we then offer the following proposition:

Proposition 6. *Suppose that economies of scale are absent and $\gamma \in (2, \infty)$. Further suppose that $\pi \in \left(\pi_F^{IV}, \frac{1}{\gamma-1}\right]$ such that run equilibria exist $\forall q \in \left(0, q_F^{IV}\right]$ or $\pi \in \left(\frac{1}{\gamma-1}, 1\right)$ and $R_L \in \left(1, R_L^V\right)$ ($R_L \in \left(R_L^V, \infty\right)$) such that run equilibria exist $\forall q \in \left(0, q_A^V\right]$ ($\forall q \in \left(0, q_F^{IV}\right]$). Under these conditions, the degree of instability is independent of the size of the bank's deposit base. In addition, there exhibits a linear relationship between the amount of bank investment and welfare.*

Intuitively, absent increasing returns, payments on deposits issued by the bank are scaled in proportion to the size of its attracted deposits. Consequently, changes in the amount of bank investment alter the sum of depositors' expected utilities proportionally but do not affect its susceptibility to a crisis. In this vein, these predictions indicate that intermediaries with different levels of investment do not exhibit any differences in efficiencies in the use of their resources for maturity transformation and the fragility is not affected by the size of assets.

In contrast, once scale economies are taken into consideration, a non-monotonicity of fragility and aggregate welfare with respect to the amount of investment (the minimum scale) arises. Rather than making every pair-wise comparison to unveil these non-linear relationships, we only focus on highlighting the positive and negative effects of increasing returns on an intermediary's vulnerability to a crisis along with the sum of agents' expected utilities.

As a starting point, we examine the role of increasing returns for financial fragility where conditional on the existence of run equilibria, we compare depositors' incentives to panic across different scale thresholds.

Proposition 7. *Suppose that $\gamma \in (2, \infty)$ and $\pi \in \left(\pi_F^{II}, 1\right)$. Under these conditions:*

i. If $R_H \in \left(R_L, (R_L)^{\gamma-1}\right)$ and $q \in \left(0, \min\left\{q_L^*, q_A^{II}\right\}\right]$ ($q \in \left(q_L^{**}, \min\left\{q_L^*, q_A^{II}\right\}\right]$), $\frac{c_1^I(R_H;R_H)}{c_{2\beta}^I(R_H)} > \frac{c_1^{II}(R_H;R_H)}{c_{2\beta}^{II}(R_H)} \geq 1$; if $R_H \in \left(R_L, \min\left\{(R_L)^{\gamma-1}, R_H^{IV}\right\}\right)$ and $q \in \left(0, \min\left\{q_L^{**}, q_A^{IV}\right\}\right]$ ($q \in \left(0, \min\left\{q_L^*, q_A^{IV}\right\}\right]$), $\frac{c_1^{III}(R_H;R_L)}{c_{2\beta}^{III}(R_L)} > \frac{c_1^{IV}(R_H;R_L)}{c_{2\beta}^{IV}(R_L)} \geq 1$.

ii. If $R_H \in \left(R_L, \min\left\{(R_L)^{\gamma-1}, R_H^{IV}\right\}\right)$ and $q \in \left(0, \min\left\{q_L^{**}, q_A^V\right\}\right]$ ($q \in \left(0, \min\left\{q_L^*, q_A^V\right\}\right]$), $\frac{c_1^{III}(R_H;R_L)}{c_{2\beta}^{III}(R_L)} > \frac{c_1^*(R_A;R_A)}{c_{2\beta}^*(R_A)} \geq 1$ for any $* = I, II, IV, V$.

To begin, as stated in case i, depositors have more incentive to run over relatively small minimum scales than do those in economies where the scale threshold is higher. This necessarily occurs as the bank implementing allocation c^{II} (c^{IV}) behaves cautiously when determining short-term liabilities before the aggregate state is revealed. Recall that the inverse of the minimum size can serve as a proxy for the relative amount of bank investment. In this regard, one may argue that for a pair of comparable intermediaries that admit the same minimum scale requirement, there exist possibilities that a smaller bank is more stable than its larger counterpart.

Nevertheless, is a relatively large minimum scale always associated with weaker degrees of instability? No. To see why a higher cutoff κ may worsen banking stability, we now focus on the results presented in case ii. For example, due to its excessively high short-term liabilities and low levels of matured output should runs occur, the bank implementing allocation c^{III} is more vulnerable to a crisis than that pursuing higher asset returns in both sunspot states. In fact, it can also be shown that the parameter space in which run equilibria exist under c^{III} is largest when comparing to those of any other feasible allocations.

Lastly, we turn to examine the impact of increasing returns on equilibrium aggregate welfare.

Proposition 8. *Suppose that* $\gamma \in (2, \infty)$, $\pi \in \left(\pi_F^{II}, 1\right)$, *and* $R_H \in \left(R_L, \min\left\{(R_L)^{\gamma-1}, R_H^{IV}\right\}\right)$. *Under these conditions:*

i. For any $q \in \left(0, \min\left\{q_L^*, q_A^V\right\}\right]$ *(*$q \in \left(0, \min\left\{q_L^{**}, q_A^V\right\}\right]$*),* $W^I > \max\left\{W^{II}, W^{III}, W^{IV}, W^V\right\}$*;*

ii. If the run probability q *is arbitrarily low and agents pool their full endowments in the bank,* $\max\left\{W^I, W^{II}, ..., W^V\right\} = W^I = W^{III}$.

In the presence of increasing returns to scale, it is straightforward that liquidity provision improves over low scale thresholds, as suggested by case i. If, instead, the minimum scale requirement is high, the bank must structure liabilities cautiously so that a higher asset return is earned or it fails to qualify for increasing returns. Consequently, consumers are only able to obtain the low welfare level due to prudent banking behavior or lower returns on bank investment.

However, the results in case ii imply that a large minimum scale is not necessarily associated with a low level of aggregate utility. To be precise, note that similar to that over low size cutoffs, the bank confronting the intermediate threshold is still qualified to earn the high productive asset return during normal times. While only low returns on investment are available in the event of a run, such effects on both ex-post payments and payoffs on the earliest demandable deposits are negligible as long as the run probability is arbitrarily low. Given that the payment scheme should no panics occur is identical to that under the low scale cutoffs, the bank with the medium minimum scale requirement is still able to deliver the highest expected utility. Yet, this only occurs at costs of exacerbated financial fragility, as implied by Proposition 7.

5 Conclusion

In recent years, there is a bulk of empirical research that shows that scale economies play a critical role in banking activity. Given the existence and prevalence of increasing returns in the banking sector, it is important to develop an empirically relevant framework to investigate how economies of scale affect maturity transformation, risk-sharing, and financial stability. For this purpose, we build on a version of Diamond and Dybvig (1983) with limited commitment where the bank subject to instability has access to an increasing-returns-to-scale investment technology.

It is worth noting that when depositors' preferences are of the constant-relative-risk-aversion variety and economies of scale are absent, the degree of financial fragility is independent of the level of investment and aggregate welfare is proportional to the size of the bank's deposit base. Yet, once the minimum scale effect is present, these conventional results may be overturned. In particular, there generally exhibits a non-monotonicity of banking fragility and consumers' expected utilities with respect to the level of bank investment. Specifically, the bank facing an intermediate scale threshold for the high asset return may deliver the highest welfare to depositors but is most prone to panics, especially if runs are highly unlikely. That is, due to increasing returns in normal times, the intermediary provides high payments on deposits in the short-term but only low output will be realized should run behavior emerge.

There are many important extensions to consider for future work. For example, banking stability may also be affected by the moral hazard problem related to some forms of public insurance such as deposit insurance and government bailouts. Moreover, in light of the recent trend towards increasing banking concentration in many advanced economies, adopting a singular focus on the sources of fragility under perfect competition may not be enough. Thus, future work which looks at these forms of the government safety net and/or rent-seeking, strategic behavior but where economies of scale exist is important.

6 Technical appendix

The best-response allocations under the scale threshold

The best-response allocations of a bank are determined by conditions (7) to (12), and the resource constraints (5) and (6). Recall that ψ_{HH}, ψ_{HL}, ψ_{LL}, and $\phi_s(R_A)$ are defined in (14) to (17) while κ_L, κ_L^*, κ_L^{**}, κ_M, κ_m, κ_S^*, q_L^*, q_L^{**}, and R_H^{IV} are defined in Lemmas 1 to 3. Notably, the solution to the bank's optimization problem belongs to either Scenario I, II, III, IV, or V, depending on the value of κ, q, and R_A.

Scenarios I, III, and V: $k_\alpha^I > k_\beta^I > \kappa$, $k_\alpha^{III} > \kappa > k_\beta^{III}$, and $\kappa > k_\alpha^V > k_\beta^V$

If $\kappa \in (0, \kappa_L)$, allocation c^I is implemented where consumers pool their full endowments, i.e., $\lambda = 1$; if $\kappa \in (\kappa_L^*, \kappa_M]$ and $q \in (0, q_L^{**}]$ or $\kappa \in (\kappa_M, \kappa_m]$ and $q \in (0, q_L^*]$, allocation c^{III} is implemented where consumers pool only $\frac{\kappa}{\kappa_M}$ units of goods for any $\kappa \in (\kappa_L^*, \kappa_M]$, i.e., $\lambda = \frac{\kappa}{\kappa_M}$, but pool their full endowments for any $\kappa \in (\kappa_M, \kappa_m]$, i.e., $\lambda = 1$; if $R_H \in \left(R_L, R_H^{IV}\right)$ and $\kappa \in (\kappa_S^*, 1)$, allocation c^V is implemented where consumers deposit their full endowments, i.e., $\lambda = 1$. The vector $c' = \left\{c_1'(R_A; R_A), c_{2\alpha}'(R_A), c_{1\beta}'(R_A), c_{2\beta}'(R_A)\right\}$, where $\prime = I, III, V$, is thus given by:

$$
\begin{aligned}
c_1'(R_A; R_A) &= \lambda \left(\frac{1}{\psi_{A(\alpha)A(\beta)} + \pi} \right), \\
c_{2\alpha}'(R_A) &= \lambda \left(\frac{R_A}{\phi_\alpha(R_A)} \right)^{\frac{1}{\gamma}} \left(1 - \pi \left(\frac{1}{\psi_{A(\alpha)A(\beta)} + \pi} \right) \right), \\
c_{1\beta}'(R_A) &= \lambda \left(\frac{1}{\phi_\beta(R_A)} \right)^{\frac{1}{\gamma}} \left(1 - \pi \left(\frac{1}{\psi_{A(\alpha)A(\beta)} + \pi} \right) \right), \\
c_{2\beta}'(R_A) &= \lambda \left(\frac{R_A}{\phi_\beta(R_A)} \right)^{\frac{1}{\gamma}} \left(1 - \pi \left(\frac{1}{\psi_{A(\alpha)A(\beta)} + \pi} \right) \right),
\end{aligned}
$$

where under Scenario I, $R_A = R_H$ for $s = \alpha, \beta$ and $\psi_{A(\alpha)A(\beta)} = \psi_{HH}$; under Scenario III, $R_A = R_H$ for $s = \alpha$, $R_A = R_L$ for $s = \beta$, and $\psi_{A(\alpha)A(\beta)} = \psi_{HL}$; under Scenario V, $R_A = R_L$ for $s = \alpha, \beta$ and $\psi_{A(\alpha)A(\beta)} = \psi_{LL}$. Note also that $\frac{c_{2\beta}'(R_A)}{c_{1\beta}'(R_A)} = (R_A)^{\frac{1}{\gamma}} > 1$. In addition, combining conditions (7) to (10) yields: $\frac{c_{2\alpha}^I(R_H)}{c_1^I(R_H;R_H)} = \left(R_H \left[(1-q) + q\frac{\phi_\beta(R_H)}{\phi_\alpha(R_H)} \right] \right)^{\frac{1}{\gamma}} > 1$, $\frac{c_{2\alpha}^{III}(R_H)}{c_1^{III}(R_H;R_L)} = \left(R_H \left[(1-q) + q\frac{\phi_\beta(R_L)}{\phi_\alpha(R_H)} \right] \right)^{\frac{1}{\gamma}} > 1$, and $\frac{c_{2\alpha}^V(R_L)}{c_1^V(R_L;R_L)} = \left(R_L \left[(1-q) + q\frac{\phi_\beta(R_L)}{\phi_\alpha(R_L)} \right] \right)^{\frac{1}{\gamma}} > 1$ and we restrict our attention to situations where $\frac{c_1'(R_A;R_A)}{c_{2\beta}'(R_A)} \geq 1$ holds.

Scenario II: $k_\alpha^{II} > k_\beta^{II} = \kappa$

If $\kappa \in [\kappa_L, \kappa_L^*]$ and $q \in (0, q_L^*]$ or $\kappa \in (\kappa_L^*, \min\{\kappa_L^{**}, \kappa_M\}]$ and $q \in (q_L^{**}, q_L^*]$, allocation c^{II} is implemented where consumers pool their full endowments, i.e., $\lambda = 1$. In this vein, the vector $c^{II} = \left\{c_1^{II}(R_H; R_H), c_{2\alpha}^{II}(R_H), c_{1\beta}^{II}(R_H), c_{2\beta}^{II}(R_H)\right\}$ is given by:

$$\begin{aligned}
c_1^{II}(R_H; R_H) &= \frac{\lambda - \left[\frac{\pi}{1-\pi}(R_H)^{\frac{\gamma-1}{\gamma}} + 1\right]\kappa}{\pi}, \\
c_{2\alpha}^{II}(R_H) &= \frac{\left[\pi + (1-\pi)(R_H)^{\frac{1-\gamma}{\gamma}}\right](R_H)^{\frac{2\gamma-1}{\gamma}}\kappa}{(1-\pi)^2}, \\
c_{1\beta}^{II}(R_H) &= \frac{(R_H)^{\frac{\gamma-1}{\gamma}}\kappa}{(1-\pi)^2}, \\
c_{2\beta}^{II}(R_H) &= \frac{R_H\kappa}{(1-\pi)^2},
\end{aligned}$$

where $\frac{c_{2\beta}^{II}(R_H)}{c_{1\beta}^{II}(R_H)} = (R_H)^{\frac{1}{\gamma}} > 1$. Further, we restrict our attention to situations where $\frac{c_{2\alpha}^{II}(R_H)}{c_1^{II}(R_H;R_H)} > 1$ and $\frac{c_1^{II}(R_H;R_H)}{c_{2\beta}^{II}(R_H)} \geq 1$ hold.

Scenario IV: $k_\alpha^{IV} = \kappa > k_\beta^{IV}$

If $\kappa \in [\kappa_m, \kappa_S^*)$ and $R_H \in \left(R_L, R_H^{IV}\right)$, allocation c^{IV} is implemented where consumers pool their full endowments, i.e., $\lambda = 1$. The vector $c^{IV} = \left\{c_1^{IV}(R_H; R_L), c_{2\alpha}^{IV}(R_H), c_{1\beta}^{IV}(R_L), c_{2\beta}^{IV}(R_L)\right\}$ is thus given by:

$$\begin{aligned}
c_1^{IV}(R_H; R_L) &= \frac{\lambda - \kappa}{\pi}, \\
c_{2\alpha}^{IV}(R_H) &= \frac{R_H\kappa}{1-\pi}, \\
c_{1\beta}^{IV}(R_L) &= \frac{\kappa}{(1-\pi)\left[\pi + (1-\pi)(R_L)^{\frac{1-\gamma}{\gamma}}\right]}, \\
c_{2\beta}^{IV}(R_L) &= \frac{(R_L)^{\frac{1}{\gamma}}\kappa}{(1-\pi)\left[\pi + (1-\pi)(R_L)^{\frac{1-\gamma}{\gamma}}\right]},
\end{aligned}$$

where $\frac{c_{2\beta}^{IV}(R_L)}{c_{1\beta}^{IV}(R_L)} = (R_L)^{\frac{1}{\gamma}} > 1$. Further, we restrict our attention to situations where $\frac{c_{2\alpha}^{IV}(R_H)}{c_1^{IV}(R_H;R_L)} > 1$ and $\frac{c_1^{II}(R_H;R_H)}{c_{2\beta}^{II}(R_H)} \geq 1$ hold.

■

Proof of Lemma 1

See the discussion in the main text.

■

Proof of Lemma 2

First, suppose that $\kappa \in [\kappa_L, \kappa_m)$. Under this condition, as κ goes to κ_L, allocation c^{II} mirrors c^I for $\lambda = 1$. Thus, $\lim_{\kappa \to \kappa_L} W^I = \lim_{\kappa \to \kappa_L} W^{II} > \lim_{\kappa \to \kappa_L} W^{III}$. To proceed, suppose that κ goes to κ_M. Further, suppose that the run probability q is sufficiently low. Under these conditions and for $\lambda = 1$, $\lim_{\kappa \to \kappa_M, q \to 0} W^{III} > \lim_{\kappa \to \kappa_M, q \to 0} W^{II}$ as $\lim_{q \to 0} \psi_{HL} = \lim_{q \to 0} \psi_{HH}$.

In what follows, we conduct the welfare comparison between $\lim_{\kappa \to \kappa_M} W^{II}$ and $\lim_{\kappa \to \kappa_M} W^{III}$ as q attains the highest value consistent with the strategy profile (2). Notably, as κ goes to κ_M, $c_1^{II}(R_H; R_H) \geq c_{2\beta}^{II}(R_H)$ if $q \leq q_F^{II}$, where $q_F^{II} \in (0,1)\ \forall \pi > \pi_F^{II}$. Moreover, $c_1^{III}(R_H; R_L) \geq c_{2\beta}^{III}(R_L)$ if $q \leq q_F^{III}$, where $q_F^{III} \in (0,1)\ \forall \pi > \pi_F^{III}$. It can also be verified that $\pi_F^{III} < \pi_F^{II} < 1$ if $\gamma > 2$ and $R_L > (R_H)^{\frac{1}{\gamma - 1}}$ and thus $q_F^{II} < q_F^{III}$. In this vein,

$$\lim_{\kappa \to \kappa_M, q \to q_F^{II}} \left(W^{II} - W^{III}\right)$$

$$= \frac{\pi^{\gamma}(\pi R_H)^{1-\gamma} + \left(\psi_{HH}\left(q_F^{II}\right)\right)^{\gamma}\left[\left(\phi_{\beta}(R_H)\right)^{\frac{1}{\gamma}}(R_H)^{\frac{\gamma-1}{\gamma}}\right]^{1-\gamma} - \left(\psi_{HL}\left(q_F^{II}\right)+\pi\right)^{\gamma}\left[\pi R_H + \left(\phi_{\beta}(R_H)\right)^{\frac{1}{\gamma}}(R_H)^{\frac{\gamma-1}{\gamma}}\right]^{1-\gamma}}{(1-\gamma)\left[\pi R_H + \left(\phi_{\beta}(R_H)\right)^{\frac{1}{\gamma}}(R_H)^{\frac{\gamma-1}{\gamma}}\right]^{1-\gamma}}$$

where the sign of $\lim_{\kappa \to \kappa_M, q \to q_F^{II}} \left(W^{II} - W^{III}\right)$ depends on parameter values. Yet, to seek for the robust results, it is helpful to derive the following properties of $\lim_{\kappa \to \kappa_M} W^{II}$ and $\lim_{\kappa \to \kappa_M} W^{III}$:

$$\frac{d}{dq}\left(\lim_{\kappa \to \kappa_M} W^{II}\right) < 0,\ \frac{d}{d^2 q}\left(\lim_{\kappa \to \kappa_M} W^{II}\right) > 0,\ \frac{d}{dq}\left(\lim_{\kappa \to \kappa_M} W^{III}\right) < 0,\ \text{and}\ \frac{d}{d^2 q}\left(\lim_{\kappa \to \kappa_M} W^{III}\right) > 0.$$

That is, both $\lim_{\kappa \to \kappa_M} W^{II}$ and $\lim_{\kappa \to \kappa_M} W^{III}$ are decreasing and strictly convex functions of q. It can also be shown that $\frac{d}{dq}\left(\lim_{\kappa \to \kappa_M} W^{III}\right)\Big|_{q=0} < \frac{d}{dq}\left(\lim_{\kappa \to \kappa_M} W^{II}\right)\Big|_{q=0} < 0$. By the strict convexity of $\lim_{\kappa \to \kappa_M} W^{II}$ and $\lim_{\kappa \to \kappa_M} W^{III}$ with respect to q, $\lim_{\kappa \to \kappa_M, q \to 0} W^{III} > \lim_{\kappa \to \kappa_M, q \to 0} W^{II}$, and $\lim_{\kappa \to \kappa_M, q \to q_F^{II}} \left(W^{II} - W^{III}\right) \lesseqgtr 0$, we then define a unique threshold probability $q_L^* \in \left(0, q_F^{II}\right]$ below which $\lim_{\kappa \to \kappa_M} W^{III} > \lim_{\kappa \to \kappa_M} W^{II}$.

In addition, differentiating W^{II} with respect to $c_1^{II}(R_H; R_H)$ yields:

$$\frac{dW^{II}}{dc_1^{II}(R_H; R_H)} = \pi\left\{u'\left(c_1^{II}(R_H; R_H)\right) - R_H\left[(1-q)\,\mu_{2\alpha}^{II}(R_H) + q\mu_{2\beta}^{II}(R_H)\right]\right\} > 0.$$

Given that $\frac{dc_1^{II}(R_H;R_H)}{d\kappa} = -\frac{\frac{\pi}{1-\pi}(R_H)^{\frac{\gamma-1}{\gamma}}+1}{\pi} < 0$, then W^{II} is a strictly decreasing function of κ. Next, it can be derived that, for any $\kappa \in [\kappa_L, \kappa_M)$ and $\lambda = \frac{\kappa}{\kappa_M}$:

$$\frac{dW^{III}}{d\kappa} = \frac{(\psi_{HL}+\pi)^{\gamma}\kappa^{-\gamma}}{(\kappa_M)^{1-\gamma}} > 0.$$

Along scale thresholds $\kappa \in [\kappa_L, \kappa_M)$, we also derive the following properties of W^{II}, where $\lambda = 1$, and W^{III}, where $\lambda = \frac{\kappa}{\kappa_M}$:

$$\frac{dW^{II}}{dq} < 0, \quad \frac{dW^{II}}{d^2q} = 0, \quad \frac{dW^{III}}{dq} < 0, \quad \text{and} \quad \frac{dW^{III}}{d^2q} > 0.$$

Now, first suppose that $q \in (0, q_L^*]$, where $q_L^* \in \left(0, q_F^{II}\right]$, and $\kappa \in [\kappa_L, \kappa_M]$. By $\lim_{\kappa\to\kappa_L, q\to 0} W^{II} > \lim_{\kappa\to\kappa_L, q\to 0} W^{III}$, $\lim_{\kappa\to\kappa_M, q\to 0} W^{III} > \lim_{\kappa\to\kappa_M, q\to 0} W^{II}$, $\frac{dW^{II}}{d\kappa} < 0$, and $\frac{dW^{III}}{d\kappa} > 0$, there exists a unique cutoff minimum scale $\kappa_L^* \in (\kappa_L, \kappa_M)$ with which $\lim_{\kappa\to\kappa_L^*, q\to 0} W^{II} = \lim_{\kappa\to\kappa_L^*, q\to 0} W^{III}$. By the strict convexity of W^{II} and W^{III} with respect to q, $\lim_{q\to 0} W^{II} < \lim_{q\to 0} W^{III}$ $\forall \kappa \in (\kappa_L^*, \kappa_M]$, and $\lim_{\kappa\to\kappa_M, q\to q_F^{II}} \left(W^{II} - W^{III}\right) \lesseqgtr 0$, we then define a unique threshold probability $q_L^{**} \in (0, q_L^*]$ below which $W^{III} > W^{II}$ $\forall \kappa \in (\kappa_L^*, \kappa_M]$.

It is also worth emphasizing that if $\lim_{\kappa\to\kappa_M, q\to q_F^{II}} W^{II} \leq \lim_{\kappa\to\kappa_M, q\to q_F^{II}} W^{III}$, then $q_L^* = q_F^{II}$ and there exists a unique cutoff minimum scale $\kappa_L^{**} \in (\kappa_L, \kappa_M)$ above which $q_L^{**} = q_L^* = q_F^{II}$. If instead $\lim_{\kappa\to\kappa_M, q\to q_F^{II}} W^{II} > \lim_{\kappa\to\kappa_M, q\to q_F^{II}} W^{III}$, $q_L^* < q_F^{II}$ and q_L^{**} converges to q_L^* as κ goes to κ_M.

To illustrate these possibilities, we resort to numerical simulations in Figure 2 below. As a start, we construct examples with parameters $(\pi, \gamma, R_L, R_H) = (0.5, 8, 1.1, 1.15)$ under which $\lim_{\kappa\to\kappa_M, q\to q_F^{II}} W^{II} < \lim_{\kappa\to\kappa_M, q\to q_F^{II}} W^{III}$. As shown in the left panel, $\kappa_L = 0.2209$, $\kappa_L^* = 0.2247$, $\kappa_L^{**} = 0.2256$, $\kappa_M = 0.2257$, and $q_L^* = q_F^{II} = 5.83\%$. We further modify the parameter set to $(\pi, \gamma, R_L, R_H) = (0.5, 8, 1.1, 1.14)$ under which $\lim_{\kappa\to\kappa_M, q\to q_F^{II}} W^{II} > \lim_{\kappa\to\kappa_M, q\to q_F^{II}} W^{III}$. As depicted in the right panel, $\kappa_L = 0.2232$, $\kappa_L^* = 0.2257$, $\kappa_M = 0.2272$, $q_L^* = 11.13\%$, and $q_F^{II} = 11.61\%$.

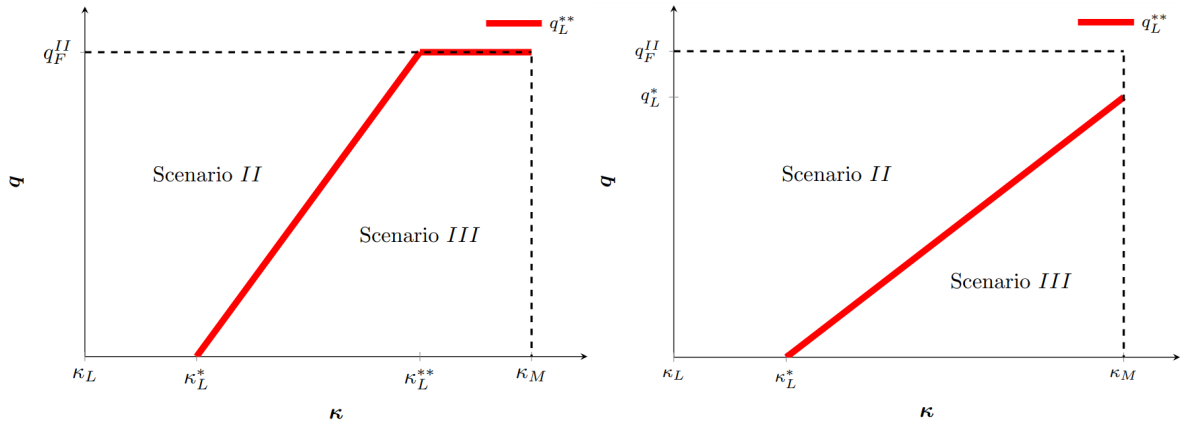

Figure 2: Possibilities with $q_L^* = q_F^{II}$ and $q_L^* < q_F^{II}$

To summarize equilibrium allocations over intermediate minimum scales more generally without relying on numerical examples above, we restrict to $q \in (0, q_L^*]$. Then, if $\kappa \in [\kappa_L, \kappa_L^*]$, allocation c^{II} is implemented and $\lambda = 1$. Next, if $\kappa \in (\kappa_L^*, \min\{\kappa_L^{**}, \kappa_M\}]$ and $q \in (q_L^{**}, q_L^*]$, allocation c^{II} is implemented and $\lambda = 1$ whereas if $\kappa \in (\kappa_L^*, \kappa_M]$ and $q \in (0, q_L^{**}]$, allocation c^{III} is implemented and $\lambda = \frac{\kappa}{\kappa_M}$. Lastly, if $\kappa \in (\kappa_M, \kappa_m)$, allocation c^{III} is implemented and $\lambda = 1$.

■

Proof of Lemma 3

To begin, suppose that $\kappa \in [\kappa_m, 1)$ such that allocations c^I and c^{III} are infeasible. Moreover, for allocation c^{II} where $\lambda = 1$, $c_1^{II}(R_H; R_H) \geq c_{2\beta}^{II}(R_H)$ if and only if $\kappa \leq \tilde{\kappa}_M$, where $\tilde{\kappa}_M > \kappa_m$ requires:

$$q < \underbrace{\frac{\left(\frac{(1-\pi)^2}{R_H + (1-\pi)(R_H)^{\frac{\gamma-1}{\gamma}}}\right)^{\gamma} - \phi_\alpha(R_H)}{\phi_\beta(R_L) - \phi_\alpha(R_H)}}_{<0}.$$

Yet, this inequality is violated for any $q > 0$ and thus $\tilde{\kappa}_M < \kappa_m$. In this vein, for any $\kappa > \kappa_m$, we omit allocation c^{II} and feasible allocations include c^{IV} and c^V only. Moreover, using the approach similar to that in the proof of Lemma 2, we can demonstrate that $\frac{dW^{IV}}{d\kappa} < 0$ for $\lambda = 1$, $\frac{dW^V}{d\kappa} > 0$ for $\kappa < \kappa_S = \frac{\psi_{LL}}{\psi_{LL}+\pi}$ and $\lambda = \frac{\kappa}{\kappa_S}$, and $\frac{dW^V}{d\kappa} = 0$ for $\kappa \geq \kappa_S$ and $\lambda = 1$.

To see which allocation will be on the equilibrium path of play, note that it can be shown that for $\lambda = 1$, $\lim_{\kappa \to \kappa_S} W^{IV} > \lim_{\kappa \to \kappa_S} W^V$ if $\psi_{LL} > \psi_{HL}$, which necessarily holds. Note also that the early payment $c_1^{IV}(R_H; R_L)$ would be below $c_{2\beta}^{IV}(R_L)$ if and only if $\kappa > \tilde{\kappa}_S = \frac{1}{\frac{\pi}{1-\pi}(R_L)^{\frac{1}{\gamma}}\left[\pi + (1-\pi)(R_L)^{\frac{1-\gamma}{\gamma}}\right]^{-1} + 1}$,

where $\tilde{\kappa}_S \geq \kappa_S$ if $q \leq q_F^{IV} = \frac{\frac{1}{R_L}\phi_\beta(R_L)-\phi_\alpha(R_L)}{\phi_\beta(R_L)-\phi_\alpha(R_L)}$. In particular, $q_F^{IV} > 0$ if $\gamma > 2$ and $\pi > \pi_F^{IV} = \frac{(R_L)^{\frac{2-\gamma}{\gamma}}-(R_L)^{\frac{1-\gamma}{\gamma}}}{1-(R_L)^{\frac{1-\gamma}{\gamma}}}$. In this vein, suppose that $\gamma > 2$, $\pi > \pi_F^{IV}$, and $q \leq q_F^{IV}$ such that $\tilde{\kappa}_S \geq \kappa_S$. Under these conditions, reducing $\lim_{\kappa \to \tilde{\kappa}_S} W^{IV} < \lim_{\kappa \to \tilde{\kappa}_S} W^V$ yields $R_H < R_H^{IV} = \left(\frac{(\psi_{LL}+\pi)^\gamma \left[\pi(R_L)^{\frac{1}{\gamma}} + \left(\phi_\beta(R_L)\right)^{\frac{1}{\gamma}}\right]^{1-\gamma} - \pi(R_L)^{\frac{1-\gamma}{\gamma}} - q\left(\phi_\beta(R_L)\right)^{\frac{1}{\gamma}}}{(1-\pi)^\gamma(1-q)\left(\phi_\beta(R_L)\right)^{\frac{1-\gamma}{\gamma}}} \right)^{\frac{1}{1-\gamma}}$. That is, for any $R_H \in \left(R_L, R_H^{IV}\right)$ and by $\lim_{\kappa \to \kappa_S} W^{IV} > \lim_{\kappa \to \kappa_S} W^V$, $\lim_{\kappa \to \tilde{\kappa}_S} W^{IV} < \lim_{\kappa \to \tilde{\kappa}_S} W^V$, $\frac{dW^{IV}}{d\kappa} < 0$ for $\lambda = 1$, $\frac{dW^V}{d\kappa} > 0 \; \forall \kappa < \kappa_S$ and $\lambda = \frac{\kappa}{\kappa_S}$, and $\frac{dW^V}{d\kappa} = 0 \; \forall \kappa \geq \kappa_S$ and $\lambda = 1$, there exists a cutoff $\kappa_S^* \in (\kappa_S, \tilde{\kappa}_S)$ below (above) which allocation c^{IV} (c^V) is obtained in equilibrium. Further, among these scenarios, it would be optimal for consumers to pool their full endowments.

■

Proof of Proposition 1

Suppose that $\kappa \in (0, \kappa_L)$ such that allocation c^I is obtained in equilibrium. To proceed,

$$\frac{d\frac{c_1^I(R_H;R_H)}{c_{2\beta}^I(R_H)}}{dq} = -\left(\frac{R_H}{\phi_\beta(R_H)}\right)^{-\frac{1}{\gamma}} (\psi_{HH})^{-2} \frac{d\psi_{HH}}{dq} < 0$$

whereas the ratio $\frac{c_1^I(R_H;R_H)}{c_{2\beta}^I(R_H)}$ is independent of κ. In this vein, reducing $c_1^I(R_H;R_H) \geq c_{2\beta}^I(R_H)$ in terms of q yields: $q \leq q_F^I$, where $q_F^I > 0$ if and only if $\gamma > 2$ and $\pi > \pi_F^I$. Next, $W^I \geq W^A$ if $q \leq q_A^I$, where $q_F^I \lesseqgtr q_A^I$ if $R_L \lesseqgtr R_L^I = \left(\frac{\left[(R_H)^{-\frac{1}{\gamma}}\left(\phi_\beta(R_H)\right)^{\frac{1}{\gamma}} + \pi\right]^\gamma - \pi}{1-\pi} \right)^{\frac{1}{1-\gamma}}$. Note that if the numerator in the parentheses denoted $\xi_0^I \equiv \left[(R_H)^{-\frac{1}{\gamma}}\left(\phi_\beta(R_H)\right)^{\frac{1}{\gamma}} + \pi\right]^\gamma - \pi$ is positive, then $R_L^I > 1$ necessarily holds. Note also that $\xi_0^I > 0$ if and only if:

$$\underbrace{(1-\pi)\left[\pi + (1-\pi)(R_H)^{\frac{1-\gamma}{\gamma}}\right](R_H)^{-\frac{1}{\gamma}}}_{\equiv \xi_1^I} > \pi^{\frac{1}{\gamma}} - \pi,$$

where $\lim_{R_H \to 1} \xi_1^I = 1 - \pi > \pi^{\frac{1}{\gamma}} - \pi$, $\lim_{R_H \to \infty} \xi_1^I = 0 < \pi^{\frac{1}{\gamma}} - \pi$, $\frac{d\xi_1^I}{dR_H} < 0$, and $\frac{d\xi_1^I}{d^2 R_H} > 0$. Alternatively, there exists a unique threshold $\underline{R}_H^I > 1$ below (above) which $\xi_0^I > 0$ and thus $R_L^I > 1$ (R_L^I is undefined).

In what follows, we show that if $\pi \in \left(\pi_F^I, \frac{1}{\gamma-1}\right]$, there exists a cutoff high asset return $R_H^I > R_L$ below (above) which $R_L^I < R_H$ ($R_L^I > R_H$). To this end, note that $R_L^I < R_H$ if and only if

$\xi_1^I > \xi_2^I \equiv \left[\pi + (1-\pi)\left(R_H\right)^{1-\gamma}\right]^{\frac{1}{\gamma}} - \pi$. It would also be helpful to see some properties:

$$\lim_{R_H \to 1} \xi_1^I = \lim_{R_H \to 1} \xi_2^I = 1 - \pi, \quad \lim_{R_H \to \infty} \xi_1^I = 0, \quad \text{and} \quad \lim_{R_H \to \infty} \xi_2^I = \left(\pi^{\frac{1-\gamma}{\gamma}} - 1\right)\pi > 0.$$

Moreover,

$$\frac{d\xi_2^I}{dR_H} < 0 \quad \text{and} \quad \frac{d\xi_2^I}{d^2 R_H} > 0.$$

That is, ξ_1^I and ξ_2^I are decreasing and strictly concave functions of R_H. To proceed,

$$\left.\frac{d\xi_1^I}{dR_H}\right|_{R_H=1} = \left(\left(\frac{1-\gamma}{\gamma}\right)(1-\pi) - \frac{1}{\gamma}\right)(1-\pi) \quad \text{and} \quad \left.\frac{d\xi_2^I}{dR_H}\right|_{R_H=1} = \left(\frac{1-\gamma}{\gamma}\right)(1-\pi).$$

In particular, $\left.\frac{d\xi_1^I}{dR_H}\right|_{R_H=1} > \left.\frac{d\xi_2^I}{dR_H}\right|_{R_H=1}$ if and only if $\pi > \frac{1}{\gamma-1}$. Furthermore, to see the relationship between $\frac{1}{\gamma-1}$ and π_F^I, we define:

$$\Delta\pi^I \equiv \frac{1}{\gamma-1} - \frac{\left(R_H\right)^{\frac{2-\gamma}{\gamma}} - \left(R_H\right)^{\frac{1-\gamma}{\gamma}}}{1 - \left(R_H\right)^{\frac{1-\gamma}{\gamma}}} = \frac{\xi_3^I}{(\gamma-1)\left[\left(R_H\right)^{\frac{\gamma-1}{\gamma}} - 1\right]},$$

where $\xi_3^I \equiv \left[\left(R_H\right)^{\frac{\gamma-1}{\gamma}} - 1\right] - (\gamma-1)\left[\left(R_H\right)^{\frac{1}{\gamma}} - 1\right]$ and the sign of $\Delta\pi^I$ depends on that of ξ_3^I. To proceed, differentiating ξ_3^I with respect to R_H yields:

$$\frac{d\xi_3^I}{dR_H} = \left(\frac{\gamma-1}{\gamma}\right)\left[1 - \left(R_H\right)^{\frac{2-\gamma}{\gamma}}\right]\left(R_H\right)^{-\frac{1}{\gamma}} > 0,$$

for any $\gamma > 2$ and $R_H > 1$. Thus, suppose that $\gamma \in (2, \infty)$. Under this condition, by $\lim_{R_H \to 1} \xi_3^I = 0$ and $\frac{d\xi_3^I}{dR_H} > 0$, $\Delta\pi^I > 0$ and hence $\frac{1}{\gamma-1} > \pi_F^I$.

To summarize, first suppose that $R_H \in \left(1, \underline{R}_H^I\right)$. Further suppose that $\pi \in \left(\pi_F^I, \frac{1}{\gamma-1}\right]$. Under these conditions, $R_L < R_H < R_L^I$, $q_F^I < q_A^I$, and run equilibria exist $\forall q \in \left(0, q_F^I\right]$. Suppose instead that $\pi \in \left(\frac{1}{\gamma-1}, 1\right)$. Under these conditions, there exists a cutoff $R_H^I > R_L$ such that: (a) if $R_L \in \left(1, R_L^I\right)$ and $R_H \in \left(R_L, \min\left\{\underline{R}_H^I, R_H^I\right\}\right)$, then $q_F^I < q_A^I$ and run equilibria exist $\forall q \in \left(0, q_F^I\right]$; (b) if $R_L \in \left(R_L^I, R_H\right)$ and $R_H \in \left(R_L, \min\left\{\underline{R}_H^I, R_H^I\right\}\right)$, then $q_A^I < q_F^I$ and run equilibria exist $\forall q \in \left(0, q_A^I\right]$; (c) if $R_H \in \left(R_H^I, \underline{R}_H^I\right)$, then $R_L < R_H < R_L^I$, $q_F^I < q_A^I$, and run equilibria exist $\forall q \in \left(0, q_F^I\right]$.

■

Proof of Proposition 2

Suppose that $\kappa \in [\kappa_L, \kappa_L^*]$ ($\kappa \in (\kappa_L^*, \min\{\kappa_L^{**}, \kappa_M\}]$), $\gamma \in (2,\infty)$, $R_L \in \left((R_H)^{\frac{1}{\gamma-1}}, R_H\right)$, $\pi \in \left(\pi_F^{II}, 1\right)$, and $q \in (0, q_L^*]$ ($q \in (q_L^{**}, q_L^*]$) such that allocation c^{II} is obtained in equilibrium. Further note that:

$$\frac{d\frac{c_1^{II}(R_H;R_H)}{c_{2\beta}^{II}(R_H)}}{d\kappa} = -\frac{(1-\pi)^2\left(\left[\frac{\pi}{1-\pi}(R_H)^{\frac{\gamma-1}{\gamma}}+1\right]\kappa + \left(1-\left[\frac{\pi}{1-\pi}(R_H)^{\frac{\gamma-1}{\gamma}}+1\right]\kappa\right)\right)\pi R_H}{(\pi R_H \kappa)^2} < 0,$$

while the ratio $\frac{c_1^{II}(R_H;R_H)}{c_{2\beta}^{II}(R_H)}$ is independent of q. In this vein, reducing $c_1^{II}(R_H;R_H) \geq c_{2\beta}^{II}(R_H)$ in terms of κ yields: $\kappa \leq \tilde{\kappa}_M$ in which $\tilde{\kappa}_M \geq \kappa_M$ if $q \leq q_F^{II}$, where q_F^{II} is defined in the proof of Lemma 2. Also recall that if $\gamma \in (2,\infty)$ and $R_L \in \left((R_H)^{\frac{1}{\gamma-1}}, R_H\right)$, $\pi_F^{II} < 1$ and $q_F^{II} < q_F^I$ if $\pi \in \left(\pi_F^{II}, 1\right)$, where $\pi_F^{II} \in \left(\pi_F^I, 1\right)$. As a result, $c_1^{II}(R_H;R_H) \geq c_{2\beta}^{II}(R_H)$ must hold for any $q \in (0, q_L^*]$. In addition, $W^{II} \geq W^A$ if and only if $q \leq q_A^{II}$.

To summarize, suppose that $\kappa \in [\kappa_L, \kappa_L^*]$ ($\kappa \in (\kappa_L^*, \min\{\kappa_L^{**}, \kappa_M\}]$), $\gamma \in (2,\infty)$, $R_L \in \left((R_H)^{\frac{1}{\gamma-1}}, R_H\right)$, $\pi \in \left(\pi_F^{II}, 1\right)$. Under these conditions, run equilibria exist $\forall q \in \left(0, \min\left\{q_L^*, q_A^{II}\right\}\right]$ ($\forall q \in \left(q_L^{**}, \min\left\{q_L^*, q_A^{II}\right\}\right]$).

■

Proof of Proposition 3

Suppose that $\kappa \in (\kappa_L^*, \kappa_M]$ ($\kappa \in (\kappa_M, \kappa_m]$), $\gamma \in (2,\infty)$, $R_L \in \left((R_H)^{\frac{1}{\gamma-1}}, R_H\right)$, $\pi \in \left(\pi_F^{II}, 1\right)$, and $q \in (0, q_L^{**}]$ ($q \in (0, q_L^*]$) such that allocation c^{III} is on the equilibrium path of play. Further,

$$\frac{d\frac{c_1^{III}(R_H;R_L)}{c_{2\beta}^{III}(R_L)}}{dq} = -\left(\frac{R_L}{\phi_\beta(R_L)}\right)^{-\frac{1}{\gamma}}(\psi_{HL})^{-2}\frac{d\psi_{HL}}{dq} < 0,$$

whereas the ratio $\frac{c_1^{III}(R_H;R_L)}{c_{2\beta}^{III}(R_L)}$ is independent of κ. In this manner, reducing $c_1^{III}(R_H;R_L) \geq c_{2\beta}^{III}(R_L)$ in terms of q yields: $q \leq q_F^{III}$, where $q_F^{III} > 0$ if $\pi > \pi_F^{III}$. Further, it can easily be verified that if $\gamma \in (2,\infty)$ and $R_L \in \left((R_H)^{\frac{1}{\gamma-1}}, R_H\right)$, $\pi_F^{III} < \pi_F^I < \pi_F^{II} < 1$, implying that $0 < q_L^{**} \leq q_L^* \leq q_F^{II} < q_F^I < q_F^{III} < 1\ \forall \pi \in \left(\pi_F^{II}, 1\right)$. Next, $W^{III} \geq W^A$ if and only if $q < q_A^{III}$.

In what follows, we examine whether run equilibria in an active bank exists. As a starting point, it is useful to see that:

$$\lim_{R_L \to R_H \to 1} W^{III}\left(q_F^{III}\right) = \lim_{R_L \to 1} W^A = \frac{1}{1-\gamma},\ \lim_{R_L \to R_H \to \infty} W^{III}\left(q_F^{III}\right) = \frac{\pi^\gamma}{1-\gamma}, \text{ and } \lim_{R_L \to \infty} W^A = \frac{\pi}{1-\gamma},$$

$\lim_{R_L \to R_H \to \infty} W^{III}\left(q_F^{III}\right) > \lim_{R_L \to R_H \to \infty} W^A$. Moreover, it is useful to derive the following properties

of $W^{III}\left(q_F^{III}\right)$ and W^A:

$$\frac{dW^{III}\left(q_F^{III}\right)}{dR_L} > 0, \quad \frac{dW^{III}\left(q_F^{III}\right)}{d^2R_L} < 0, \quad \frac{dW^A}{dR_L} > 0, \quad \text{and} \quad \frac{dW^A}{d^2R_L} < 0.$$

That is, $W^{III}\left(q_F^{III}\right)$ and W^A are increasing and strictly concave functions of R_L. In addition,

$$\left.\frac{dW^{III}\left(q_F^{III}\right)}{dR_L}\right|_{R_L=R_H=1} = \left(\frac{1-\pi}{\gamma-1}\right)\left[\gamma\left(1-\pi\right)+\pi\right] \quad \text{and} \quad \left.\frac{dW^A}{dR_L}\right|_{R_L=1} = 1-\pi.$$

In this vein, $\left.\frac{dW^{III}\left(q_F^{III}\right)}{dR_L}\right|_{R_L=1} \gtreqless \left.\frac{dW^A}{dR_L}\right|_{R_L=1}$ if and only if $\pi \lesseqgtr \frac{1}{\gamma-1}$. Further, it can be shown that $\pi_F^{II} \lesseqgtr \frac{1}{\gamma-1}$ if $R_L \gtreqless R_L^{II} = \left(\frac{(R_H)^{\frac{\gamma-1}{\gamma}}+(\gamma-2)}{R_H(\gamma-1)}\right)^{\frac{\gamma}{1-\gamma}}$, where $R_L^{II} > (R_H)^{\frac{1}{\gamma-1}}$ necessarily holds. To show that $R_L^{II} < R_H$, the proof below is by contradiction. Suppose that $R_L^{II} > R_H$ such that $R_L < R_L^{II}$ necessarily holds, implying that there is no room for the condition $\pi_F^{II} < \frac{1}{\gamma-1}$ to hold. However, this contradicts the results that $\lim_{R_L \to R_H} \pi_F^{II} = \pi_F^{I} < \frac{1}{\gamma-1}$, as demonstrated in the proof of Proposition 1. Consequently, we must have $(R_H)^{\frac{1}{\gamma-1}} < R_L^{II} < R_H$.

To summarize, suppose that $\gamma \in (2,\infty)$, $R_L \in \left((R_H)^{\frac{1}{\gamma-1}}, R_H\right)$, and $\pi \in \left(\pi_F^{II}, 1\right)$. Further suppose that $R_L \in \left((R_H)^{\frac{1}{\gamma-1}}, R_L^{II}\right)$ such that $\pi_F^{II} > \frac{1}{\gamma-1}$. Under these conditions, by $\left.\frac{dW^{III}\left(q_F^{III}\right)}{dR_L}\right|_{R_L=R_H=1} < \left.\frac{dW^A}{dR_L}\right|_{R_L=1}$ and $\lim_{R_L \to R_H \to \infty} W^{III}\left(q_F^{III}\right) > \lim_{R_L \to R_H \to \infty} W^A$, there exists a threshold R_L^{III} such that: (a) if $R_L \in \left((R_H)^{\frac{1}{\gamma-1}}, \min\left\{R_L^{II}, R_L^{III}\right\}\right)$ and $\kappa \in (\kappa_L^*, \kappa_M]$ or $\kappa \in (\kappa_M, \kappa_m)$, then $W^{III}\left(q_F^{III}\right) < W^A$, suggesting that $q_A^{III} < q_F^{III}$ and thus run equilibria exist $\forall q \in \left(0, \min\left\{q_L^{**}, q_A^{III}\right\}\right]$ or $\forall q \in \left(0, \min\left\{q_L^{*}, q_A^{III}\right\}\right]$; (b) if $R_L \in \left(\max\left\{(R_H)^{\frac{1}{\gamma-1}}, R_L^{III}\right\}, R_L^{II}\right)$ and $\kappa \in (\kappa_L^*, \kappa_M]$ or $\kappa \in (\kappa_M, \kappa_m)$, then $W^{III}\left(q_F^{III}\right) > W^A$, implying that $q_L^* < q_F^{III} < q_A^{III}$ and therefore run equilibria exist $\forall q \in (0, q_L^{**}]$ or $\forall q \in (0, q_L^*]$.

Suppose instead that $R_L \in \left(R_L^{II}, R_H\right)$ such that $\pi_F^{II} < \frac{1}{\gamma-1}$. Then, if $\pi \in \left(\pi_F^{II}, \frac{1}{\gamma-1}\right)$ and $\kappa \in (\kappa_L^*, \kappa_M]$ or $\kappa \in (\kappa_M, \kappa_m)$, $W^{III}\left(q_F^{III}\right) > W^A$, implying that $q_F^{III} < q_A^{III}$ and therefore run equilibria exist $\forall q \in (0, q_L^{**}]$ or $\forall q \in (0, q_L^*]$. Yet, if $\pi \in \left[\frac{1}{\gamma-1}, 1\right)$, there exists a threshold R_L^{III} such that: (a) if $R_L \in \left(R_L^{II}, \min\left\{R_L^{III}, R_H\right\}\right)$ and $\kappa \in (\kappa_L^*, \kappa_M]$ or $\kappa \in (\kappa_M, \kappa_m)$, then $W^{III}\left(q_F^{III}\right) < W^A$, suggesting that $q_A^{III} < q_F^{III}$ and hence run equilibria exist $\forall q \in \left(0, \min\left\{q_L^{**}, q_A^{III}\right\}\right]$ or $\forall q \in \left(0, \min\left\{q_L^{*}, q_A^{III}\right\}\right]$; (b) if $R_L \in \left(\max\left\{R_L^{II}, R_L^{III}\right\}, R_H\right)$ and $\kappa \in (\kappa_L^*, \kappa_M]$ or $\kappa \in (\kappa_M, \kappa_m)$, then $W^{III}\left(q_F^{III}\right) > W^A$, implying that $q_F^{III} < q_A^{III}$ and thus run equilibria exist $\forall q \in (0, q_L^{**}]$ or $\forall q \in (0, q_L^*]$.

■

Proof of Proposition 4

To begin, suppose that $\kappa \in [\kappa_m, \kappa_S^*]$, $\gamma \in (2, \infty)$, $\pi \in \left(\pi_F^{IV}, 1\right)$, $q \in \left(0, q_F^{IV}\right]$, and $R_H \in \left(R_L, R_H^{IV}\right)$ such that allocation c^{IV} is implemented in equilibrium. Further note that:

$$\frac{d\frac{c_1^{IV}(R_H;R_L)}{c_{2\beta}^{IV}(R_L)}}{d\kappa} = -\frac{(1-\pi)\left[\pi + (1-\pi)(R_L)^{\frac{1-\gamma}{\gamma}}\right]}{\pi (R_L)^{\frac{1}{\gamma}} \kappa^2} < 0$$

but the ratio $\frac{c_1^{IV}(R_H;R_L)}{c_{2\beta}^{IV}(R_L)}$ is independent of q. In this vein, reducing $c_1^{IV}(R_H;R_L) \geq c_{2\beta}^{IV}(R_L)$ in terms of κ yields: $\kappa \leq \frac{1}{\frac{\pi}{1-\pi}(R_L)^{\frac{1}{\gamma}}\left[\pi+(1-\pi)(R_L)^{\frac{1-\gamma}{\gamma}}\right]^{-1}+1} = \tilde{\kappa}_S$ in which $\tilde{\kappa}_S \geq \kappa_S$ if $q \leq q_F^{IV}$, where q_F^{IV} is defined in the proof of Lemma 3. Also recall that if $\gamma \in (2, \infty)$ and $\pi \in \left(\pi_F^{IV}, 1\right)$, $q_F^{IV} \in (0, 1)$ and thus the fragility condition holds for any $q \in \left(0, q_F^{IV}\right]$. In addition, $W^{IV} \geq W^A$ if and only if $q \leq q_A^{IV}$.

To summarize, suppose that $\kappa \in [\kappa_m, \kappa_S^*]$, $\gamma \in (2, \infty)$, $\pi \in \left(\pi_F^{IV}, 1\right)$, $q \in \left(0, q_F^{IV}\right]$, and $R_H \in \left(R_L, R_H^{IV}\right)$. Under these conditions, run equilibria exist $\forall q \in \left(0, \min\left\{q_F^{IV}, q_A^{IV}\right\}\right]$.

■

Proof of Proposition 5

Suppose that $\kappa \in (\kappa_S^*, 1)$, $\gamma \in (2, \infty)$, $\pi \in \left(\pi_F^{IV}, 1\right)$, $q \in \left(0, q_F^{IV}\right]$, and $R_H \in \left(R_L, R_H^{IV}\right)$ such that allocation c^V is obtained in equilibrium. Note also that:

$$\frac{d\frac{c_1^V(R_L;R_L)}{c_{2\beta}^V(R_L)}}{dq} = -\left(\frac{R_L}{\phi_\beta(R_L)}\right)^{-\frac{1}{\gamma}} (\psi_{LL})^{-2} \frac{d\psi_{LL}}{dq} < 0.$$

whereas the ratio $\frac{c_1^V(R_L;R_L)}{c_{2\beta}^V(R_L)}$ is independent of κ. Thus, reducing $c_1^V(R_L;R_L) \geq c_{2\beta}^V(R_L)$ if and only if $q < q_F^{IV}$, where $q_F^{IV} > 0$ if $\gamma > 2$ and $\pi > \pi_F^{IV}$. To proceed, $W^V \geq W^A$ if $q < q_A^V$.

In what follows, we examine whether run equilibria in an active bank exist. To this end, it is useful to see that:

$$\lim_{R_L \to 1} W^V\left(q_F^{IV}\right) = \lim_{R_L \to 1} W^A = \frac{1}{1-\gamma} \quad \text{and} \quad \lim_{R_L \to R_H \to \infty} W^V\left(q_F^{IV}\right) = \frac{\pi^\gamma}{1-\gamma}.$$

Therefore, it can be shown that $\lim_{R_L \to R_H \to \infty} W^A < \lim_{R_L \to R_H \to \infty} W^V\left(q_F^{IV}\right)$. Further, it is helpful to derive the following properties of $W^V\left(q_F^{IV}\right)$:

$$\frac{dW^V(q_F)}{dR_L} > 0 \quad \text{and} \quad \frac{dW^V(q_F)}{d^2R_L} < 0.$$

Thus, both W^A and $W^V\left(q_F^{IV}\right)$ are strictly increasing and concave functions of R_L. It can also be verified that $\left.\frac{dW^V\left(q_F^{IV}\right)}{dR_L}\right|_{R_L=1} \gtreqless \left.\frac{dW^A}{dR_L}\right|_{R_L=1}$ if $\pi \lesseqgtr \frac{1}{\gamma-1}$ and we next demonstrate that $\frac{1}{\gamma-1} > \pi_F^{IV}$ holds for any $\gamma > 2$ and $R_L > 1$. To be specific, we define:

$$\Delta\pi^V \equiv \frac{1}{\gamma-1} - \frac{(R_L)^{\frac{2-\gamma}{\gamma}} - (R_L)^{\frac{1-\gamma}{\gamma}}}{1-(R_L)^{\frac{1-\gamma}{\gamma}}} = \frac{\xi_1^V}{(\gamma-1)\left[(R_L)^{\frac{\gamma-1}{\gamma}}-1\right]},$$

where $\xi_1^V \equiv \left[(R_L)^{\frac{\gamma-1}{\gamma}}-1\right] - (\gamma-1)\left[(R_L)^{\frac{1}{\gamma}}-1\right]$. That is, the sign of $\Delta\pi^V$ depends on the sign of ξ_1^V. To proceed, differentiating ξ_1^V with respect to R_L yields:

$$\frac{d\xi_1^V}{dR_L} = \left(\frac{\gamma-1}{\gamma}\right)\left[1-(R_L)^{\frac{2-\gamma}{\gamma}}\right](R_L)^{-\frac{1}{\gamma}} > 0,$$

for any $\gamma > 2$ and $R_L > 1$. By $\lim_{R_L \to 1} \xi_1^V = 0$ and $\frac{d\xi_1^V}{dR_L} > 0$, $\Delta\pi^V > 0$ and hence $\frac{1}{\gamma-1} > \pi_F^{IV}$.

To summarize, suppose that $\kappa \in (\kappa_S^*, 1)$, $\gamma \in (2, \infty)$, and $R_H \in \left(R_L, R_H^{IV}\right)$. Further suppose that $\pi \in \left(\pi_F^{IV}, \frac{1}{\gamma-1}\right]$. Under these conditions, $W^V\left(q_F^{IV}\right) > W^A$, implying that $q_F^{IV} < q_A^V$ for any $R_L > 1$ and thus run equilibria exist $\forall q \in \left(0, q_F^{IV}\right]$. Suppose, however, that $\pi \in \left(\frac{1}{\gamma-1}, 1\right)$. Under these conditions, there exists a threshold R_L^V such that: (a) if $R_L \in \left(1, \min\left\{R_L^V, R_H\right\}\right)$, $W^V\left(q_F^{IV}\right) < W^A$, suggesting that $q_A^V < q_F^{IV}$ and hence run equilibria exist $\forall q \in \left(0, q_A^V\right]$; (b) if $R_L \in \left(R_L^V, R_H\right)$, $W^V\left(q_F^{IV}\right) > W^A$, implying that $q_F^{IV} < q_A^V$ and run equilibria exist $\forall q \in \left(0, q_F^{IV}\right]$.

Given the existence of run equilibria across different minimum scales, we next summarize the

sum of all consumers' expected utilities denoted by $\mathcal{W}^*$ for $* = I, II, ..., V$ as follows:

$$
\begin{cases}
\mathcal{W}^{I} = \frac{(\psi_{HH}+\pi)^{\gamma}}{1-\gamma} & \text{if } \kappa \in (0, \kappa_L),\ \gamma \in (2, \infty), \\
 & \pi \in \left(\pi_F^I, 1\right), \text{and } q \in \left(0, \min\left\{q_F^I, q_A^I\right\}\right]; \\
\mathcal{W}^{II} = \frac{\pi^{\gamma}\left(1-\left[\frac{\pi}{1-\pi}(R_H)^{\frac{\gamma-1}{\gamma}}+1\right]\kappa\right)^{1-\gamma}+(\psi_{HH})^{\gamma}\left(\left[\frac{\pi}{1-\pi}(R_H)^{\frac{\gamma-1}{\gamma}}+1\right]\kappa\right)^{1-\gamma}}{1-\gamma} & \text{if } \kappa \in [\kappa_L, \kappa_L^*]\ (\kappa \in (\kappa_L^*, \min\{\kappa_L^{**}, \kappa_M\}]), \\
 & \gamma \in (2, \infty),\ R_L \in \left((R_H)^{\frac{1}{\gamma-1}}, R_H\right), \\
 & \pi \in \left(\pi_F^{II}, 1\right), \text{and } q \in \left(0, \min\left\{q_L^*, q_A^{II}\right\}\right] \\
 & (q \in \left(q_L^{**}, \min\left\{q_L^*, q_A^{II}\right\}\right]); \\
\mathcal{W}^{III} = \begin{array}{l} \frac{\pi\left[\frac{\kappa}{\kappa_M}\left(\frac{1}{\psi_{HL}+\pi}\right)+\left(1-\frac{\kappa}{\kappa_M}\right)\right]^{1-\gamma}}{1-\gamma} \\ +\frac{(1-q)(1-\pi)\left[\frac{\kappa}{\kappa_M}\left(\frac{R_H}{\phi_\alpha(R_H)}\right)^{\frac{1}{\gamma}}\left(\frac{\psi_{HL}}{\psi_{HL}+\pi}\right)+R_L\left(1-\frac{\kappa}{\kappa_M}\right)\right]^{1-\gamma}}{1-\gamma} \\ +\frac{q(1-\pi)\pi\left[\frac{\kappa}{\kappa_M}\left(\frac{1}{\phi_\beta(R_L)}\right)^{\frac{1}{\gamma}}\left(\frac{\psi_{HL}}{\psi_{HL}+\pi}\right)+\left(1-\frac{\kappa}{\kappa_M}\right)\right]^{1-\gamma}}{1-\gamma} \\ +\frac{q(1-\pi)^2\left[\frac{\kappa}{\kappa_M}\left(\frac{R_L}{\phi_\beta(R_L)}\right)^{\frac{1}{\gamma}}\left(\frac{\psi_{HL}}{\psi_{HL}+\pi}\right)+R_L\left(1-\frac{\kappa}{\kappa_M}\right)\right]^{1-\gamma}}{1-\gamma} \end{array} & \text{if } \kappa \in (\kappa_L^*, \kappa_M],\ \gamma \in (2, \infty), \\
 & R_L \in \left((R_H)^{\frac{1}{\gamma-1}}, R_H\right), \pi \in \left(\pi_F^{II}, 1\right), \\
 & \text{and } q \in \left(0, \min\left\{q_L^{**}, q_A^{III}\right\}\right]; \\
\mathcal{W}^{III} = \frac{(\psi_{HL}+\pi)^{\gamma}}{1-\gamma} & \text{if } \kappa \in (\kappa_M, \kappa_m],\ \gamma \in (2, \infty), \\
 & R_L \in \left((R_H)^{\frac{1}{\gamma-1}}, R_H\right), \pi \in \left(\pi_F^{II}, 1\right), \\
 & \text{and } q \in \left(0, \min\left\{q_L^*, q_A^{III}\right\}\right]; \\
\mathcal{W}^{IV} = \frac{\pi^{\gamma}(1-\kappa)^{1-\gamma}+(\psi_{HL})^{\gamma}\kappa^{1-\gamma}}{1-\gamma} & \text{if } \kappa \in [\kappa_m, \kappa_S^*],\ \gamma \in (2, \infty), \\
 & R_H \in \left(R_L, R_H^{IV}\right), \pi \in \left(\pi_F^{IV}, 1\right), \\
 & \text{and } q \in \left(0, \min\left\{q_F^{IV}, q_A^{IV}\right\}\right]; \\
\mathcal{W}^{V} = \frac{(\psi_{LL}+\pi)^{\gamma}}{1-\gamma} & \text{if } \kappa \in (\kappa_S^*, 1),\ \gamma \in (2, \infty), \\
 & R_H \in \left(R_L, R_H^{IV}\right), \pi \in \left(\pi_F^{IV}, 1\right), \\
 & \text{and } q \in \left(0, \min\left\{q_F^{IV}, q_A^{V}\right\}\right].
\end{cases}
$$

■

Proof of Proposition 6

Suppose that the bank only has access to a constant-returns-to-scale technology. Under this condition, $R_A = R_L = R_H$ and the best-response allocation is equivalent to the allocation vector c^V. To succinctly summarize the existence of run equilibria, note that the results are the same to

those in the proof of Proposition 5 with the exception that restrictions on R_H and κ are not needed as c^V is the unique optimal allocation.

In this regard, as we argued in the main text and given that consumers will pool their full endowments, i.e., $\lambda = 1$, the relative bank size or investment can be gauged by the inverse of the minimum scale, $\frac{1}{\kappa}$. Yet, as $\frac{d\frac{c_1^V(R_L;R_L)}{c_{2\beta}^V(R_L)}}{d\kappa} = 0$, the degree of instability is then independent of the size of the bank's deposit base. To proceed, it is trivial to show that $W^V = \mathcal{W}^V = \frac{\lambda^{1-\gamma}(\psi_{LL}+\pi)^\gamma}{1-\gamma}$, implying that changes in the amount of banks' initial investment alter the sum of depositors' expected utilities proportionally.

■

Proof of Proposition 7

i. As before, to compare depositors' incentives to panic across minimum scales, we begin by characterizing under what conditions run equilibria across allocations of interest exist. Notably, it follows from proofs of Propositions 1 to 5 that if $\gamma \in (2,\infty)$, $\pi \in \left(\pi_F^{II}, 1\right)$, $R_H \in \left(R_L, (R_L)^{\gamma-1}\right)$ ($R_H \in \left(R_L, \min\left\{(R_L)^{\gamma-1}, R_H^{IV}\right\}\right)$), and $q \in \left(0, \min\left\{q_L^*, q_A^{II}\right\}\right]$ or $q \in \left(q_L^{**}, \min\left\{q_L^*, q_A^{II}\right\}\right]$ ($q \in \left(0, \min\left\{q_L^{**}, q_A^{IV}\right\}\right]$ or $q \in \left(0, \min\left\{q_L^*, q_A^{IV}\right\}\right]$), run equilibria exist under allocations c^I and c^{II} (c^{III} and c^{IV}).

To proceed, note that to implement allocation c^{II} while allocation c^I is not feasible as the scale threshold κ is higher, the bank needs to reduce the early payment made to depositors so that the bank still qualifies for the higher asset return in both sunspot states. In this manner, $c_1^{II}(R_H;R_H) < c_1^I(R_H;R_H)$, implying that $c_{2\beta}^{II}(R_H) > c_{2\beta}^I(R_H)$. Consequently, we have $\frac{c_1^I(R_H;R_H)}{c_{2\beta}^I(R_H)} > \frac{c_1^{II}(R_H;R_H)}{c_{2\beta}^{II}(R_H)} \geq 1$. Similar reasoning applies to the comparison between $c_1^{III}(R_H;R_L)$ and $c_1^{IV}(R_H;R_L)$. That is, $c_1^{IV}(R_H;R_L) < c_1^{III}(R_H;R_L)$, implying that $c_{2\beta}^{IV}(R_L) > c_{2\beta}^{III}(R_L)$. In this vein, we have $\frac{c_1^{III}(R_H;R_L)}{c_{2\beta}^{III}(R_L)} > \frac{c_1^{IV}(R_H;R_L)}{c_{2\beta}^{IV}(R_L)} \geq 1$.

ii. It can be shown that if $\gamma \in (2,\infty)$, $\pi \in \left(\pi_F^{II}, 1\right)$, $R_H \in \left(R_L, \min\left\{(R_L)^{\gamma-1}, R_H^{IV}\right\}\right)$, and $q \in \left(0, \min\left\{q_L^{**}, q_A^V\right\}\right]$ or $q \in \left(0, \min\left\{q_L^*, q_A^V\right\}\right]$, run equilibria exist across all allocations c^I to c^V.

Next, to establish $\frac{c_1^I(R_H;R_H)}{c_{2\beta}^I(R_H)} < \frac{c_1^{III}(R_H;R_L)}{c_{2\beta}^{III}(R_L)}$, note that $\frac{\frac{c_1^{III}(R_H;R_L)}{c_{2\beta}^{III}(R_L)}}{\frac{c_1^I(R_H;R_H)}{c_{2\beta}^I(R_H)}} = \left(\frac{R_H}{R_L}\frac{\phi_\beta(R_L)}{\phi_\beta(R_H)}\right)^{\frac{1}{\gamma}} \frac{\psi_{HH}}{\psi_{HL}}$ and differentiating $\frac{\psi_{HH}}{\psi_{HL}}$ with respect to q yields:

$$\frac{d\frac{\psi_{HH}}{\psi_{HL}}}{dq} = \frac{1}{\gamma}\left(\frac{(1-q)\,\phi_\alpha(R_H) + q\phi_\beta(R_H)}{(1-q)\,\phi_\alpha(R_H) + q\phi_\beta(R_L)}\right)^{\frac{1-\gamma}{\gamma}} \frac{\phi_\alpha(R_H)\left[\phi_\beta(R_H) - \phi_\beta(R_L)\right]}{\left[(1-q)\,\phi_\alpha(R_H) + q\phi_\beta(R_L)\right]^2} < 0.$$

Thus, as q goes to 1, $\frac{\frac{c_1^{III}(R_H;R_L)}{c_{2\beta}^{III}(R_L)}}{\frac{c_1^{I}(R_H;R_H)}{c_{2\beta}^{I}(R_H)}}$ attains the minimum value where $\lim_{q\to 1}\frac{\frac{c_1^{III}(R_H;R_L)}{c_{2\beta}^{III}(R_L)}}{\frac{c_1^{I}(R_H;R_H)}{c_{2\beta}^{I}(R_H)}} = \left(\frac{R_H}{R_L}\right)^{\frac{1}{\gamma}} > 1$, which implies that there exist economies for which $c_1^I(R_H;R_H)$ is slightly lower than $c_{2\beta}^I(R_H)$, but $c_1^{III}(R_H;R_L) \geq c_{2\beta}^{III}(R_L)$ still holds. In addition, it can be easily shown that $\frac{\frac{c_1^{III}(R_H;R_L)}{c_{2\beta}^{III}(R_L)}}{\frac{c_1^{V}(R_L;R_L)}{c_{2\beta}^{V}(R_L)}} = \frac{\psi_{LL}}{\psi_{HL}} > 1$. Combining the results in this case and those in case i yields $\frac{c_1^{III}(R_H;R_L)}{c_{2\beta}^{III}(R_L)} > \frac{c_1^{*}(R_A;R_A)}{c_{2\beta}^{*}(R_A)} \geq 1$ for any $* = I, II, IV, V$.

■

Proof of Proposition 8

i. To begin, we compare the welfare levels across minimum scales. To this end, it is worth emphasizing that the expected utility is realized should the partial-run strategy profile specified in (2) be part of an equilibrium. Recall from the proof of Proposition 3 that if $\gamma > 2$, $R_H < (R_L)^{\gamma-1}$, and $\pi > \pi_F^{II}$, $q_L^* \leq q_F^{II} < q_F^I < q_F^{III}$. Further, we next establish $\pi_F^{IV} < \pi_F^{II}$. Note that if R_H is arbitrarily close to its lower bound R_L, $\pi_F^{II} = \pi_F^{IV} < 1$ and as R_H goes to its upper bound $(R_L)^{\gamma-1}$, $\pi_F^{IV} < \pi_F^{II} = 1$. Note also that for any $R_H \in \left(R_L, (R_L)^{\gamma-1}\right)$:

$$\frac{d\pi_F^{II}}{dR_H} = \frac{\left[\left(\frac{R_H}{R_L}\right)^{\frac{\gamma-1}{\gamma}} - (R_L)^{\frac{1-\gamma}{\gamma}}\right] - \left[\left(\frac{R_H}{R_L}\right)^{\frac{\gamma-1}{\gamma}} - (R_H)^{-\frac{1}{\gamma}}\right]\left(\frac{\gamma-1}{\gamma}\right)}{\left[(R_H)^{\frac{\gamma-1}{\gamma}} - 1\right]^2} > 0,$$

implying that $\pi_F^{II} > \pi_F^{IV}$ and thus $q_F^{II} < q_F^{IV}$. It also follows from the results summarized in Propositions 1 to 5 that if $\gamma \in (2, \infty)$, $\pi \in \left(\pi_F^{II}, 1\right)$, $R_H \in \left(R_L, \min\left\{(R_L)^{\gamma-1}, R_H^{IV}\right\}\right)$, and $q \in \left(0, \min\left\{q_L^{**}, q_A^V\right\}\right]$, run equilibria exist across allocations c^* for $* = I, II, ..., V$. Given that the bank implementing c^I delivers the highest welfare level to depositors, $W^I > \max\left\{W^{II}, W^{III}, W^{IV}, W^V\right\}$.

ii. Furthermore, consider cases in which agents pool their full endowments in the bank. As payment schemes c^I and c^{III} converge if the run probability is arbitrarily close to zero, $\max\left\{W^I, W^{II}, ..., W^V\right\} = W^I = W^{III}$.

■

References

- Andolfatto, D. and Nosal, E. (2018) "Bank Panics and Scale Economies." Working Paper 2018-016A, Federal Reserve Bank of St. Louis.
- Beccalli, E., Anolli, M., and Borello, G. (2015) "Are European Banks Too Big? Evidence On Economies of Scale." *Journal of Banking and Finance*, 58, 232-246.
- Bertolai, J.D.P., Cavalcanti, R. de O. and Monteiro, P.K. (2014) "Run Theorems for Low Returns and Large Banks." *Economic Theory*, 57, 223-252.
- Cooper, R. and Ross, T.W. (1998) "Bank Runs: Liquidity Costs and Investment Distortions." *Journal of Monetary Economics*, 41, 27-38.
- Diamond, D. and Dybvig, P. (1983) "Bank Runs, Deposit Insurance, and Liquidity." *Journal of Political Economy*, 91, 191-206.
- Ennis, H.M. and Keister, T. (2003) "Economic Growth, Liquidity, and Bank Runs." *Journal of Economic Theory*, 109, 220-245.
- Ennis, H.M. and Keister, T. (2009) "Bank Runs and Institutions: The Perils of Intervention." *American Economic Review*, 99, 1588-1607.
- Ennis, H.M. and Keister, T. (2010) "Banking Panics and Policy Responses." *Journal of Monetary Economics*, 57, 404-419.
- Gao, J. and Reed, R.R. (2021) "Sunspot Bank Runs and Fragility: The Role of Financial Sector Competition." *European Economic Review*, 139, 103877.
- Green, E.J. and Lin, P. (2000) "Diamond and Dybvig's Classic Theory of Financial Intermediation: What's Missing?" *Federal Reserve Bank of Minneapolis Quarterly Review*, Winter, 2-13.
- Green, E.J. and Lin, P. (2003) "Implementing Efficient Allocations in a Model of Financial Intermediation." *Journal of Economic Theory*, 109, 1-23.
- Gu, C., Monnet, C., Nosal, E. and Wright, R. (2020) "On the Instability of Banking and Financial Intermediation." BIS Working Papers No. 862.
- Keister, T. (2016) "Bailouts and Financial Fragility." *Review of Economic Studies*, 83, 704-736.

- Li, Y. (2017) “Interest Rates and Financial Fragility.” *Journal of Economic Dynamics and Control*, 82, 195-205.
- Peck, J. and Setayesh, A. (2022) “Bank Runs and the Optimality of Limited Banking.” *Review of Economic Dynamics*, forthcoming.
- Peck, J. and Shell, K. (2003) “Equilibrium Bank Runs.” *Journal of Political Economy*, 111, 103-123.
- Wallace, N. (1988) “Another Attempt to Explain an Illiquid Banking System: the Diamond and Dybvig Model with Sequential Service Taken Seriously.” *Federal Reserve Bank of Minneapolis Quarterly Review* Fall, 3-16.
- Wallace, N. (1990) “A Banking Model in Which Partial Suspension Is Best.” *Federal Reserve Bank of Minneapolis Quarterly Review*, Fall, 11-23.
- Wheelock, D. and Wilson, P. (2012) “Do Large Banks Have Lower Costs? New Estimates of Returns to Scale for U.S. Banks.” *Journal of Money, Credit and Banking*, 44, 171-199.
- Wheelock, D. and Wilson, P. (2018) “The Evolution of Scale Economies in US Banking.” *Journal of Applied Econometrics*, 33, 16-28.