Bank panics and scale economies

Bank panics and scale economies[*]

David Andolfatto
Federal Reserve Bank of St. Louis
Simon Fraser University

Ed Nosal
Federal Reserve Bank of Chicago

March 2017

Abstract

A *bank panic* is an expectation-driven redemption event that results in a self-fulfilling prophecy of losses on demand deposits. From the standpoint of theory in the tradition of Diamond and Dybvig (1983) and Green and Lin (2003), it is surprisingly difficult to generate bank panic equilibria if one allows for a plausible degree of contractual flexibility. A common assumption employed in the standard banking model is that returns are linear in the scale of investment. Instead, we assume the existence of a fixed investment cost, so that a higher risk-adjusted rate of return is available only if investment exceeds a minimum scale requirement. With this simple and empirically-plausible modification to the standard model, we find that bank panic equilibria emerge easily and naturally, even under highly flexible contractual arrangements. While bank panics can be eliminated through an appropriate policy, it is not always desirable to do so. We use our model

to examine a number of issues, including the likely effectiveness of recent financial market regulations. Our model also lends some support for the claim that low-interest rate policy induces a "reach-for-yield" phenomenon resulting in a more panic-prone financial system.

1 Introduction

Investors in market economies have access to an array of financial products with implicit or explicit options exercisable at their discretion. Bryant (1980) suggests that options like demandable and callable debt are a means to insure investors against unobservable liquidity risks. The same is likely true of any short-term financial instrument, like commercial paper and repo, where rollover is frequently an option.

The benefit of liquidity insurance, however, is undermined if the option to withdraw funding is triggered *en masse* by fear instead of fundamentals. Indeed, the notion that psychological factors are responsible for triggering financial crises has a long tradition in the history in economic thought. Diamond and Dybvig (1983) formalize this idea by demonstrating how simple bank deposit contracts can induce a coordination game exhibiting two equilibrium outcomes. In the *fundamental* equilibrium, all depositors represent their liquidity needs truthfully, so that options are exercised for fundamental economic reasons only. In the *bank panic* equilibrium, depositors not in need of liquidity pretend that they are. In this case, options are exercised out of a fear that little will be left for latecomers if other depositors are similarly misrepresenting themselves. In this way, the mere expectation of widespread redemptions can become a self-fulfilling prophecy.

While the notion of a panic-induced crisis has certain appeal, the phenomenon is difficult to identify empirically. An alternative and equally plausible view asserts that financial instability and its associated emotional trauma is merely symptomatic of deteriorating fundamentals experienced in the broader economy prior to an economic downturn; see Gorton (1988) and Allen and Gale (1998). While there is merit to this view, it is not inconsistent with the possibility that some crises are panic-driven. In particular, not all financial crises are associated with recessions; see Capiro and Klingebiel (1997).

2

Because it is difficult to discriminate empirically between panic-based and fundamental-based explanations of crises, policymakers should hedge their bets when designing financial regulation.[1] We think that the proper hedge in this case might be usefully informed by theoretical, as well as empirical, plausibility. Given the current state of theory, a case could be made for adjusting posterior odds in favor of fundamentals over panics. The basis for this assessment rests on the apparent difficulty of generating bank panics in model economies, at least for economies that permit an empirically plausible degree of contractual flexibility.

To explain what we mean by this, note that the seminal model of Diamond and Dybvig (1983) does *not* exhibit bank panics when banks adopt a simple suspension scheme,[2] a device that was actually used—sometimes suuccessfully—to halt runs. Nor do Diamond and Dybvig (1983) establish the existence of bank panics when they extend their basic model to include stochastic withdrawals.[3] In fact, Green and Lin (2000, 2003) demonstrate that bank panics do not exist in a version of the Diamond and Dybvig (1983) with stochastic withdrawals, even when sequential service is modeled seriously (Wallace, 1988). Peck and Shell (2003) modify informational assumptions in Green and Lin (2003) but must resort to a numerical example to demonstrate the existence of an equilibrium bank panic.[4] Andolfatto *et. al.* (2016) show that it can be extremely difficult to generate bank panics under general contractual mechanisms. The generic non-existence of panic equilibria in this body of theory could reasonably be taken by policymakers as evidence suggesting that bank panics can be safely ignored as a practical policy concern. The results of our present paper suggest that such a conclusion may not be warranted.

We pose a theory in which bank panics that can arise easily and natu-

---

[1] Ideally, one can hope for legislative solutions that remain robust to the actual driving mechanism underlying financial crises. But many policies rely heavily on what on what is assumed about the underlying model of crisis; see, Keister and Narasiman (2016).

[2] Diamond and Dybvig (1983) make this point in Section III of their article.

[3] Diamond and Dybvig (1983) demonstrate that the first-best state-contingent allocation is not attainable under a sequential service constraint. They then go on to show how deposit insurance can be used to resurrect efficiency, assuming that the government is not restricted by sequential service.

[4] Their example arguably relies on a parameter living in a remote region of the parameter space. In particular, the marginal utility of consumption for impatient depositors is an order of magnitude larger than for patient depositors.

rally whenever short-term debt is used to finance investments characterized by even a modest degree of increasing returns to scale. And, while our explanation does not require sequential service, it is certainly not compromised if sequential service is imposed. Our idea is based on the notion that many investments entail some fixed costs. A large commercial development project, for example, requires a significant outlay in capital and labor services, e.g., cranes and crane operators, that must be paid regardless of how much construction activity is actually taking place on site. Because of the existence of these fixed costs, it is not implausible that higher risk-adjusted returns are available for larger scale projects. An alternative and mathematically equivalent interpretation of our scale economy assumption is that the unit cost of liquidating capital rises measurably if the volume of liquidation exceeds a given threshold. Such a threshold may be very low for some asset classes and higher for others. As we explain below, the specter of panics arises when this latter class of investments is funded with short-term debt.

The basis of our analysis is the Green and Lin (2000, 2003) framework, which is a finite-trader version of the Diamond and Dybvig (1983) model with aggregate liquidity risk.[5] The standard formulation assumes a constant returns to scale investment technology and a sequential service constraint. By contrast, we assume an increasing returns to scale investment technology and no sequential service restriction. As is also standard, we assume that all agents in the model have access to the same investment technology and that the distinguishing characteristic of a bank is how it funds investments. In particular, banks offer their depositor base products that provide the flexibility to withdraw funding on short notice.

We show that when private information over depositor type and a modest scale economy are present, bank panic equilibria are likely to exist even under highly flexible contractual structures. Eliminating panic equilibria in our model is possible by segregating some assets on bank balance sheets or through the use of liquidity fees in when redemption activity is unusually high. However, eliminating panic equilibria necessarily comes at the cost of reduced risk-sharing and, as such, may not always be optimal. On the other hand, we show how the expected cost of eliminating bank panics is decreasing in bank size. Hence, larger or more connected banking systems can be made more stable at a lower cost—a prediction that bears up well

---

[5]For a comprehensive review of this literature, see Ennis and Keister (2010).

against the historical evidence, e.g., Williamson (1989).

The paper is organized as follows. Section 2 describes the economic environment. In Section 3, we characterize the set of efficient incentive-compatible allocations for economies subject to private information and economies of scale. We establish the existence of panic equilibria in Section 4. Section 5 investigates an extension of the model that permits two asset classes: a narrow bank and a shadow bank. The conclusions of this section provide some support for the notion that low real rates of return on safe asset classes can promote financial instability through a reach-for-yield behavior. Finally, we provide a discussion of pertinent issues in Section 6.

2   The model

Our framework of analysis is the finite-trader version of the Diamond and Dybvig (1983) model with aggregate liquidity risk developed by Green and Lin (2000, 2003). The economy has two dates, $t = 1, 2$, and a finite number $N \geq 3$ of *ex ante* identical individuals. Individuals are subject to a shock at date $t = 1$ that determines their preference type: *impatient* or *patient*. Let $0 < \pi < 1$ denote the probability that an individual is impatient. Let π_n denote the probability that $0 \leq n \leq N$ individuals are impatient. We assume that individual types are *i.i.d.* so that $\pi_n = \binom{N}{n}\pi^n(1 - \pi)^{N-n}$. The distribution of types has full support since $0 < \pi_n < 1$ for all n.

Impatient individuals want to consume at date 1 only. Patient individuals are willing to defer consumption to date 2; technically, they are indifferent between consuming at dates 1 and 2. Let c_t represent the consumption of an individual at date t. *Ex ante* preferences are given by

$$U(c_1, c_2) = \pi u(c_1) + (1 - \pi)u(c_1 + c_2), \tag{1}$$

where $u(c) = c^{1-\sigma}/(1 - \sigma)$, $\sigma > 1$.

Each individual is endowed with a claim to y units of date 1 output. There is a technology that transforms k units of date 1 output into $F_\kappa(k)$ units of date 2 output according to

$$F_\kappa(k) = \begin{cases} rk & \text{if} \quad k < \kappa \\ Rk & \text{if} \quad k \geq \kappa \end{cases}, \tag{2}$$

where $0 < r < 1 < R$ and $0 \leq \kappa < Ny$. The high rate of return R is available only if the level of investment exceeds a minimum scale requirement of κ.[6] When the minimum scale is not met the rate of return reflects the cost of intermediated storage, indexed by the parameter $1 - r$. Technology (2) is a generalization of the standard specification used in the literature, which assumes $\kappa = 0$ so that $F_0(k) = Rk$ for all $k > 0$.

There are benefits to cooperation. First, there are the usual gains associated with sharing risk. Second, and absent from the standard model, scale economies are more easily attained when resources are pooled.

In what follows, we refer to a risk-sharing arrangement that pools resources and exploits scale economies as a *bank*. Individuals who deposit resources with the bank are referred to as *depositors*. A bank can be viewed as a resource-allocation mechanism that pools the resources of the N depositors before they learn their types. In exchange for deposits, the bank issues state and time-contingent deposit liabilities. What makes our environment a bank problem, instead of a standard insurance problem, is that the bank cannot verify depositor types. Specifically, we assume that the depositor's type—his liquidity preference—is revealed only to the depositor. Because liquidity preference is private information, the optimal risk-sharing arrangement will embed options to withdraw funds on demand at either date. As a result, the optimal contract will resemble conventional demand deposit liabilities.

Depositors are spatially separated in the sense of Townsend (1987); that is, they are unable to communicate with each other. However, depositors can communicate with the bank at date 1 and date 2, subject to some restrictions. In particular, following their initial deposit, we assume that it is prohibitively costly for depositors to contact the bank more than one time. Given this communication protocol, depositors communicate with the bank either at date 1 or at date 2, but not at both dates. This protocol is adopted in part for its descriptive realism. A more general communications protocol would allow depositors to communicate with the bank at all dates and, in particular, on dates where they do not want to withdraw funds. As it turns out, our assumed communications protocol may or may not be restrictive, depending

---

[6] One could easily generalize the analysis to permit multiple threshold levels and associated rates of return. For most of our analysis, we stick to one threshold level for the sake of making our point in the simplest manner possible.

on circumstances. We identify in due course when the protocol is binding and when it is not.

At each date the bank collects all deposit requests and then sets a withdrawal limit for each request conditional on that period's aggregate withdrawal demand. So, unlike Wallace (1988), Green and Lin (2003) and many others, we do *not* assume that depositors are served sequentially. We abstract from sequential service not because we think it is unrealistic or problematic, but simply because we do not need it for our results.[7]

Our communications protocol implies that the date 1 consumption payments specified in the bank contract need only be conditioned on the number of depositors m who contact the bank at date 1, where $m \in \{0, 1, ..., N\}$. In particular, if m depositors contact the bank at date 1, then each depositor receives $c_1(m)$ units of date 1 consumption. Depositors who contact the bank at date 2 each receive $c_2(m) = F_\kappa[Ny - mc_1(m)]/(N - m)$ units of date 2 consumption. Hence, the bank offers depositors an allocation or *contract* $(\mathbf{c}_1, \mathbf{c}_2)$, where $\mathbf{c}_1 = [c_1(1), \dots, c_1(N)]$ and $\mathbf{c}_2 = [c_2(0), c_2(1), \dots, c_2(N - 1)]$.

By construction contract $(\mathbf{c}_1, \mathbf{c}_2)$ is feasible. To ensure that contract $(\mathbf{c}_1, \mathbf{c}_2)$ promotes efficient resource allocation, it should be structured in a way that gives depositors an incentive to represent their preferences truthfully. But, since liquidity preferences are private information, depositors may have an incentive to misrepresent themselves to the bank. Since depositor behavior will generally depend on how they believe other depositors are likely to behave, private information renders the environment strategic. We now describe the game played by depositors.

A *depositor game* is played after individuals deposit their resource endowments with the bank. Specifically, after they learn their types, depositor $j \in \{1, 2, ..., N\}$ chooses an action $t_j \in \{1, 2\}$, where t_j denotes the date depositor j contacts the bank. Depositor j knows only his own type when he chooses t_j. In particular, depositor j does not know the number of impatient depositors n in the economy when choosing his action. A *strategy profile* $\boldsymbol{t} \equiv \{t_1, t_2, ..., t_N\}$ implies a number $m \in \{0, 1, ..., N\}$, the number of depositors that contact the bank at date 1. Since efficiency dictates that impatient depositors consume at date 1 and patient depositors consume at date 2, a *truth-telling strategy* is a strategy profile that has impatient depositors contacting the bank at date 1 and patient depositors contacting the bank at

---

[7]Our results remain robust to modeling sequential service.

date 2. A truth-telling strategy implies that $m = n$.

A strategy profile $\mathbf{t}$ and its associated m constitutes a Bayes-Nash *equilibrium* of the depositor game with allocation $(\mathbf{c}_1, \mathbf{c}_2)$ if $t_j \in \mathbf{t}$ is a best response for depositor j against $\mathbf{t}_{-j} \equiv \{t_1, ..., t_{j-1}, t_{j+1}, ..., t_N\}$ for all $j \in \{0, 1, ..., N\}$. An allocation $(\mathbf{c}_1, \mathbf{c}_2)$ is said to be *incentive compatible* (IC) if a truth-telling strategy is an equilibrium for the depositor game.

Since, in any equilibrium $c_1(m) > 0$, impatient depositors always tell the truth—they contact the bank at date 1. A patient depositor tells the truth by contacting the bank at date 2; he has an incentive to do so iff

$$\sum_{n=0}^{N-1} \Pi^n u\left[c_2(n)\right] \geq \sum_{n=0}^{N-1} \Pi^n u\left[c_1(n+1)\right], \tag{3}$$

where Π^n is the conditional probability that there are n impatient individuals given there is at least one patient individual, i.e.,

$$\Pi^n = \frac{\binom{N-1}{n}(1-\pi)^{N-n-1}\pi^n}{\sum_{n=0}^{N-1}\binom{N}{n}(1-\pi)^{N-n-1}\pi^n}.$$

If feasible contract $(\mathbf{c}_1, \mathbf{c}_2)$ satisfies (3), then there exists an equilibrium where all depositors play the truth-telling strategy.

If an allocation $(\mathbf{c}_1, \mathbf{c}_2)$ satisfies (3), there *may* exist other equilibrium outcomes. In particular, there may exist an equilibrium in which depositors play a *panic strategy*. A panic strategy is a strategy profile that has all depositors contacting the bank at date 1, i.e., $t_j = 1$ for all j and, as a result, $m = N$ for any $n \leq N$.

3 Efficient incentive-compatible allocations

In this section we characterize the properties of efficient incentive-compatible allocations. We begin with the standard case where the return to investment is invariant to its scale, $\kappa = 0$. We then study the case in which the return to investment is subject to a scale economy, $\kappa > 0$.

8

3.1 Linear technology

We first characterize the unconstrained efficient allocation. The unconstrained allocation *assumes* that impatient depositors contact the bank at date 1 and patient depositors contact the bank at date 2. This implies that n impatient depositors contact the bank at date 1. The unconstrained efficient allocation is given by an allocation $(\mathbf{c}_1, \mathbf{c}_2) \equiv \{c_1(n), c_2(n)\}_{n=0}^{N}$ that maximizes the expected utility of the representative, *ex ante* identical depositor,[8]

$$\max_{\{c_1(n)\}} \sum_{n=0}^{N} \pi_n \{nu\left[c_1(n)\right] + (N-n)u\left[c_2(n)\right]\} \tag{4}$$

subject to the resource constraints

$$nc_1(n) = Ny - k(n) \tag{5}$$
$$Rk(n) = (N-n)c_2(n) \tag{6}$$

which, when combined, yields

$$nc_1(n) + \frac{(N-n)c_2(n)}{R} = Ny, \tag{7}$$

for all $n \in \{0, 1, ..., N\}$. Let $(\mathbf{c}_1^*, \mathbf{c}_2^*)$ denote the solution to the problem above. It is easy to show that a unique solution exists and satisfies

$$u'[c_1^*(n)] = Ru'[c_2^*(n)] \; \forall n < N, \tag{8}$$

with $c_2^*(N) = 0$ and the resource constraint (7). Given our CES preference specification, the solution is available in closed-form,

$$c_1^*(n) = \frac{Ny}{n + (N-n)R^{1/\sigma - 1}} \tag{9}$$
$$c_2^*(n) = R^{1/\sigma} c_1^*(n), \tag{10}$$

for all $n < N$ with $c_1^*(N) = y$ and $c_2^*(N) = 0$. Note that for all $n < N$ depositors engage in risk-sharing since $y < c_1^*(n) < c_2^*(n) < Ry$. Moreover, because $\sigma > 1$ and $R > 1$ imply $R^{1/\sigma - 1} < 1$, it follows that both $c_1^*(n)$ and $c_2^*(n)$ are decreasing in n.

---

[8] Green and Lin (2000, 2003) provide a characterization of the efficient allocation when there is no sequential service and the investment technology is linear. We present this allocation here because it is relevant for the efficient allocation for the scale economy.

Property 1 $c_2^*(n) > c_1^*(n) > c_1^*(n+1)$ for all $n \in \{0, 1, ...N - 1\}$.

One implication of Property 1 is that the short and long-term rates of return on deposits, $c_1^*(n)/y$ and $c_2^*(n)/y$, respectively, are both decreasing in the level of date 1 redemption activity, n. Wallace (1988) interprets $c_1^*(n) > c_1^*(n+1)$ as a partial suspension scheme which, by construction, is efficient here.

Using (6) and (9), the efficient investment or capital schedule, $k^*(n)$, is

$$k^*(n) = \frac{(N-n)R^{1/\sigma - 1}Ny}{n + (N-n)R^{1/\sigma - 1}}. \tag{11}$$

Notice that $k^*(n)$ is decreasing n and that $k^*(n) < (N-n)y$.[9] A large value for n means that the aggregate demand for early withdrawals is high. In this case, it makes sense to devote less per capita resources to investment, reducing the effective return for late withdrawals, spreading the additional early resources more thinly among the more numerous impatient depositors. Note that high realizations for n can be interpreted as recessionary events or investment collapses associated with large numbers of depositors making early withdrawals. These events, however, are driven by economic fundamentals. A bank could mitigate the economic impact of these "fundamental runs" by expanding its depositor base, N. Our full support assumption, however, implies that the probability that all depositors desire early withdrawal, π_N, will remain strictly positive, even if $\pi_N \to 0$ as $N \to \infty$.

There are two important results associated with the solution $(\mathbf{c}_1^*, \mathbf{c}_2^*)$. First, it follows immediately from Property 1 that it $(\mathbf{c}_1^*, \mathbf{c}_2^*)$ is incentive compatible. In particular, Property 1 implies that a patient depositor's consumption is higher than an impatient depositor's consumption in a truth-telling equilibrium. And, if a patient depositor chooses to contact the bank at date 1, then date 1 consumption for all depositors is lowered relative to what they would have enjoyed if the patient depositor had instead contacted the bank at date 2. Therefore, assuming that all other depositors are playing truthfully, a patient depositor has no incentive to contact the bank at date 1, since doing so would result in a strictly lower payoff. Hence, allocation $(\mathbf{c}_1^*, \mathbf{c}_2^*)$ satisfies the incentive compatibility condition (3).

---

[9]The latter inequality is a direct implication of risk-sharing. Capital per patient depositor, $k^*(n)/(N-n)$ is also a decreasing function of n.

Second, the truth-telling equilibrium that implements $(\mathbf{c}_1^*, \mathbf{c}_2^*)$ in the depositor game is *unique*. To see this, first note that it is a dominant strategy for impatient depositors to contact the bank at date 1 since $c_1(m) > 0$ for all $m \in \{1, 2, ..., N\}$. It is also a dominant strategy for the patient depositor to contact the bank at date 2. To understand this, let the patient depositor conjecture *anything* about the behavior of other depositors. Any such conjecture generates an $m \in \{1, 2, ..., N\}$ for the patient depositor. If the patient depositor plays truthfully, he gets $c_2^*(m) > c_1^*(m) > c_1^*(m+1)$, where these inequalities again follow from Property 1. Hence, regardless of the volume of early withdrawals, a patient depositor will always receive a higher payoff by postponing his withdrawal to the later date, i.e., $c_2^*(m) > c_1^*(m+1)$. Since it is a dominant strategy for a patient individual to contact the bank at date 2, the allocation $(\mathbf{c}_1^*, \mathbf{c}_2^*)$ can be uniquely implemented as an equilibrium in dominant strategies.

Proposition 1 *[Green and Lin, 2003] The unconstrained efficient allocation $(\mathbf{c}_1^*, \mathbf{c}_2^*)$ is uniquely implementable as a Bayes-Nash equilibrium of the depositor game when depositor types are private information.*

Proposition 1 implies that private information is not an obstacle to implementing the first-best allocation uniquely. Notice too our assumed communications protocol does not hinder the implementation of the unconstrained efficient allocation under private information. An implication of Proposition 1 is that bank panics do not exist.

Green and Lin (2003) impose a sequential service constraint on the above environment and ask if a bank panic equilibrium is possible. They demonstrate that while sequential service alters the properties of the efficient allocation, it does not overturn the uniqueness result.[10] Peck and Shell (2003) study Green and Lin (2003) with sequential service and assume that depositors do not know their place in the date 1 service queue. They demonstrate that a panic equilibrium can exist, at least, in some regions of the parameter space.[11] Below, we ignore sequential service and ask whether scale economies might be a more plausible source of bank panics.

---

[10]The Green and Lin (2003) results are generalized in Andolfatto, Nosal and Wallace (2007).

[11]Peck and Shell (2003) require that the marginal utility of the impatient depositor be very high compared to the marginal utility of the patient depositor. This assumption implies that the date 1 consumption payoff will be relatively high, which provides an

3.2 Scale economies

We now examine an economy where $0 < \kappa < Ny$ and $N \geq 3$. Note that for a given κ, the investment technology is approximately linear in large economies; that is, $\kappa/(Ny) \to 0$ as $N \to \infty$. Since our results below hold for an arbitrarily large but finite N, there is a sense in which our model constitutes an extremely mild departure from the standard linear-technology case.

To make our point in the simplest possible way, we assume that $\kappa = 2y$.[12] That is, we assume that the investment technology requires the resources of at least two depositors to achieve minimum scale. As above, we begin by characterizing the unconstrained efficient allocation for the scale economy, denoted by $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$, by assuming that depositors play truthfully.

One can demonstrate that there exists an R_0 such that for all $1 < R \leq R_0$, the inequality $2y < k^*(N-3)$ holds.[13] Assume for simplicity that $R \leq R_0$, in which case we have $k^*(N-2) < 2y < k^*(N-3)$, where $k^*(n)$ is given by (11). Therefore, if there are $n \leq N-3$ impatient depositors, then the level of investment associated with allocation $(\mathbf{c}_1^*, \mathbf{c}_2^*)$ for all $n \leq N-3$ will have the high rate of return R because $k^*(n) > \kappa = 2y$. However, if the number of impatient depositors is $n \in \{N-2, N-1\}$, then the unconstrained efficient level of investment for the linear economy is unable to achieve the high rate of return R since $k^*(n) < \kappa = 2y$. Achieving scale is irrelevant when $n = N$ because all depositors are impatient, in which case $\hat{k}(N) = k^*(N) = 0$. Consequently, we have

Property 2 $\{\hat{c}_1(n), \hat{c}_2(n)\} = \{c_1^*(n), c_2^*(n)\}$ for all $n \in \{0, 1, ..., N-3, N\}$.

We now characterize the unconstrained efficient allocation in the scale economy for the remaining two cases: $n = N-1$ and $n = N-2$. Notice that for $j = 1, 2$, $\hat{k}(N-j) = k^*(N-j) < \kappa$ cannot be a solution to the bank's problem since these levels of investment do not earn the high return R. The

---

incentive for patient depositors to panic, assuming that all other patient depositors are panicking.

[12] We explain below why our results hold more generally.

[13] For $R > R_0$, $k^*(N-3) < 2y$. Qualitatively speaking, none of our results are affected in this case; see footnote 14. Efficient risk sharing implies that $k^*(N-2) < 2y$, independent of the value of $R > 1$.

unconstrained efficient contract will take one of two forms in these states: (i) invest at level $k \geq \kappa$ and earn the rate of return $R > 1$; or (ii) invest at level $k < \kappa$ and earn the rate of return $r < 1$. The first option provides depositors with a high rate of return but at the cost of less risk sharing. The second option provides depositors with efficient risk sharing but at the cost of a low rate of return.

Let's first examine the case of two patient depositors, $n = N-2$. Consider first the high-return option, which requires $\hat{k}^H(N-2) = \kappa = 2y$. Using (5), we see that $\hat{k}^H(N-2) = 2y$ implies $\hat{c}_1^H(N-2) = y$ and (6) implies that $\hat{c}_2^H(N-2) = Ry$. This option exploits scale but fails to provide risk-sharing. In particular, note that the impatient depositor receives his autarkic payoff.

Consider now the low-return option, $\hat{k}^L(N-2) < \kappa$. In this case, $\hat{c}_1^L(N-2)$ and $\hat{c}_2^L(N-2)$ are determined by replacing R with r in (9) and (10), respectively, with $n = N-2$, i.e.,

$$\hat{c}_1^L(N-2) = \frac{N}{N-2+2r^{1/\sigma-1}}y \tag{12}$$

$$\hat{c}_2^L(N-2) = r^{1/\sigma}\hat{c}_1^L(N-2) \tag{13}$$

Since $\sigma > 1 > r$, it follows that $\hat{c}_2^L(N-2) < \hat{c}_1^L(N-2) < y$. Because $\hat{c}_1^H(N-2) > \hat{c}_1^L(N-2)$ and $\hat{c}_2^H(N-2) > \hat{c}_1^L(N-2)$, the high-return option dominates the return-return option and we have

Property 3 When $n = N-2$, the unconstrained efficient allocation in the scale economy is given by the high-return option, where $\hat{c}_1^H(N-2) = y$ and $\hat{c}_2^H(N-2) = Ry$.[14]

---

[14]Property 3 is relevant for the case where $R > R_0$, which implies that $k^*(N-3) < 2y$. Suppose that $R > R_0$ and the bank chooses $\hat{k}(N-3) = 2y$, the high-return option. This level of investment implies that date 1 consumption is given by $c_1^*(N-3) > \hat{c}_1^H(N-3) = [(N-2)/(N-3)]y > y$ and date 2 consumption given by $\hat{c}_2^H(N-3) = 2yR/3 > c_2^*(N-3) > y$. This investment strategy provides a higher expected utility than autarky. Notice also that this allocation is incentive compatible in the sense that a patient depositor strictly prefers the date 2 allocation to the date 1 allocation. Suppose, instead, that the bank tries to provide better risk-sharing for the depositors by investing in the low return r technology. Since $r < 1$, both the date 1 and date 2 efficient consumption allocations will be less than y. But if the date 1 consumption allocation is less than y, then capital, $\hat{k}(n)$, must exceed $3y > \kappa$ for $n = N-3$. Hence, it is not possible to construct an efficient risk-sharing allocation that generates the low return r. Any allocation that generates a low investment

We now examine the case of one patient depositor, $n = N-1$. Once again we evaluate the payoffs associated with the high- and low-return options. To achieve the high return, the level of investment must be set to $\hat{k}^H(N-1) = \kappa = 2y$. Using (5) and (6), we see that $\hat{k}^H(N-1) = 2y$ implies $\hat{c}_1^H(N-1) = y(N-2)/(N-1)$ and $\hat{c}_2^H(N-2) = 2Ry$. Since $\hat{c}_1^H(N-1) < y < R < \hat{c}_2^H(N-2)$, the high-return option again comes at the cost of poorer risk-sharing. Notice that the payoff to the impatient depositor is in this case less than his autarkic payoff.

Consider next the low-return option, which implies that $\hat{k}^L(N-1) < \kappa$. In this case, $\hat{c}_1^L(N-1)$ and $\hat{c}_2^L(N-1)$ are determined by replacing R with r in (9) and (10), respectively, with $n = N-1$, i.e.,

$$\hat{c}_1^L(N-1) = \frac{N}{N-1+r^{1/\sigma-1}}y, \tag{14}$$

$$\hat{c}_2^L(N-1) = r^{1/\sigma}c_1^L(N-1). \tag{15}$$

Inspecting conditions (14) and (15), leads us to

Lemma 1 *For r arbitrarily close to (but less than) unity, $\hat{c}_1^L(N-1) \approx \hat{c}_2^L(N-1) \approx y = \hat{c}_1(N)$, where $\hat{c}_2^L(N-1) < \hat{c}_1^L(N-1) < y$.*

Lemma 1 tells us that if r is close to unity, then the payoffs to patient and impatient depositors are about equal to y. Let us assume that $r < 1$ is arbitrarily close to unity. Then by Lemma 1, the expected utility payoff associated with the low-return option is approximately $u(y)$. The expected utility associated with the the high-return option is

$$\left(\frac{N-1}{N}\right) u\left(\frac{N-2}{N-1}y\right) + \left(\frac{1}{N}\right) u\left(2Ry\right).$$

Since the high-return option entails poorer risk-sharing than the low-return option, we expect the benefit of the former option to diminish with depositors'

---

return (by simply having the bank choose capital less than $2y$) will result in an expected utility that is less than that associated with an allocation that has capital investment equal to $2y$. Therefore, when $R > R_0$, the bank will always undertake a capital investment of $2y$ for $n = N-3$ and the resultant allocation is incentive compatible. It is this latter feature, incentive compatibility, that implies our main results remain qualitatively unaffected when $R > R_0$.

appetite for risk. Indeed, we can demonstrate that for preferences with $\sigma \geq 2$, the expected utility associated with the low-return option exceeds that of the high-return option, i.e.,

$$\left(\frac{N-1}{N}\right) u \left(\frac{N-2}{N-1}y\right) + \left(\frac{1}{N}\right) u\left(2Ry\right) < u\left(y\right).$$

To see this, note that our functional form for $u(\cdot)$ implies that the above inequality can be written as

$$(N-1)(1-\sigma)^{-1} \left(\frac{N-2}{N-1}y\right)^{1-\sigma} + (1-\sigma)^{-1}\left(2Ry\right)^{1-\sigma} < N(1-\sigma)^{-1}(y)^{1-\sigma}.$$

Since $1 - \sigma < 0$, the above inequality becomes

$$(N-1)\left(\frac{N-2}{N-1}\right)^{1-\sigma} + (2R)^{1-\sigma} > N. \tag{16}$$

Furthermore, since $\sigma \geq 2$, it follows that

$$\left(\frac{N-1}{N-2}\right)^{\sigma-1} > \frac{N}{N-1}, \tag{17}$$

since $(N-1)/(N-2) > N/(N-1)$. Condition (17) in turn implies condition (16). Therefore, we have the following result,

Property 4 When $n = N - 1$, $r < 1$ sufficiently close to unity and $\sigma \geq 2$, the unconstrained efficient allocation in the scale economy is given by the low-return option, where $\hat{c}_1^L\left(N-1\right)$ and $\hat{c}_2^L(N-1)$ are determined by (14) and (15), respectively.

Property 4 has the feature that when $n = N - 1$ the bank "breaks the buck" in the sense that for every unit that individuals deposit at the bank, they receive less than a unit payoff at date 1, as well as date 2. The empirical relevance of this observation is discussed in Section 5, below.

Properties 1-4 fully characterize the unconstrained efficient allocation in a scale economy where $\kappa = 2y$, $r < 1$ sufficiently close to unity, $\sigma \geq 2$ and $R \leq R_0$. In particular, the unconstrained efficient allocation, $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$, is given by

$$\{\{c_1^*(n), c_2^*(n)\}_{n=0}^{N-3}, \hat{c}_1^H(N-2), \hat{c}_2^H(N-2), \hat{c}_1^L(N-1), \hat{c}_2^L(N-1), c_1^*(N), c_2^*(N)\}$$

15

We now demonstrate that the unconstrained efficient allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ is incentive compatible. As before, impatient depositors do not have an incentive to misrepresent themselves, so they always contact the bank at date 1. With respect to the patient depositors, recall that in states $n \leq N - 2$, we have $\hat{c}_2(n) > \hat{c}_1(n) > \hat{c}_1(n+1)$ from Properties 1 and 2; and in state $n = N - 1$, we have $\hat{c}_2(N-1) < \hat{c}_1(N) = y$. Assuming that all other patient depositors contact the bank at date 2, a patient depositor will contract the bank at date 2 if the unconstrained efficient allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ satisfies (3) or equivalently if

$$\sum_{n=0}^{N-2} \Pi^n u\left[\hat{c}_2(n)\right] - \sum_{n=0}^{N-2} \Pi^n u\left[\hat{c}_1(n+1)\right] \geq \Pi^{N-1} u\left[c_2(N-1)\right] - \Pi^{N-1} u\left[c_1(N)\right].$$
(18)

Since $\hat{c}_2(n) > \hat{c}_1(n) > \hat{c}_1(n+1)$ for all $n \leq N - 2$, the left side is strictly greater than zero. When $r < 1$ is arbitrarily close to unity, the right side is negative but arbitrarily close to zero. Hence, (18) is satisfied with a strict inequality.[15] Therefore, we have the following result,

Proposition 2 *The unconstrained efficient allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ in the scale economy characterized by $\kappa = 2y$, $\sigma \geq 2$ and $r < 1$ arbitrarily close to 1 can be implemented as a truth-telling equilibrium of the depositor game.*

A few remarks are in order.

1. Allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ respects the assumed communication protocol, namely, that depositors can contact the bank at date 1 *or* date 2 but not both. It turns out that this protocol imposes restrictions on allocations in one state of the states of the world, state $n = N - 1$. To see this, suppose that depositors could contact the bank at dates 1 *and* 2. Then in state $n = N - 1$, the efficient bank contract would provide a payment y to *all* depositors at date 1. Notice that this payment dominates the payments from allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ since $y > \hat{c}_1^L(N-1) > \hat{c}_2^L(N-1)$.

---

[15]Since $\hat{c}_2(n) > \hat{c}_1(n) > \hat{c}_1(n+1)$ when $n \leq N-2$, $r < 1$ need *not* be arbitrarily close to unity to have allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ satisfy incentive-compatibility. The condition that $r < 1$ is arbitrarily close to unity simply guarantees that the incentive-compatibility condition (3) will hold with strict inequality. We discuss this in more detail, below.

2. It is straightforward to show that the expected utility of allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ exceeds autarky.[16]

3. We assume that $\kappa = 2y$. We emphasize that the qualitative properties of the unconstrained efficient allocation remain valid for more general cases. In particular, consider an arbitrary minimum scale $2y < \kappa < (N-1)y$. Then, there exists some $3 < j \leq N$ for which Property 2 continues to hold for $n \in \{0, 1, ..., N-j, N\}$. The remaining state-contingent allocations $N - j < n < N$ will then be characterized either the high-return investment option (and poorer risk-sharing) or the low-return investment option (and efficient risk-sharing). For minimum scales $\kappa \geq 2y$, it will always be the case that the low-return option is chosen in state $n = N - 1$. One of the main results of this paper, in Section 4, depends on the low-return investment option being used in state $n = N - 1$. Since all minimum scales $\kappa \geq 2y$ have the low-return option being used in state $n = N - 1$, we have assumed $\kappa = 2y$ because this case is the most straightforward to characterize.

4. Lemma 1 and Proposition 2 assume that $r < 1$ is arbitrarily close to 1. In this case we are able to show that the low-return investment option is strictly preferred to the high-return investment option and that allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ is strictly preferred to autarky. Clearly, r can be made less than unity without upsetting either of these results.

5. If $R > R_0$, then the discussion in footnote 14 implies that the unconstrained efficient allocation is still characterized by $\hat{c}_2(n) > \hat{c}_1(n) > \hat{c}_1(n+1)$ for all $n \leq N - 2$. Hence, Proposition 2 remains valid.

Proposition 2 tells us that allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ can be implemented as a truth-telling equilibrium. We now examine whether the truth-telling equilibrium is the *unique* equilibrium for the depositor game.

---

[16]In autarky all agents consume their endowment at date 1 and receive an expected utility equal to $u(y)$. Allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ delivers depositors an expected utility that strictly exceeds $u(y)$ in states $n \in \{0, \ldots, N - 2\}$, an expected utility that equals $u(y)$ in state $n = N$ and an expected utility that is arbitrarily less than $u(y)$ in state $n = N - 1$.

4 Panic equilibria

All N individuals deposit their endowment of claims to y units of date 1 output with the bank before date 1 in exchange for the deposit contact $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$. We investigate if deposit contact $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ generates equilibria other than the truth-telling equilibrium. Our main result is stated in the following proposition.

Proposition 3 *The unconstrained efficient allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ admits a panic equilibrium, where all N individuals contact the bank at date 1 regardless of their true liquidity needs, i.e., $m = N$ for all $n \in \{0, 1, \ldots, N-1\}$.*

To check the validity of Proposition 4, we begin by proposing a strategy profile where all N depositors contact the bank at date 1 and ask whether a patient depositor has an incentive to follow the crowd and contact the bank at date 1. If a patient depositor plays the proposed strategy profile and contacts the bank at date 1, he receives a payoff $\hat{c}_1(N) = y$. If he instead defects from the proposed strategy profile and contacts the bank at date 2, then, since the other $N-1$ depositors contact the bank at date 1, he receives a payoff of $\hat{c}_2^L(N-1) < y$. Clearly, a patient depositor does not have an incentive to defect from the strategy of contacting the bank at date 1 and a bank panic equilibrium exists.[17]

When the unconstrained efficient contract is characterized by the low return investment option in state $n = N - 1$, a bank panic equilibrium always exists. Hence, bank panics appear to be a robust feature of banking environments that are characterized by scale economies. Compare this with standard banking environments that feature linear investment technologies and sequential service. In these economies, whether or not bank runs exist depend critically on model parameterizations; see, for example, Peck and Shell (2003).

---

[17]If $r = 1$, a patient depositor would be indifferent between misrepresenting himself or not. Thus, a panic equilibrium remains possible even if $r = 1$, though it seems unlikely to survive any reasonable equilibrium refinement.

4.1 Eliminating panics

Following Peck and Shell (2003) we can construct sunspot equilibria using allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ by introducing a parameter $0 \leq \theta \leq 1$, the probability of a sunspot. A sunspot is an extrinsic event observed after individuals make their bank deposits but before they learn their type. With probability $1 - \theta$ the sunspot is not observed and depositors play the truthtelling equilibrium described in Proposition 2. With probability θ the sunspot is observed and all depositors play the panic equilibrium described in Proposition 3. Let $V(\kappa, R, \theta)$ denote the expected utility associated with allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ when a sunspot occurs with probability θ, i.e.,

$$V(2y, R, \theta) \equiv (1 - \theta)E\left[U(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)\right] + \theta u(y).$$

Clearly, $V(2y, R, \theta)$ is strictly decreasing and continuous in θ for all $\theta \in (0, 1]$, with $V(2y, R, \theta = 0) = u(y)$. Notice that, in contrast to Peck and Shell (2003), a banking arrangement will always emerge as an equilibrium since in our environment $V(2y, R, \theta) > u(y)$ for all $\theta > 0$.[18]

The risk-sharing arrangement that prevails depends on the magnitude of θ, the probability of a sunspot. Just as in Peck and Shell (2003), for θ sufficiently large, the bank may want to eliminate panics by offering a contract that is *panic proof*. The most obvious way to render the risk-sharing arrangement panic proof is for the bank to invest in at least κ units of capital in all states $m < N$. (In state $m = N$, all depositors contact the bank at date 1 so the bank will return all of its Ny potential investment assets to the depositors.) When $\kappa = 2y$ the best panic-proof contract, $(\bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2)$, is given by $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ for all $m \neq N - 1$ and, when $m = N - 1$,

$$[\bar{c}_1(N - 1), \bar{c}_2(N - 1)] = [y(N - 2)/(N - 1), 2Ry].$$

Contract $(\bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2)$ is panic proof because it is incentive compatible for each state of the world $n \in \{0, \ldots, N - 1\}$. In particular, a patient individual has no incentive to contact the bank at date 1 even if he thinks that the remaining $N - 1$ depositors contact the bank at date 1 since $\bar{c}_2(N - 1) = 2Ry > y = \bar{c}_1(N)$. Because $\bar{c}_1(N - 1) < \hat{c}_1^L(N - 1) < y = \hat{c}_1^L(N - 2)$,

---

[18]In Peck and Shell (2003), if θ is sufficiently high agents will prefer autarky since their in environment, which assumes a sequential service constraint, the expected utility associated with playing the panic equilibrium is strictly less than $u(y)$.

modifying the efficient contract $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ to the panic-proof contract $(\bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2)$ can be interpreted as imposing an additional "haircut" on early redemptions in the very high redemption state $m = N - 1$. Eliminating bank panics in this manner necessarily implies poorer risk sharing, since the allocation in state $n = N - 1$ is inefficient *ex post*.

Let $Q(\kappa, R)$ denote the expected utility associated with the panic-proof allocation $(\bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2)$, i.e.,

$$Q(2y, R) \equiv EU[(\bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2)].$$

Notice that for $n \leq N - 2$

$$\left(\frac{n}{N}\right) u[\bar{c}_1(n)] + \left(\frac{N-n}{N}\right) u[\bar{c}_2(n)] > u(y)$$

and that for $\sigma \geq 2$,

$$\left(\frac{N-1}{N}\right) u\left(\frac{N-2}{N-1}y\right) + \left(\frac{1}{N}\right) u(2Ry) < u(y).$$

If π_{N-1} is sufficiently large, then it is possible that $Q(2y, R) < u(y)$, which implies that autarky is preferred to the panic-proof allocation $(\bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2)$. In this situation, the equilibrium outcome for the economy will be for the bank to offer allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ to depositors, which they will accept. This allocation carries with it the risk of bank panics occurring with probability θ.

If, instead, $Q(2y, R) > u(y)$, then for all $N < \infty$ there exists a $\theta_0 \in (0, 1)$ such that

$$V(2y, R, \theta_0) = Q(2y, R) \tag{19}$$

since $V(2y, R, \theta = 0) > Q(2y, R)$ and $V(2y, R, \theta = 1) = u(y)$. For any $\theta < \theta_0$, depositors strictly prefer allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ to $(\bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2)$, which leaves them exposed to a bank panic. For any $\theta > \theta_0$, depositors will instead prefer the panic-free allocation $(\bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2)$ to the panic-prone allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$.

To close this section, note that the expected cost, in the form of poorer risk-sharing, associated with eliminating panics is decreasing in N since the probability π_{N-1} is strictly decreasing in the size of the depositor base N. It follows then that larger banks or a more interconnected banking system can be made more stable at a lower expected cost. Thus, the constrained-efficient risk-sharing arrangement may entail a panic-proof contract if the

susceptibility of a society to coordination failure, as indexed here by θ, is sufficiently high, or if the banking system has a sufficiently large depositor base or is otherwise sufficiently well diversified as indexed by the parameter N. We summarize this results in the section in the following proposition,

Proposition 4 *If $Q(2y, R) < u(y)$ or $Q(2y, R) > u(y)$ and $\theta < \theta_0$, then depositors prefer the panic-prone allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$; otherwise, depositors prefer the panic-proof allocation $(\bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2)$.*

5 Narrow banks, shadow banks and reach-for-yield

Assume now that there are two fundamentally risk-free investments in the economy parameterized by $\{(\kappa_1, R_1), (\kappa_2, R_2)\}$, where $(\kappa_1, \kappa_2) = (0, 2y)$, $(R_1, R_2) = (\delta, R)$ and $1 \leq \delta < R$. The reader is invited to think of (κ_1, R_1) as a money market fund or bank invested in safe government bonds and (κ_2, R_2) as representing a fundamentally safe class of business investments. Alternatively, one might think of (κ_1, R_1) as being funded by narrow banks and (κ_2, R_2) by shadow banks.

Let $(\mathbf{c}_1^{1*}, \mathbf{c}_2^{1*})$ denote the unconstrained efficient allocation associated with the money fund investment $(\kappa_1, R_1) = (0, \delta)$. Since $\kappa_1 = 0$, the money fund is panic free. Let $(\hat{\mathbf{c}}_1^2, \hat{\mathbf{c}}_2^2)$ denote the unconstrained efficient allocation associated with the bank funding business investments, where $(\kappa_2, R_2) = (2y, R)$. Assume that the probability of a sunspot θ is characterized by $0 < \theta < \theta_0$, where θ_0 is defined in (19). Since $0 < \theta < \theta_0$, bank contract $(\hat{\mathbf{c}}_1^2, \hat{\mathbf{c}}_2^2)$ is exposed to panic attacks and is preferred to a panic-free allocation.

Let $V(0, \delta, 0)$ represent the expected utility associated with allocation $(\mathbf{c}_1^{1*}, \mathbf{c}_2^{1*})$ and let $V(2y, R, \theta)$ denote the expected utility associated with allocation $(\hat{\mathbf{c}}_1^2, \hat{\mathbf{c}}_2^2)$. Clearly, $V(0, \delta, 0)$ is strictly increasing in δ with $V(0, 1, 0) = u(y)$. Given the properties of $V(0, \delta, 0)$ and $V(2y, R, \theta)$, it is evident that for a given $\theta \in (0, \theta_0)$, there exists a $\delta_0 \in (1, R)$ such that

$$V(0, \delta_0, 0) = V(2y, R, \theta). \tag{20}$$

Hence, there exists a risk-return trade off, where the risk is extrinsic here. Because our model assumes that investors have identical preferences, only

one of the two funds will emerge in equilibrium; that one which generates the highest expected utility for depositors.[19]

Our two investment economy can help us put some structure of the concept of "reach for yield," a term that is used extensively in the popular press. To begin, consider an environment in which $\delta_0 < \delta < R$. In this case, depositors prefer the narrow banking arrangement $(\mathbf{c}_1^{1*}, \mathbf{c}_2^{1*})$ because $V(0, \delta, 0) > V(2y, R, \theta)$. Now suppose that the rate of return on investment funded in the narrow banking sector declines to δ', where $1 < \delta' < \delta_0$. The cause of this decline in the "safe real interest rate" is immaterial here.[20] The decline in the safe real interest rate will induce a portfolio reallocation. In particular, since $V(0, \delta', 0) < V(2y, R, \theta)$, depositors are motivated to move their resources out of the narrow bank and into the shadow bank with associated bank contract $(\hat{\mathbf{c}}_1^2, \hat{\mathbf{c}}_2^2)$. But the new risk-sharing arrangement $(\hat{\mathbf{c}}_1^2, \hat{\mathbf{c}}_2^2)$ is subject to panics. Hence, there is a sense in which our model supports the notion of low real interest rates motivating a reach-for-yield behavior that renders the financial system less stable and more prone to panic attacks; see, for example, Stein (2013).

6 Discussion

6.1 Relation to the literature on liquidation costs

We can reinterpret our model in a manner that makes it more comparable to the related literature.[21] Suppose that the aggregate endowment Ny is invested entirely as capital at date 0. At date 1, capital can be *liquidated* at a unit rate of return. As long as the level of liquidation does not exceed a threshold, $Ny - \kappa$, any remaining capital yields a rate of return $R > 1$ in date 2. For levels of liquidation that exceed this threshold, any remaining

---

[19]Modeling preference heterogeneity with respect to risk-tolerance would produce a model in which risk-sharing arrangements with different risk-return characteristics could coexist. We believe that much of the following intuition would survive such a generalization.

[20]Perhaps it is induced by central bank policy as an attempt to bolster the economy in the face of an imminent recession. Alternatively, imagine a wave of pessimism in a part of the globe resulting in a flight to the safety of U.S. treasury debt, leading to a decline in real yields.

[21]We thank Todd Keister for suggesting this interpretation.

capital yields a rate of return $r < 1$. Our scale economy can be reinterpreted as costly liquidation, where the unit cost of liquidation rises if undertaken at a sufficiently high level and this happens when there is heavy redemption activity.

Costly liquidation is prominently featured in the models of Cooper and Ross (1998) and Ennis and Keister (2006). In their environments, liquidating capital at date 1 reduces capital by more than one unit. Any remaining capital provides a rate of return equal to $R > 1$. Relative to our model, this specification for costly liquidation has a very different implication for the existence of bank panics.

In our environment, assuming that $\kappa = 2y$, if a patient depositor believes that the other $N - 1$ depositors are contacting the bank at date 1, he is compelled to follow the crowd. In doing so, the patient depositor increases his payoff from $\hat{c}_2^L(N-1) < y$ to $\hat{c}_1(N) = y$. Or, put another way, when the bank liquidates an additional amount of capital, $\hat{k}^L(N-1) > 0$, it provides the patient depositor with an unambiguously *higher* rate of return, since $\hat{c}_1(N) > \hat{c}_2^L(N-1)$.

In the models of Cooper and Ross (1998) and Ennis and Keister (2006), if a patient depositor believes that all other depositors are contacting the bank at date 1, he may *not* have an incentive to follow the crowd.[22] If the patient depositor does in fact contact the bank at date 1, then an additional amount of capital k is liquidated, but only a fraction $0 < \lambda < 1$ of this liquidated capital, λk, is available for consumption. If the patient investor instead contacts the bank at date 2, his payoff will be Rk. Because $R > \lambda$, costly liquidation *reduces* early consumption relative to late consumption; that is, it is a force that works *against* a patient depositor misrepresenting himself.[23] In contrast, in our model efficient liquidation increases early consumption relative to late consumption; that is, it is it *reinforces* the incentive for a patient depositor to misrepresent himself.

---

[22]Cooper and Ross (1998) and Ennis and Keister (2004) assume inefficient contracts, a continuum of depositors, and sequential service. The present discussion assumes a finite number of depositors, that the bank contract is efficient, and that depositors do not know their place in the sequential service queue, as assumed in Peck and Shell (2003).

[23]Whether the patient depositor chooses to travel at date 1 or date 2 entails the same sort of calculus as in Peck and Shell (2003).

6.2 Recent financial stability regulations

Our theory suggests that bank panics are very likely to exist in any environment where there is: (i) a private information over liquidity needs, and (ii) a desire to finance fundamentally safe investments beyond a minimum scale with demandable or short-term debt. Organizations that fund their working capital using bank credit lines or commercial paper seem particularly vulnerable. If funding in this form is suddenly pulled in sufficient volume, these organizations could see the value of their long-term operations significantly decline. This, in turn, can reinforce a bleak outlook on the backing of their remaining debt. Something along these lines seems to have occurred on September 16, 2008, when the Reserve Primary Fund broke the buck. News of this event triggered a large wave of redemptions in the money market sector, especially from funds invested in commercial paper. The wave of redemptions ceased only after the U.S. government announced it would insure deposits in money market funds, essentially rendering them panic-free.[24]

Even though at that time mutual funds allowed their depositors to withdraw their investments on demand with impunity at a fixed par exchange rate, our model suggests that if these funds were priced using a net-asset-valuation (NAV) method, there might still have been a panic. In our model, promised rates of return are made contingent on market conditions, i.e., aggregate redemption demand, and this can be interpreted as a form of NAV pricing of liabilities. Our efficient risk-sharing arrangement is permitted to break the buck in very heavy redemption states. But note that our flexible NAV-like pricing structure does not eliminate panics.

On July 23, 2014, the Securities Exchange Commission announced money market reforms that included the requirement of a floating NAV for institutional money market funds, as well as the use of liquidity fees and redemption gates to be administered in periods of stress or heavy redemption.[25] While our model suggests that NAV pricing of demandable liabilities by itself is not sufficient to prevent panics, the use of liquidity fees and redemption gates is consistent with eliminating panics in our model economy. For example, the difference in consumption levels between panic-prone and panic-free allocations, $\hat{c}_1(N-1) - \bar{c}_1(N-1) > 0$ described in Section 4.1, can be interpreted

---

[24]See Kacperczyk and Schnabl (2010).
[25]See: https:// www.sec.gov/ News/ PressRelease/ Detail/ PressRelease/ 1370542347679

as a liquidity fee that depositors pay to obtain funds when redemption activity is judged by the directors of a market fund to be unusually high. This liquidity free prevents the bank-panic equilibrium.

Other post-financial crisis regulations also take aim at reducing banks' reliance on short-term borrowing. For example, recent Basil *liquidity ratio* regulations are designed to incentivize banks to borrow longer term. The liquidity ratio requires that banks be able to withstand a significant liquidity outflow for a period of 30 days. A bank is better able to survive such a liquidity event if it lends short, which means that it will receive cash during the liquidity event, and borrows long, which means there is a high probability that is will not have to pay off loans during the liquidity event. In the context of our model, this regulation can be interpreted as providing the bank an incentive to have at least κ units of its loans in the form of long-term 2-period debt. This long term debt pays off at date 2. This implies that the bank will always has at least κ invested in the high return project. An implication of this long-term borrowing is that economy will be panic free. There is, however, an additional cost is associated with long-term borrowing. In the event that all N depositors withdraw early for fundamental reasons, which is an event that occurs with low probability when N is large, the bank will only be able to distribute $(N-2)y$ units date 1 consumption. This implies that $2y$ units will be wasted. If, however, the probability of a sunspot, θ, is sufficiently high, then a regulation that requires banks to undertake some long-term borrowing can generate higher welfare than the allocation associated with only short-term borrowing that is subject to a panic.[26]

6.3 Bank panics and bank size

Based on an examination of the historical record of financial crises, Calomiris and Gorton (1991) conclude that institutional factors, such as branch bank

---

[26]The bank may want to invest in long-term debt, even though it is costly, for commitment reasons. For example, if the bank invests in short term debt and promises to keep at least κ invested for all $m \leq N - 1$, as in allocation $(\bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2)$, then if $N - 1$ depositors contact the bank at date 1 and the bank lacks commitment, it may have an incentive to to let the investment fall below κ. A bank will have such an incentive if its objective is to maximize the welfare of the entire community, conditional of the number of contacts it receives at date 1.

laws and bank cooperation arrangements, play an important role in determining both the frequency and magnitude of bank panics. Williamson (1989) studies the institutional differences between the banking systems in Canada and the United States during the U.S. National Banking era. The upshot of the historical evidence is that banking systems with fewer but larger banks appear more resilient to panics. Note that "largeness" here should be taken to include the propensity for banks to engage in branch banking activities and to cooperate with other banks in an interbank market, possibly in conjugation with a central bank. In our model, the parameter indexing the size of a bank or banking system is N. Our model predicts that larger banks or a more interconnected banking system are better because larger and more diversified banks are less likely to experience extremely high redemption states, thereby lowering the expected cost associated with extra haircuts in high redemptions states needed to discourage panic behavior.

References

1. Allen, Franklin and Douglas Gale (1998). "Optimal Financial Crises," *The Journal of Finance*, 53(4): 1245-1284.

2. Andolfatto, David, Nosal, Ed and Neil Wallace (2007). "The Role of Independence in the Green-Lin Diamond-Dybvig Model," *Journal of Economic Theory*, 137(1): 709–715.

3. Andolfatto, David, Nosal, Ed and Bruno Sultanum (2016). "Preventing Bank Runs," Forthcoming in *Theoretical Economics*.

4. Bryant, John (1980). "A Model of Reserves, Bank Runs, and Deposit Insurance," *Journal of Banking Finance*, 43: 749–761.

5. Calomiris, Charles and Gary Gorton (1991). "The Origins of Banking Panics: Models, Facts, and Bank Regulation," in *Financial Markets and Financial Crises*, edited by Glenn Hubbard, University of Chicago Press, 109–174.

6. Capiro, Gerard, Jr. and Daniela Klingebiel (1997). "Bank Insolvency: Bad Luck, Bad Policy, or Bad Banking?" in M. Bruno and B. Pleskovic, eds., *Annual World Bank Conference on Development Economics*. Washington DC: World Bank, 79–104.

7. Cooper, Russell and Ross, Tom (1998). "Bank Runs: Liquidity Costs and Investment Distortions," *Journal of Monetary Economics* 41, 27–38.

8. Diamond, Douglas and Philip Dybvig (1983). "Deposit Insurance and Liquidity," *Journal of Political Economy*, 91(June): 401–419.

9. Ennis, Huberto M. and Todd Keister (2006). "Bank Runs and Investment Decisions Revisited," *Journal of Monetary Economics*, 52(2): 217–232.

10. Ennis, Huberto M. and Todd Keister (2010). "On the Fundamental Reasons for Bank Fragility," FRB Richmond *Economic Quarterly* 96, First Quarter: 33-58.

11. Gorton, Gary (1988). "Banking Panics and Business Cycles," *Oxford Economic Papers*, 40: 751-781.

12. Green, Edward and Ping Lin (2000). "Diamond and Dybvig's Classic Theory of Financial Intermediation: What's Missing?" Federal Reserve Bank of Minneapolis *Quarterly Review*, 24 (Winter): 2–13.

13. Green, Edward and Ping Lin (2003). "Implementing Efficient Allocations in a Model of Financial Intermediation," *Journal of Economic Theory*, 109(1); 1–23.

14. Kacperczyk, Marcin and Philipp Schnabl (2010). "When Safe Proved Risky: Commercial Paper during the Financial Crisis 2007–2009," *Journal of Economic Perspectives*, 24(1): 29–50.

15. Keiste, Todd and Vijay Narasiman (2016). "Expectations vs. Fundamentals-based Bank Runs: When Should Bailouts be Permitted?" *Review of Economic Dynamics*, 21: 89-104.

16. Peck, James and Karl Shell (2003). "Equilibrium Bank Runs," *Journal of Political Economy*, 111(1): 103–123.

17. Stein, Jeremy (2013). "Yield-Oriented Investors and the Monetary Transmission Mechanism," remarks presented at "Banking, Liquidity and Monetary Policy," a Symposium Sponsored by the Center for Financial Studies in Honor of Raghuram Rajan.

18. Townsend, Robert M. (1987). "Economic Organization with Limited Communication," *The American Economic Review*, 77(5): 954–971.

19. Wallace, Neil (1988). "Another Attempt to Explain an Illiquid Banking System: The Diamond and Dybvig Model with Sequential Service Taken Seriously," *Quarterly Review* of the Federal Reserve Bank of Minneapolis, 12(4): 3-16.

20. Williamson, Stephen (1989). "Bank Failures, Financial Restrictions, and Aggregate Fluctuations: Canada and the United States, 1870–1913," Federal Reserve Bank of Minneapolis *Quarterly Review*, Summer: 20–40.