Bank panics without sequential service

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

Ed Nosal
Federal Reserve Bank of Chicago

October 2016

Abstract

Sequential service plays an important role in theory of bank panics. The interaction of sequential service with private information over liquidity needs evidently opens the door for bank panics in Diamond and Dybvig (1983). These authors go on to suggest that deposit insurance eliminates bank panics. Wallace (1988) has questioned the logic of their prescription since it appears to rely on deposit insurance somehow overcoming sequential service. In our paper, we provide a theory of bank panics that is based on the interaction of private information with scale economies–we make no use of sequential service. Panics in our model occur when funding is pulled in sufficient quantities from the banking system to precude it from operating at scale. Deposit insurance is feasible in our environment and may be desirable.

1 Introduction

Investors in market economies have access to a wide array of financial products with contractual options that can be exercised at investor discretion. Bryant (1980) suggests that such options are a means to insure investors against unobservable risks. Market economies are also subject to recurring financial crises. It has long been suspected that the two phenomena are linked, with the direction of causality running from financial product design to financial instability. Establishing this causal link–either empirically or theoretically–however, has proven elusive.[1] This paper concerns the theory of bank panics, by which we take to mean financial instability as a by-product of financial product design interacting with psychological (rather than fundamental) factors influencing market expectations.

[1] An alternative view is that financial instability is merely symptomatic of difficulties experienced in the broader economy. Gorton (1988), for example, notes that banking crises are frequently preceded by economic recession.

Central to the theory of bank panics in the tradition of Diamond and Dybvig (1983) is the notion of *sequential service*–the practice of satisfying depositor withdrawal demands on a first-come-first-served basis. Absent sequential service, Green and Lin (2000) demonstrate how the optimal banking arrangement in Diamond and Dybvig (1983) conditions early withdrawal amounts on aggregate withdrawal demand. Optimality dictates that the "haircut" on early redemptions is increasing in the aggregate volume of early withdrawals. These state-contingent haircuts can be structured so that sufficient resources remain in the bank to satisfy the promises made to those who are willing to delay redemption at a higher rate of return. As a consequence, patient depositors have no incentive to misrepresent their liquidity needs. If depositors collectively understand this, then they need never fear a bank panic.

Sequential service is clearly not sufficient to produce an equilibrium bank panic (Green and Lin, 2000). Even in the basic Diamond and Dybvig (1983) model, where aggregate uncertainty over liquidity demand is absent, sequential service only produces a panic equilibrium if the deposit contract fails to suspend payments in the event that reserves are depleted–a practice that is evidently observed in historical banking crises. On the other hand, Peck and Shell (2003) show how a bank panic is possible even under a contractual arrangement that is optimal in a class of direct mechanisms. Sequential service is evidently necessary to produce this result. Andolfatto, Nosal and Sultanum (2016), however, demonstrate that sequential service need not impede efficient implementation in a broader class of mechanisms.

Thus, the role of sequential service in the theory of bank panics is not entirely clear. Wallace (1988) forcefully contends that some form of communication barrier in the early withdrawal period is at the very least necessary to explain illiquid bank structures in the Diamond and Dybvig (1983) framework. If all depositors wanting to make an early withdrawal were instead in communication with each other and the bank, then the liquidity insurance problem is solved through the use of state-contingent haircuts. Moreover, absent some communication barrier, Wallace (1988) points out that depositors could–and generally would want to–participate in a one-period credit market after their liquidity needs are realized. As demonstrated by Jacklin (1987), the opportunity for *ex post* trade renders illiquid banking an unsustainable enterprise.[2]

In this paper, we revisit the question of whether bank panics are theoretically possible in a version of the Diamond and Dybvig (1983) model where sequential service for early arrivals is absent and the banking arrangement is optimal given the properties of the environment. While the practice of first-come-first-served is common in retail settings, we do not believe it constitutes an essential ingredient for financial instability. The phenomenon of panic-induced runs on local bakeries, for example, is not something that resonates with most people. A more

[2]Although depositors would not voluntarily participate in an illiquid bank (sequential rationality constraints are violated), involuntary participation through a trading restriction can improve *ex ante* welfare.

promising avenue, in our view, is to explore the role that scale economies in investment and intermediation may play in determining the fragility of funding structures.

There is some evidence to suggest that banking is subject to scale economies; see, for example, Hughes and Mester (2013), Beccali, Anolli, and Borello (2015), and Wheelock and Wilson (2016). We model the scale economy as being rooted in the nature of the investment technology itself. That is, the return to investment is subject to thresholds that, if surpassed, permit larger scale investments to earn higher rates of return. All firms, including non-bank enterprises, potentially have access to the same investment technology. The distinguishing characteristic of banks is how these investments are funded. In particular, banks are compelled (presumably through competition) to offer their depositor base structured liability products that provide the flexibility to withdraw funding on short notice (e.g., demand deposit liabilities).[3] It is the interaction of this liability structure (itself a by-product of asymmetric information) with the underlying non-convexity in asset returns that potentially opens the door to a bank panic.

2 The model

The model is based on Green and Lin (2000, 2003), which is a finite-trader version of the Diamond and Dybvig (1983) model with aggregate liquidity risk. The economy has two dates, $t = 1, 2$, and a finite number $N \geq 3$ of *ex ante* identical agents who are subject to a shock at date $t = 1$ that determines their preference type $\omega \in \{i, p\}$. We label a type $\omega = i$ agent *impatient* and a type $\omega = p$ agent *patient*. Let $0 < \pi < 1$ denote the probability that an agent becomes impatient. Let π_n denote the probability that $0 \leq n \leq N$ agents are impatient. We assume that agent types are *i.i.d.* so that $\pi_n = \binom{N}{n}\pi^n(1-\pi)^{N-n}$. Note that $0 < \pi_n < 1$ for all n (the distribution of types has full support).

Impatient agents want to consume at date 1 while patient agents 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 agent at date t. *Ex ante* preferences are given by

$$U(c_1, c_2) = \left\{ \begin{array}{lll} u(c_1) & \text{w.p.} & \pi \\ u(c_1 + c_2) & \text{w.p.} & 1-\pi \end{array} \right| \tag{1}$$

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

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

$$F_\kappa(k) = \left\{ \begin{array}{lll} rk & \text{if} & k < \kappa \\ Rk & \text{if} & k \geq \kappa \end{array} \right| , \tag{2}$$

[3] Moreover, because a larger more diversified depositor base lowers the cost of insurance, there is a second dimension along which scale economies matter for banks.

where $0 < r < 1 < R$ and $0 \leq \kappa < Ny$.[4] The function (2) is a generalization of the standard technology used in this literature, which assumes that $\kappa = 0$, so that $F_0(k) = Rk$ for all $k \geq 0$. Here, the high rate of return R is available only the level of investment exceeds a minimum scale requirement of κ.[5] When the minimum scale is not met there is a cost to storage, indexed by the parameter r.

The benefits of cooperation in this economy are twofold. First, there are the usual gains associated with sharing risk. Second, and absent from the usual model specification, scale economies are more accessible when resources are pooled. In what follows, we relabel agents as *depositors* and call any contractual arrangement among depositors to share risk and exploit scale economies a *bank*. Formally, the bank is a resource-allocation mechanism that pools the resources of $0 \leq \hat{N} \leq N$ depositors before they learn their types in exchange for a structured liability product that delivers consumption (withdrawal limits) at dates 1 and 2, conditional on individual and aggregate information available at each date.

An important consideration in the design of the contract is the nature of private information and communications. What makes our environment a bank problem—as opposed to a standard insurance problem—is that the demand for early consumption (withdrawals) is based on private information. Consequently, the contract must embed an option that is exercisable at the depositor's discretion at the early date—it will, in this sense, resemble a standard demand deposit liability. Moreover, for the purpose of ensuring efficient resource allocation, the contract should be structured in a way that gives depositors incentives to represent their liquidity preferences truthfully—that is, the allocation implied by the contract should be incentive-compatible.

What one assumes about communication generally matters for allocations. Here we adopt the simple direct revelation mechanism in Peck and Shell (2013). The mechanism works as follows: Agents agree to participate in the bank contract by depositing their claims to date 1 output. Depositors then observe their preference shock; they are either impatient or patient. Depositors can visit the bank either at date 1 or at date 2, but not in both periods. Notice that this assumed "travel itinerary" bears a close resemblance to actual depositor behavior. That is, if a depositor wants to make a withdrawal, they visit the bank and make their withdrawal, or else they wait to make their withdrawal at a later date. This protocol, however, places a potentially binding restriction on communications. For example, in Green and Lin (2003), efficient risk-sharing under a sequential service constraint requires that even patient agents communicate their preferences to the bank in the early period—a period in which they do not wish to make a withdrawal. We discuss the implications that our communication restriction (or friction) has on outcomes in due course.

[4] Note that the return r is realized only if k is of insufficient scale. This poor return is not to be confused with the scrapping cost modeled in Cooper and Ross (1998).

[5] One could easily generalize the analysis to permit multiple threshold levels and associated rates of return, but here we stick to one threshold level for simplicity.

Given our restriction on depositors' travel itineraries, communications concerning preference type are transmitted indirectly via the timing of a depositor's visit to the bank. As a result, consumption at each date need only be conditioned on the number of agents m visiting the bank at date 1, where $m \in \{0, 1, ..., \hat{N}\}$. In particular, if m agents travel to the bank at date 1, then each agent receives $c_1(m)$ units of date 1 consumption. Agents who visit the bank at date 2 each receive $c_2(m) = F_\kappa[\hat{N}y - mc_1(m)]/(\hat{N} - m)$ units of date 2 consumption. Hence, the bank offers depositors a contract (or allocation) $(\mathbf{c}_1, \mathbf{c}_2)$, where $\mathbf{c}_1 = [c_1(1), \ldots, c_1(\hat{N})]$ and $\mathbf{c}_2 = [c_2(0), c_2(1), \ldots, c_2(\hat{N} - 1)]$.

In what follows, we restrict attention to equilibria in which $\hat{N} = N$. It seems clear enough that whenever a scale economy is present, $\hat{N} = 0$ can be an equilibrium outcome. But it is an outcome that is attributable solely to the assumed non-convexity in production and has little—if anything—to do with how investment is financed. Our concern in what follows is to examine the potential indeterminacy introduced by the funding structure itself.

3 Efficient incentive-compatible allocations

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

3.1 Linear technology

Assume, for the moment, that impatient depositors visit the bank at date 1 and that patient depositors visit at date 2. In this "truth-telling" scenario, $m = n$. Recall that there is no sequential service constraint which implies that the efficient allocation depends on the number of depositors that visit the bank at date 1.[6]

Collectively, depositors seek a contract to maximize their (*ex ante* identical) expected utility,

$$\max \sum_{n=0}^{N} \pi_n \{nu\left[c_1(n)\right] + (N - n)u\left[c_2(n)\right]\} \tag{3}$$

subject to the resource constraint,

$$nc_1(n) + \frac{(N - n)c_2(n)}{R} = Ny \tag{4}$$

[6] Green and Lin (2000, 2003) also provide a characterization of the efficient allocation when there is no sequential service constraint and when the investment technology is linear. We repeat these for the reader's convenience.

Let $(\mathbf{c}_1^*, \mathbf{c}_2^*)$ denote the solution to the problem above. Then it is easy to see that there is a unique solution that satisfies

$$u'[c_1^*(n)] = Ru'[c_2^*(n)] \ \forall n \tag{5}$$

together with the resource constraint (4). Given our CES preference specification, the solution is available in closed-form,

$$c_1^*(n) = \frac{Ny}{n + (N-n)R^{1/\sigma - 1}} \tag{6}$$

$$c_2^*(n) = R^{1/\sigma} c_1^*(n) \tag{7}$$

Notice that $\sigma > 1$ implies $\left(1 - R^{1/\sigma - 1}\right) \gtreqless 0$ as $R \gtreqless 1$. Since $R > 1$ here, it follows that both $c_1^*(n)$ and $c_2^*(n)$ are decreasing in n. Using (6) and $k^*(n) = Ny - nc_1^*(n)$ yields the associated investment schedule,

$$k^*(n) = \frac{(N-n)R^{1/\sigma - 1}}{n + (N-n)R^{1/\sigma - 1}} Ny \tag{8}$$

Observe here that $k^*(n)$ and $k^*(n)/(N-n)$ is decreasing n and that $k^*(n) < (N-n)y$.[7]

A large value for n means that the aggregate demand for early withdrawals is high. In this case, it makes sense to devote less resources to investment, reducing the effective return for late withdrawals, spreading the additional early resources more thinly among the more numerous impatient, reducing their return as well. Note that high realizations for n are recessionary events (or investment collapses) associated with large numbers of depositors making early withdrawals. These events, however, are driven by economic fundamentals. A bank (or the banking sector) could mitigate the economic impact of these "fundamental runs" by expanding its depositor base (i.e., increasing 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 $N \to \infty$.

If depositor preferences were observable, then this would be the end of the story. (Up to this point we have assumed that only impatient depositors visit the bank early, which is equivalent to assuming that depositor preferences are observable.) However, when liquidity preferences are private information, then depositors may misrepresent themselves for private gain and at the expense of the community. Optimal depositor behavior will generally depend on how they believe other depositors might behave. Hence, private information renders the environment strategic, where depositors play a game among themselves.

The game that depositors play in this environment is simple. In the *depositor game*, which is played after agents accept the contract by depositing their date 1 endowment claims at the bank and learn their type, depositor $j \in \{1, 2, ..., N\}$

[7] The latter inequality is a direct implication of risk sharing. If we replace R with r in (8), then $k^*(n)$ is still decreasing in n but $k^*(n)/(N-n)$ is increasing.

must choose an action $t_j \in \{1, 2\}$, where t_j denotes the date depositor j chooses to visit the bank. A given *strategy profile* $\mathbf{t} \equiv \{t_1, t_2, ..., t_N\}$ implies a number $m \in \{0, 1, ..., N\}$, the number of depositors visiting the bank at date 1. We define a *truth-telling strategy* to be a strategy profile consisting of impatient depositors visiting the bank at date 1 and patient depositors visiting the bank at date 2. For a truth-telling strategy, $m = n$. We define a *panic strategy* to be a strategy profile consisting of all depositors visiting the bank at date 1. For a panic strategy, $m = N$.

A strategy profile $\mathbf{t}$ (and its associated m) is said to be an (Bayes-Nash) *equilibrium* of the depositor game with allocation $(\mathbf{c}_1, \mathbf{c}_2)$ if $t_j \in \mathbf{t}$ constitutes 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. Mathematically, incentive-compatibility requires

$$\sum_{n=0}^{N-1} \Pi^n u\left[c_2(n)\right] \geq \sum_{n=0}^{N-1} \Pi^n u\left[c_1(n+1)\right] \tag{9}$$

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

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

There are two important results associated with the solution $(\mathbf{c}_1^*, \mathbf{c}_2^*)$. First, it is incentive-compatible. To see this, note that when $R > 1$, (6) and (7) imply that $c_2^*(n) > c_1^*(n) > c_1^*(n+1)$. In words, a patient agent's consumption is higher than an impatient agent's consumption in a truthtelling equilibrium. And, if a patient agent chooses to visit the bank at date 1, then date 1 consumption for all agents is lower compared to what they would enjoy if the patient agent instead traveled at date 2. Therefore, assuming that all other depositors are playing truthfully, a patient agent has no incentive to travel at date 1 since doing so would result in a strictly lower payoff.

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 visit the bank at date 1. In doing so, they represent themselves truthfully. The question, as usual, concerns the behavior of patient depositors. Let the patient depositor conjecture anything about the behavior of other depositors; any such conjecture generates an $m \in \{1, 2, ..., N\}$. If the patient depositor plays truthfully, he gets $c_2^*(m) > c_1^*(m) > c_1^*(m+1)$, where these inequalities follow from the stated properties of the allocation (6) and (7) when $R > 1$. In other words, truth-telling for the patient depositor is a dominant strategy. Regardless of the volume of early withdrawals, a patient depositor will always earn a higher return by leaving his resources with the bank and letting the deposit earn a high rate of return.

3.2 Scale economies

Consider now the same economy but with $0 < \kappa < Ny$ (recall that $N \geq 3$) and let's characterize the efficient incentive compatible allocation. We proceed by first assuming truth-telling on the part of depositors. Let $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ denote the efficient truth-telling allocation under our assumed scale economy. We later verify that the allocation is incentive-compatible.

A given $0 < \kappa < Ny$ will lie in the interval $[k^*(n-1), k^*(n)]$ for some $n \geq 1$. Let us without loss of generality assume that $\kappa = k^*(2)$.[8] This implies that in any economy for which there are at least two patient depositors, the investment level $k^*(n)$ is feasible. It immediately follows that $\{\hat{c}_1(n), \hat{c}_2(n)\} = \{c_1^*(n), c_2^*(n)\}$ for $n \in \{0, 1, ..., N-2\}$. Moreover, it is also clear that $\hat{c}_1(N) = y = c_1^*(N)$. Consequently, we have

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

We now characterize the efficient allocation in the event of a single patient depositor, $[\hat{c}_1(N-1), \hat{c}_2(N-1)]$. Since $k^*(1) < \kappa$, the efficient allocation will be characterized by either $\hat{k}(N-1) < \kappa$, a project return of r and efficient risk sharing or $\hat{k}(N-1) = \kappa$, a project return of R and "inefficient" risk sharing. We examine each in turn.

and Recall that any investment $k < \kappa$ delivers a return rk, where $r < 1$. The benefit associated with increasing investment is realized only when the threshold level is reached. Therefore, the constrained-efficient level of investment will be determined either by (8) with r replacing R, or by the minimum level of investment needed to achieve scale, κ. We can rule out this latter case by considering economies with N sufficiently large render $k^*(N-2) = \kappa > 2y$. In particular, note from (8) that $k^*(n)$ is monotonically increasing in N for any n.

Consider first a contract allocation that chooses $\hat{k}(N-1) < \kappa$. Then, using (6) and (7) with r replacing R, early and late consumption allocation are given by

$$\hat{c}_1(N-1) = \left[\frac{N}{N-1+r^{1/\sigma-1}}\right] y \tag{10}$$

$$\hat{c}_2(N-1) = r^{1/\sigma}\hat{c}_1(N-1) \tag{11}$$

Recall that $A = 1 - r^{1/\sigma} < 0$. Consequently, $-A = -1 + r^{1/\sigma-1} = B > 0$ so

$$\hat{c}_1(N-1) = \left[\frac{N}{N+B}\right] y$$

$\hat{c}_1(N-1) > y$. To determine the properties of $\hat{c}_2(N-1)$, define $\alpha \equiv r^{1/\sigma} < 1$. Thus

$$\hat{c}_2(N-1) = \left[\frac{\alpha N}{N-1+\alpha r^{-1}}\right] y$$

[8] We discuss below how this assumption is without loss of generality.

For $\hat{c}_2(N-1) < y$ to be true, we need the following to hold,

$$\begin{aligned} \left[\frac{\alpha N}{N-1+\alpha r^{-1}}\right] &< 1 \\ \alpha N &< N-1+\alpha r^{-1} \\ 1-\alpha r^{-1} &< N(1-\alpha) \end{aligned}$$

or

$$N > \frac{1-\alpha/r}{1-\alpha}$$

This condition will hold for sufficiently large N.

Property 2 $\hat{c}_2(N-1) < y = \hat{c}_1(N)$.

Property 2 means is that *if* a depositor knew he was the only patient one—this is not something he knows until all bank visits are completed—he would misreport himself as impatient. Notice that as $r \to 1$, $\hat{c}_2(N-1) \to y$. Hence, if r is arbitrarily close to 1, $\hat{c}_2(N-1) \approx \hat{c}_1(N) = y$.

Consider now a contract allocation $[\hat{c}_1(N-1), \hat{c}_2(N-1)]$ that chooses $\hat{k}(N-1) = \kappa = k^*(N-2)$. Such an allocation implies that the project return is high, $R > 1$, as well as $\hat{c}_1(N-1) < c_1^*(N-1)$ and $\hat{c}_2(N-1) > c_2^*(N-1)$. So, although allocation $[\hat{c}_1(N-1), \hat{c}_2(N-1)]$ gives a high project rate of return is high, it does so at the cost of (worse) risk-sharing. Since $\hat{k}(N-1) = k^*(N-2)$, $\hat{c}_1(N-1) = YN - \hat{k}(N-1)$ and $\hat{c}_2 = R\hat{k}(N-1)$, we have

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

$$\hat{c}_2(N-1) = \frac{2R^{1/\sigma}}{N-2+2R^{1/\sigma-1}}y \tag{13}$$

Proposition 1 *The allocation* $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2) = (\mathbf{c}_1^*, \mathbf{c}_2^*)$ *for all* $n \in \{0, 1, ..., N-2, N\}$ *with* $[\hat{c}_1(N-1), \hat{c}_2(N-1)]$ *given by (10) and (11) is constrained efficient incentive-compatible if* $r < 1$ *is sufficiently close to unity and if* N *is sufficiently big.*

Proof. We first show that allocation (10)-(11) generates higher welfare than allocation (12)-(13). The we show that the allocation stated in the Proposition is incentive compatible. When r is sufficiently close to 1, then allocations (10)-(11) are sufficiently close to y, i.e., $\hat{c}_1(N-1) = \hat{c}_2(N-1) = y$ and social welfare is arbitrarily close to $(1-\sigma)^{-1}Ny^{1-\sigma} \equiv W_1$. Social welfare associated allocation (12)-(13) is

$$(N-1)[\frac{N-2}{N-2+2R^{1/\sigma-1}}y]^{1-\sigma} + [\frac{2R^{1/\sigma}}{N-2+2R^{1/\sigma-1}}y]^{\sigma} \equiv W_2.$$

Since $\sigma > 1$, $W_1 > W_2$ if

$$(N-1)[\frac{N-2+2R^{1/\sigma-1}}{N-2}]^{\sigma-1} + [\frac{N-2+2R^{1/\sigma-1}}{2R^{1/\sigma}}]^{\sigma-1} > N. \tag{14}$$

Notice that: (1) $\sigma - 1 > 0$; (2) the first fraction on the left side of (14) is greater than 1; and (3) the second fraction is greater than 1 if

$$N - 2 + 2R^{1/\sigma-1} > 2R^{1/\sigma}. \tag{15}$$

Clearly, if N is sufficiently large (for reasonable values of R, $N = 3$ is sufficiently large), then (15) is valid, as is (14), all of which implies that $W_1 > W_2$. Hence, the efficient truth-telling contract is the one stated in the Proposition. Since $\hat{c}_2(n) > \hat{c}_1(n+1)$ for all $n \neq 1$ and $\hat{c}_2(N-1) \approx \hat{c}_1(N) = y$, allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2) = (\mathbf{c}_1^*, \mathbf{c}_2^*)$ for all $n \in \{0, 1, ..., N-2, N\}$ with $[\hat{c}_1(N-1), \hat{c}_2(N-1)]$ given by (10) and (11) is incentive compatible, i.e., it satisfies (9). ■

4 Panic equilibria

Before agents fund a bank and become depositors, they make the usual participation decision: should they join the risk-sharing arrangement or not? Collectively, this participation decision determines $0 \leq \hat{N} \leq N$. Autarky is always a feasible option and so places a lower bound on *ex ante* utility. Assuming that no individual agent possesses sufficient resources to operate investment at scale, i.e., $\kappa > y$, the autarkic payoff is $u(y)$. Autarky is always an equilibrium in the participation game when $\kappa > y$. That is, if an agent believes that all other agents are choosing not to participate in the banking arrangement, then it is optimal for an agent to follow the crowd. However, this is not the indeterminacy we are interested in. In what follows, we assume depositors can coordinate on the participation outcome.

To highlight the role of private information in this economy, let us imagine for the moment that it is absent. In this case, depositors traveling to the bank at date 1 automatically reveal their type. A patient depositor visiting at date 1 is not supposed to be there, so the bank can refuse to service him (in accordance to the contractual terms which would have specified this denial-of-service stipulation *ex ante*). Assuming that the mechanism can commit to the threats it makes along off-equilibrium paths, no patient depositor would ever have an incentive to misrepresent himself. That is, in the absence of private information, and conditional on participation, $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ is uniquely implementable as a truth-telling equilibrium.

We now state our main result.

Proposition 2 *If liquidity preference is private information, then allocation* $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ *admits a panic equilibrium, where all N agents exercise their option to withdraw funding from the bank at date 1, regardless of their true liquidity needs.*

To check the validity of the statement in Proposition 2, begin by proposing a strategy profile in which all N depositors visit the bank at date 1. The question is whether a patient depositor has an incentive to follow the crowd, or leave his money in the bank, with the intent of withdrawing at the later date. If he plays the proposed panic strategy, he receives a payoff $\hat{c}_1(N) = y$. If he instead defects from the panic strategy and travels to the bank at date 2, by Property 2 he receives $\hat{c}_2(N-1) \approx y$ but $\hat{c}_2(N-1) < y$. Consequently, a patient depositor is strictly motivated to panic if he conjectures that other patient depositors are as well.

5 Eliminating panics

Let θ denote the probability of a sunspot–an event that coordinates depositors to panic. Then the expected utility of participating in the bank is computed using a modified probability distribution function. The probability of no panic is $(1-\theta)$ and the probability of the full redemption state N is π_N. The probability of the full redemption state regardless of the true N is θ. So, the probability of the full redemption state is $(1-\theta)\pi_N + \theta$. Or something like this (we can work it out).

The mechanism could eliminate panic equilibria by guaranteeing that a minimum of κ units of output will always be held back for investment. This leads to risk-sharing inefficiency in state $n = N - 1$. Even worse, I wonder if the κ units must be wasted in state $n = N$. Whatever the case may be, eliminating panics comes at a cost. The cost must be weighed against the potential benefits which, in turn, will depend on θ.

6 Discussion

To summarize, the scenario we model is as follows. Think of a local economy consisting of a number of people who are in a position to fund a local investment project, like a shopping mall. A minimum level of investment is needed if the project is to offer a high return–building one or two retail spaces is pointless. Somewhat unrealistically, the return to the project does not diminish with scale, but this is easily relaxed. The problem is that local depositors are subject to "unanticipated expenditure events" that they will have to funded with the resources tied up in the investment project.[9] A standard insurance contract is not possible because there is no way for a third party to verify the expenditure event–it is private information. The bank is a mechanism that provides the desired liquidity insurance by permitting depositors to withdraw their funding on demand (on their initiative). Because we do not assume sequential service, early

[9] Implicitly, the investment project is illiquid in the sense that stakes in the project cannot be easily traded on a secondary market.

withdrawal limits can be conditioned on the aggregate demand for early funds. The bank proceeds with the investment project according a scale that depends on the amount of funds remaining after early redemption requests are satisfied. Notice that if the available funding falls short of the minimum scale requirement, the bank is unable to fund a high rate of return investment. Hence, in addition to the usual reason for pooling resources in a Diamond-Dyvbig environment—risk sharing—depositors will also want to pool to reduce *funding risk*, i.e., the risk that there is insufficient resources to fund the capital project at a scale sufficient to realize a high rate of return.

References

1. Beccali, Elena, Mario Anolli and Giuliana Borello (2015). "Are European banks too big? Evidence on economies of scale," *Journal of Banking & Finance*, 58(C): 232–246.

2. Bryant, John (1980). "A model of reserves, bank runs, and deposit insurance," *Journal of Banking Finance*, 43: 749–761.

3. Cooper, Russell and Ross, Tom W. (1998). "Bank runs: liquidity costs and investment distortions", *Journal of Monetary Economics* 41, 27–38.

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

5. Hughes, Joseph P. and Loretta J. Mester (2013). "Who said large banks don't experience scale economies? Evidence from a risk-return-driven cost function," Federal Reserve Bank of Philadelphia Working Paper No. 13-13/R (forthcoming in the *Journal of Financial Intermediation*).

6. Wheelock, David and Paul W. Wilson (2016). "The evolution of scale economies in U.S. banking," Federal Reserve Bank of St. Louis Working Paper 2015-021.